

ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES

THÈSE

Pour obtenir le grade de
Docteur de l'École des Hautes Études en Sciences Sociales en Sciences Économiques

Présentée et soutenue publiquement le 5 décembre 2008 par

Julien GRENET

DÉMOCRATISATION SCOLAIRE, POLITIQUES ÉDUCATIVES
ET INÉGALITÉS : UNE ÉVALUATION ÉCONOMIQUE

Directeur de thèse : M. Thomas PIKETTY

Composition du jury :

M. Robert GARY-BOBO	Professeur à l'Université de Paris I - Panthéon-Sorbonne
M. Francis KRAMARZ	Inspecteur général de l'Insee
M. Stephen MACHIN	Professeur à University College London
M. Éric MAURIN	Directeur d'études à l'EHESS
M. Thomas PIKETTY	Directeur d'études à l'EHESS
Mme Fabienne ROSENWALD	Administratrice de l'Insee

Remerciements

Je souhaiterais avant tout remercier Thomas Piketty, qui m'a encouragé à travailler sur les questions éducatives et a dirigé cette thèse avec rigueur et exigence. Ses conseils avisés, sa confiance et son enthousiasme m'ont été de précieux auxiliaires. Je lui suis tout particulièrement reconnaissant d'avoir su me transmettre sa conception d'une science économique ouverte sur le monde et en dialogue avec les autres sciences sociales.

Je remercie Robert Gary-Bobo et Francis Kramarz d'avoir accepté d'être rapporteurs de cette thèse et Steve Machin, Éric Maurin et Fabienne Rosenwald d'avoir accepté d'être membres du jury.

Des remerciements spéciaux vont à Steve Machin, qui m'a permis de réaliser une partie de mes travaux dans le cadre d'un séjour prédoctoral particulièrement fructueux au Centre for Economic Performance de la London School of Economics.

Je tiens également à remercier Gabrielle Fack, avec qui j'ai produit une partie de ce travail et qui m'a convaincu de l'importance de la dimension urbaine des politiques éducatives.

J'ai bénéficié dans la poursuite de mes recherches de l'écoute et des conseils bienveillants de nombreux économistes français et étrangers que je tiens ici à remercier, en particulier Philippe Askenazy, Sandra Black, Esther Duflo, Steve Gibbons, Caroline Hoxby, Marc Gurgand, Steve Machin, Éric Maurin, Harris Selod, Miguel Urquiola, Etienne Wasmer ainsi que l'ensemble des chercheurs de l'École d'Économie de Paris et du Centre for Economic Performance.

Cette thèse n'aurait pas pu voir le jour sans l'aide de ceux qui ont accepté de me fournir les données nécessaires à la réalisation de mes travaux empiriques : Alexandre Kych et Roxanne Silberman, du centre Maurice Halbwachs ; Fabienne Rosenwald et Françoise Œuvrard, de la Direction des Études, de la Performance et de la Prospective du ministère de l'Éducation nationale ; Patrick Caille, du Rectorat de l'Académie de Paris, ainsi que la Chambre des Notaires de Paris et d'Île-de-France.

Je tiens aussi à remercier les personnes responsables des différentes institutions qui m'ont permis d'acquérir une solide formation en sciences économiques et de réaliser cette thèse dans les meilleures conditions : l'École Normale Supérieure, où j'ai effectué mes études avant de m'initier à l'enseignement dans le cadre du monitorat ; l'École

Nationale de la Statistique et de l'Administration Économique et le Master « Analyse et Politique Économiques », où j'ai commencé ma formation doctorale ; l'École des Hautes Études en Sciences Sociales, qui a financé mon allocation de recherche ; la London School of Economics, où j'ai séjourné pendant un an dans le cadre du Programme doctoral européen en économie quantitative ; University College London et le Centre for Economic Performance, où j'ai pu effectuer un séjour prédoctoral de six mois ; le Cepremap, qui a financé une partie de mes recherches ; le laboratoire Paris-Jourdan Sciences Économiques, enfin, qui m'a accueilli dans ses locaux et m'a offert le soutien matériel nécessaire à la réalisation de cette thèse. Je tient à remercier tout particulièrement Cécile Alrivie, France Artois, Marie-Christie Paoletti, Nicole Ruster, Bénédicte Sabatier, José Sastre ainsi que l'ensemble des membres du personnel administratif et technique du site Jourdan pour leur assistance compétente et efficace.

J'adresse également mes remerciements à Daniel Cohen et Dominique Goux, qui m'ont permis d'encadrer pendant trois ans les travaux dirigés de leurs enseignements à l'École Normale Supérieure.

Je remercie en outre les nombreux participants aux séminaires et conférences à l'occasion desquels mes travaux ont été présentés, à Paris, Lille, Dijon, Londres, Barcelone, Padoue, Kiel, Steyr et Athènes. Les commentaires et suggestions de l'ensemble des participants, en particulier ceux du Lunch séminaire d'économie appliquée, où j'ai eu la chance de pouvoir présenter l'ensemble de mes travaux, m'ont été particulièrement profitables.

Si le campus Jourdan est un lieu de recherche vivant et convivial, il le doit en grande partie à ses doctorants et anciens doctorants que j'ai eu le plaisir de côtoyer au cours des nombreuses années passés dans ce lieu attachant : Pierre André, Antonin Aviat, Francesco Avvisati, Maria Bas, Pierre-Yves Cabannes, Juan Carluccio, Nicolas Coeurdacier, Stéphane Guibaud, Elvire Guillaud, Élise Huillery, Muriel Jakubowicz, Rémi Jedwab, François Keslair, Ivan Ledezma, François Legrand, Carolina Lennon, Dimitri Mavridis, Florian Mayneris, Amine Ouazad, Gwenn Parent, Abla Safir et Holger Stichnoth. Qu'il me soit permis de décerner une mention spéciale aux résidents passés et présents du bureau B030 ainsi qu'aux membres du blog *Ecopublix*, pour avoir su marier le débat économique avec la bonne humeur, la bonne chère et le bon vin : Laurent Bach, Antoine Bozio, Clément Carbonnier, Guilhem Cassan, Fabien Dell, Gabrielle Fack, Emmanuel Frot, Camille Landais et Mathieu Valdenaire.

J'aimerais encore remercier mes parents et mon frère pour leur soutien attentionné, ainsi que mes amis de toujours qui m'ont chaleureusement entouré tout au long de ces années. Je suis également reconnaissant à mes grands-parents et à ma belle famille de m'avoir encouragé dans mes recherches, de près ou de loin. Ma pensée va enfin à Irène, pour sa douceur, sa patience et la confiance qu'elle a su me témoigner à chaque instant.

Table des matières

Remerciements	i
Introduction générale	1
I OBLIGATION SCOLAIRE ET RENDEMENTS DE L'ÉDUCATION	25
1 Suffit-il d'allonger la scolarité obligatoire pour augmenter les salaires ? Une comparaison France – Royaume-Uni	31
1.1 Introduction	31
1.2 Compulsory schooling laws and educational systems in France and Britain	37
1.2.1 France	37
1.2.2 Britain	40
1.3 Estimation strategy: a Regression Discontinuity Design	41
1.3.1 A “sharp” discontinuity design	41
1.3.2 Reduced form estimation	42
1.3.3 IV estimation of the returns to compulsory schooling	46
1.4 Data and summary statistics	47
1.4.1 France	48
1.4.2 Britain	50
1.5 Results	52
1.5.1 OLS estimates of the returns to schooling	52
1.5.2 Impact of compulsory schooling laws on school leaving age	53
1.5.3 Impact on wages	68
1.5.4 IV estimates of the returns to compulsory schooling	70
1.5.5 Robustness checks	72
1.6 Interpretation	74
1.6.1 Comparison with existing estimates	79
1.6.2 Discussion of potential explanations	80
1.6.3 Institutional features of schooling systems	85
1.6.4 The role of academic credentials	95
1.6.5 The nature of the monetary gains from better academic credentials	105
1.7 Conclusion	110

II SECTORISATION, MARCHÉ IMMOBILIER ET INÉGALITÉS SCOLAIRES 113

2 Politiques de sectorisation, inégalité scolaires et ségrégation résidentielle : analyse théorique et simulations sur données parisiennes 119

2.1 Introduction 119

2.2 Review of literature: theoretical models of school choice 120

2.3 Modelling alternative school enrollment schemes 130

2.3.1 Strict school zoning 130

2.3.2 School choice 138

2.4 Introducing private schools 140

2.5 Simulations 145

2.5.1 Simulations set-up 145

2.5.2 Alternative school admission rules 148

2.5.3 The simulation algorithm 152

2.5.4 An example 159

2.5.5 Changing the values of the parameters 173

2.6 Calibrations for the city of Paris 175

2.6.1 Calibration of the key parameters 175

2.6.2 School enrollment policy simulations 179

2.7 Conclusion 184

3 Sectorisation et marché immobilier : analyse empirique de l'impact de la performance des collèges parisiens sur le prix des logements 187

3.1 Introduction 187

3.2 Estimation strategy 191

3.2.1 The model 191

3.2.2 The endogeneity of school performance 192

3.2.3 School attendance boundaries fixed effects and their limitations 193

3.2.4 An alternative estimator 195

3.3 Public and private middle schooling in Paris 200

3.3.1 The French educational system 200

3.3.2 Public and private school admission rules 202

3.3.3 The importance of middle schooling 205

3.4 Data and summary statistics 206

3.4.1 School catchment areas in Paris 206

3.4.2 Public and private schools 212

3.4.3 Housing prices 219

3.4.4 Neighborhood socio-demographic characteristics 222

3.5 Results 224

3.5.1 Naive estimates 226

3.5.2 Sales located near school attendance boundaries 228

3.5.3 Matching sales across school attendance boundaries 228

3.5.4 How large is the effect? 231

3.6 Robustness checks 231

3.6.1 The boundary fixed effects alternative specification 232

3.6.2 Testing the validity of the identifying assumption 234

3.7	The mitigating effect of private schools	240
3.7.1	Theoretical predictions of school choice models	240
3.7.2	Estimation strategy	244
3.7.3	Results	247
3.7.4	Robustness of findings to alternative interpretations	251
3.8	Conclusion	256
III REDOUBLEMENT, ORIENTATION ET INÉGALITÉS SELON LE MOIS DE NAISSANCE		261
4	L’impact du mois de naissance sur les trajectoires scolaires et la vie professionnelle	267
4.1	Introduction	267
4.1.1	Les effets potentiels du mois de naissance	267
4.1.2	Difficultés méthodologiques	269
4.1.3	Que sait-on sur la part des différents effets du mois de naissance?	273
4.1.4	Études sur données françaises	281
4.1.5	Contribution	284
4.2	Données	287
4.3	Le mois de naissance : un instrument à manier avec précaution	290
4.3.1	Un biais de mesure : la surreprésentation du mois de janvier	291
4.3.2	Une saisonnalité différenciée sociologiquement	293
4.3.3	Peut-on limiter les biais induits par la composante sociologique du mois de naissance?	303
4.4	L’effet du mois de naissance sur les performances scolaires	304
4.4.1	Mois de naissance et durée de la scolarité pré-élémentaire	304
4.4.2	Comment estimer l’impact du mois de naissance sur les performances scolaires?	306
4.4.3	Résultats des estimations	311
4.4.4	Comparaison avec les estimations sur données britanniques	320
4.5	L’effet du mois de naissance sur les trajectoires éducatives	323
4.5.1	Impact sur le redoublement	323
4.5.2	Mois de naissance et orientation scolaire	328
4.5.3	Mois de naissance et obligation scolaire	334
4.6	Les effets de la date de naissance persistent-ils à l’âge adulte?	336
4.6.1	Mois de naissance et niveau de formation	337
4.6.2	Mois de naissance et situation d’emploi	345
4.6.3	Mois de naissance et salaires	345
4.7	Conclusion	349
Conclusion générale		351
Bibliographie		381
Liste des tableaux		388
Liste des figures		395

Introduction générale

« Les distinctions entre primaire, secondaire et supérieur n'ont plus de sens. Séparer, dès l'origine, les Français en deux classes et les y fixer pour toujours par une éducation différente, c'est aller à l'encontre du bon sens, de la justice et de l'intérêt national. Parlons de l'enseignement tout court, de l'enseignement unique. »

L'Université Nouvelle, Tome 1 : Les Principes, 1918.

Lorsqu'au lendemain de la victoire de 1918, les Compagnons de l'Université Nouvelle lancèrent le thème de l'école unique pour donner à tous les Français la même formation de base et élever le niveau général d'instruction, le système éducatif français était encore structuré en deux ordres distincts et largement hermétiques : le primaire, qui fonctionnait comme une « école du peuple », et le secondaire, où étaient scolarisés les enfants issus des couches sociales privilégiées depuis l'école élémentaire jusqu'au lycée. Il aura fallu près d'un demi-siècle de luttes politiques pour que vers le milieu des années 1970, cette hiérarchisation soit définitivement abolie et laisse place à un cursus éducatif unifié de l'école élémentaire jusqu'à la fin du premier cycle de l'enseignement secondaire.

De grandes réformes ont jalonné cette transformation : la réforme Berthoin de 1959, qui créa les Collèges d'enseignement général et les Collèges d'enseignement technique et allongea de deux ans l'obligation scolaire en la portant de 14 à 16 ans révolus ; la réforme Capelle-Fouchet de 1963, qui mit en place les Collèges d'enseignement secondaires, première étape vers l'unification du premier cycle de l'enseignement secondaire ; la réforme Haby de 1975, enfin, qui finalisa le processus en créant le Collège unique, conçu alors comme le seul moyen de donner de mêmes chances à tous

les élèves, de révéler leurs aptitudes et intérêts différents et de les préparer à des orientations ultérieures adaptées.

Le terme ambivalent de « démocratisation scolaire » a été fréquemment utilisé pour qualifier la transformation du système éducatif français au cours de la seconde moitié du XX^e siècle. Ce concept recouvre deux idées distinctes. La démocratisation scolaire renvoie d'abord à une réalité statistique, que l'on désigne parfois sous le nom de « massification » : par ce terme, on décrit le fait que depuis la fin de la Seconde guerre mondiale, l'accès à l'éducation s'est élargi et que le système éducatif français a scolarisé un nombre croissant d'élèves pour des durées d'études de plus en plus longues. La démocratisation scolaire constitue par ailleurs l'expression d'un jugement porté sur les bénéfices de ce processus du point de vue de l'égalité des chances : en s'ouvrant davantage aux catégories les moins favorisées, les transformations du système éducatif français auraient permis de réduire les inégalités face à l'école et d'affaiblir en partie les mécanismes de la reproduction sociale.

Du point de vue quantitatif, le bilan de la démocratisation scolaire est spectaculaire. En l'espace d'un siècle, l'âge moyen de fin d'études est passé de 11,5 à 21,3 ans (Marchand et Thélot, 1997 ; MEN-DEP, 2007). La proportion d'individus quittant l'école sans diplôme est passée de 51,5 % pour les générations nées au début du siècle à 17 % pour les générations nées au début des années 1980 (Thélot et Vallet, 2000 ; MEN-DEP, 2007). Dans le même temps, la proportion d'individus quittant l'école avec un diplôme de l'enseignement secondaire (CAP, BEP ou Baccalauréat) est passée de 9,6 à 41 % et la proportion de diplômés de l'enseignement supérieur de 2,4 à 42 %. Plus que tout autre, le diplôme du Baccalauréat manifeste l'ampleur du processus de démocratisation du système français : la proportion de bacheliers a en effet été multipliée par 10 en moins d'un siècle, en passant de moins de 6 % pour les générations nées dans les années 1900 à près de 64 % pour les générations nées au début des années 1980. L'allongement des études a été perceptible dans tous les milieux sociaux et l'accès au baccalauréat ou aux études supérieures, qui était fermé aux enfants issus des couches populaires, s'est largement élargi : la proportion de bacheliers parmi les enfants d'ouvriers, qui n'était que de 1 % au début du siècle, atteint aujourd'hui près de 50 %.

Du point de vue de la réduction des inégalités, le bilan de la démocratisation scolaire a fait l'objet d'analyses contrastées. D'un côté, certains auteurs considèrent que l'allongement des études et l'augmentation du niveau scolaire dans la population n'a opéré qu'une « translation » des inégalités sans effet réel sur l'égalité des chances. Dans un ouvrage au titre provocateur, Marie Duru-Bellat (2006) défend la thèse de l'« inflation scolaire », selon laquelle la démocratisation scolaire aurait entraîné une dévalorisation des diplômes suffisamment importante pour réduire à zéro les bénéfices de l'allongement des études pour les couches sociales les moins favorisées. Le sociologue Stéphane Beaud (2003) insiste quant à lui sur les frustrations que les promesses déçues d'ascension sociale par les études auraient suscitées chez les jeunes issus de l'immigration et qui ont passé le bac dans les années 1980.

Cette thèse d'une démocratisation scolaire de façade a néanmoins été mise à mal par un certain nombre d'évaluations statistiques récemment réalisées en France. Les estimations réalisées par Thélot et Vallet (2000) à partir des données de l'enquête Formation et Qualification Professionnelles et de l'enquête Emploi semblent indiquer que la démocratisation du système éducatif français a eu un impact appréciable sur la mobilité sociale : d'après ces auteurs, l'allongement général des études expliquerait les trois quarts de la réduction observée du lien entre origine sociale et destinée scolaire pour les générations nées entre 1908 et 1972, et l'affaiblissement du lien spécifique unissant origine sociale et diplôme expliquerait un septième environ de cette réduction des écarts. Plus récemment, les travaux de Gurgand et Maurin (2007) ont révélé que l'expansion scolaire d'après-guerre a coïncidé avec une amélioration très sensible de la situation salariale des individus issus de milieux sociaux modestes : d'après leurs calculs, le surcroît de formation initiale dont ont bénéficié les 90 % de la population qui ne sont passés ni par les grandes écoles ni par les derniers cycles universitaires s'est traduit par un gain salarial total d'environ 7 % en moyenne sur une période de dix-huit ans. En s'appuyant sur les données de l'enquête Emploi, Goux et Nouveau (2007) montrent quant à eux que les générations qui ont bénéficié de la démocratisation du lycée au début des années 1980 sont moins confrontées au chômage entre 27 et 32 ans et ont un taux d'emploi significativement plus élevé entre ces âges que les générations nées plus tôt et qui n'ont pas été exposées à la démocratisation de l'enseignement

secondaire. Une étude de Rosenwald (2006) montre que la plus grande ouverture du système éducatif français a également contribué à réduire les inégalités scolaires entre hommes et femmes. Au cours du XX^e siècle, la scolarité des filles a connu une évolution remarquable et a rattrapé celle des garçons : en 2005, une jeune femme sur deux qui sort du système éducatif est diplômée du supérieur contre quatre garçons sur dix. Les conclusions de ces études rejoignent plus généralement celles des travaux empiriques menés en Angleterre, en Irlande ou dans les pays scandinaves, qui montrent que l’allongement des études et le passage au collège unique ont participé à la réduction des inégalités dans ces pays (Maurin, 2007).

Si le débat sur les bénéfices de la démocratisation scolaire est aussi vif, c’est que la confiance dans la promesse d’émancipation sociale par l’école est aujourd’hui sérieusement entamée. Les inquiétudes et les doutes qui s’expriment sur la capacité de l’institution scolaire à réduire les inégalités se nourrissent d’un triple constat d’échec. Le premier tient au fait que la dynamique d’ouverture du système éducatif français a marqué le pas depuis une dizaine d’années : la durée moyenne des études, la proportion de bacheliers et le taux d’accès à l’enseignement supérieur ne progressent plus depuis le milieu des années 1990. Le second constat tient au fait que l’institution scolaire ne parvient pas à réduire l’échec scolaire d’une fraction considérable de la jeunesse : près de 100 000 élèves quittent chaque année l’école sans diplôme, soit près de 15 % d’une classe d’âge, et ce chiffre ne diminue pas depuis dix ans. Enfin, un certain nombre d’outils mis en place pour tenter de corriger les inégalités scolaires ont déçu leurs promesses : le principe de sectorisation des établissements scolaires, conçu comme un instrument au service de la mixité sociale à l’école, est aujourd’hui perçu comme hypocrite et incapable d’endiguer l’émergence de ghettos scolaires ; la politique d’éducation prioritaire mise en place au début des années 1980 pour cibler les moyens éducatifs sur les établissements les plus difficiles, n’a pas quant à elle apporté les résultats escomptés et semble avoir surtout contribué à stigmatiser les publics visés (Bénabou et al., 2005).

La crise de confiance qui mine aujourd’hui l’institution scolaire suscite un certain nombre d’interrogations sur l’opportunité de poursuivre le mouvement d’ouverture du système éducatif amorcé dans l’immédiat après-guerre. Alors que certains obser-

vateurs n'hésitent plus à prôner le retour à un certain malthusianisme scolaire, afin que l'école s'adapte à l'hétérogénéité des aptitudes individuelles et au caractère limité des débouchés professionnels, d'autres pensent au contraire qu'il est urgent de donner un second souffle à la démocratisation, en élargissant notamment l'accès à l'enseignement supérieur.

Pour que ce débat ne se cantonne pas à des pétitions de principe mais repose sur des chiffres précis et des évaluations solides, il est indispensable de tirer avantage du recul historique dont on dispose aujourd'hui pour dresser un bilan approfondi des bénéfices et des limites de la démocratisation scolaire en France.

Cet exercice se heurte néanmoins à deux difficultés importantes. La première est qu'il est encore trop tôt pour évaluer les effets de certaines des politiques qui ont accompagné ce processus : par exemple, le résultat des efforts éducatifs entrepris dans les années 1980 pour augmenter la proportion de bacheliers et élargir l'accès aux études supérieures est difficile à évaluer de manière complète, dans la mesure où ces efforts ont concerné des générations qui ne sont entrées sur le marché du travail que très récemment. Une seconde source de difficultés tient au fait que les politiques éducatives qui ont jalonné la démocratisation du système scolaire français ne représentent pas un bloc homogène, mais se déclinent en une multiplicité d'interventions publiques, qui vont de l'obligation scolaire à la politique d'éducation prioritaire, en passant par la mise en place du collège unique. Compte tenu de la diversité des enjeux associés à ces différentes politiques, on ne peut espérer mettre à jour les raisons de leur aptitude ou de leur incapacité à avoir réduit les inégalités scolaires et socio-économiques sans s'atteler à une évaluation spécifique de leurs effets.

Dans cette perspective, l'objectif de cette thèse est double. Il s'agit d'une part de mobiliser les outils de l'analyse économique pour contribuer à l'évaluation du bilan de la démocratisation scolaire en France, en concentrant notre analyse sur trois des principales politiques éducatives qui ont accompagné l'ouverture du système éducatif à une fraction plus importante de la population : l'allongement de la scolarité obligatoire, la sectorisation des établissements d'enseignement public et l'orientation scolaire des élèves. L'impact de chacune de ces politiques sur les inégalités met en jeu des mécanismes distincts, qu'il convient d'analyser séparément. Surtout, leurs effets

ne peuvent être pleinement mesurés qu'à la condition d'adopter une perspective de long terme. C'est pourquoi nous avons choisi autant que possible de mobiliser des sources statistiques qui permettent de suivre les trajectoires scolaires et professionnelles des individus sur une période relativement longue. Pour autant, l'ambition de cette thèse ne se résume pas à la prise en compte de la dimension historique des politiques éducatives étudiées. En effet, la scolarité obligatoire, la sectorisation des établissements et l'orientation scolaire demeurent à bien des égards des piliers importants du système éducatif français tel qu'il fonctionne aujourd'hui et elles font régulièrement l'objet de débats passionnés. L'évaluation de l'impact de ces politiques constitue par conséquent un élément essentiel de la réflexion qui doit guider l'intervention publique dans la sphère éducative. Dans cette perspective, le second objectif de cette thèse consiste à tenter d'identifier et de quantifier les principaux paramètres qui conditionnent le succès ou l'échec des politiques scolaires évaluées, afin de proposer un certain nombre de pistes de réformes susceptibles d'améliorer l'efficacité et l'équité de notre système éducatif.

Dans la suite de cette introduction, nous discutons l'intérêt d'une évaluation économique des politiques éducatives, avant de présenter l'approche méthodologique que nous avons adoptée et les principales contributions de cette thèse.

L'intérêt d'une approche économique des politiques éducatives

Contrairement à d'autres objets d'étude, tels que la politique monétaire ou le commerce international, l'éducation ne constitue pas un domaine réservé de l'économie. On peut même considérer que l'intérêt des économistes pour les questions éducatives ne s'est manifesté qu'assez tardivement, alors que les historiens et les sociologues ont depuis longtemps investi ce champ de recherche. Les politiques éducatives que nous analysons dans cette thèse ont d'ailleurs plus souvent été abordées du point de vue de l'histoire, de la sociologie ou des sciences de l'éducation que du point de vue économique à proprement parler. L'extension de la scolarité obligatoire en France a été étudiée par les historiens dans le contexte plus large du processus d'unification

du système éducatif après la Seconde guerre mondiale, avec une attention particulière portée aux débats politiques et aux résistances qui accompagnèrent ce puissant mouvement de transformation de l'école (Prost, 2004). La politique de sectorisation des établissements scolaires publics a quant à elle donnée lieu à de très nombreuses études sociologiques qui ont examiné le lien entre ségrégation résidentielle et ségrégation sociale à l'école et cherché à identifier les déterminants sociaux de l'évitement scolaire (Van Zanten, 2001 ; Oberti et Préteceille, 2004 ; François et Poupeau, 2004). L'orientation scolaire, enfin, a été étudiée par les sciences de l'éducation avec pour objectif de mettre en lumière le rôle joué par les représentations des élèves sur leurs propres aptitudes, la pression exercée par les parents et les pairs ou encore l'influence du corps enseignant (Le Bastard-Landrier, 2005 ; Odry, 2006).

L'intérêt relativement tardif des économistes pour l'éducation s'explique en grande partie par le fait que la dimension économique de cet objet d'étude n'apparaît pas *a priori* essentielle, dans la mesure où l'école est généralement perçue comme étant d'abord un lieu de culture destiné à former des citoyens responsables et capables de s'intégrer dans la vie sociale, ce qui dépasse de beaucoup les aspects économiques de l'insertion professionnelle. Un certain nombre d'économistes ont néanmoins contribué à montrer que contrairement aux apparences, la composante économique de l'éducation était loin d'être négligeable et que la discipline économique aurait tort de se désintéresser d'un secteur qui est devenu le premier budget de l'État dans la plupart des pays développés¹. Sous l'impulsion décisive de Gary Becker et Jacob Mincer, l'économie de l'éducation s'est constituée au début des années 1960 comme un domaine de recherche autonome au sein de la discipline économique². Dans le cadre de la théorie du capital humain, ces auteurs ont montré que les choix éducatifs individuels pouvaient être analysés comme une forme d'investissement qui s'exerce en prenant en compte les coûts et les bénéfices qui sont associés à la poursuite des études. Dans le prolongement de cette approche, de très nombreuses analyses économiques ont été consacrées à l'estimation empirique de la valeur des rendements privés de l'éduca-

¹En France, le budget de l'Éducation nationale s'élève en 2007 à 77 milliards d'euros, soit près de 28 % du budget de l'État. En 2006, la dépense intérieure d'éducation a atteint 121,4 milliards d'euros, soit environ 6,8 % de la richesse nationale (MEN-DEP, 2007).

²Pour une présentation synthétique des recherches en économie de l'éducation, voir Gurgand (2005).

tion, en tentant d'identifier l'effet causal du nombre d'années d'études sur le salaire indépendamment des nombreux biais de sélection qui affectent la relation entre ces deux variables. La rentabilité de l'investissement éducatif a également été envisagée du point de vue macroéconomique dans le cadre de la réflexion sur le lien entre éducation et croissance.

Si l'économie de l'éducation s'est longtemps structurée autour de la seule théorie du capital humain, elle s'en est progressivement émancipée sous l'effet d'une double révolution intervenue au milieu des années 1980. D'une part, la mise à disposition des chercheurs de données statistiques individuelles d'une très grande richesse a entraîné une multiplication des études empiriques consacrées aux questions éducatives. D'autre part, ce bouleversement s'est accompagné d'une révolution méthodologique sans précédent, à travers l'élaboration de techniques économétriques d'évaluation toujours plus sophistiquées. L'économie de l'éducation a joué un rôle moteur dans ce renouveau empirique qui n'a épargné aucune branche de la science économique, dans la mesure où les défis empiriques rencontrés par les économistes de l'éducation sont à l'origine d'une partie importante des avancées méthodologiques réalisées en économétrie au cours des vingt dernières années. Le regain d'intérêt pour l'analyse empirique a surtout contribué à un profond renouvellement des problématiques de l'économie de l'éducation. Celle-ci s'est ouverte à une pléiade de nouveaux sujets, qui vont de l'étude de l'efficacité et de l'équité de la dépense d'éducation (Fernández et Rogerson, 2003 ; Gary-Bobo et Trannoy, 2008) à l'estimation de l'effet de la taille des classes sur la réussite des élèves (Angrist et Lavy, 1999 ; Piketty et Valdenaire, 2006 ; Gary-Bobo et Mahjoub, 2006 ; Bressoux et al., 2008), en passant par l'analyse des relations entre éducation et ségrégation résidentielle (Black, 1999 ; Gibbons et Machin, 2006 ; Bénabou et al., 2005) ou encore la mesure des effets de pairs à l'école (Hoxby, 2001 ; Angrist et Lang, 2004 ; Kramarz et al., 2007). L'élargissement du champ d'investigation des économistes de l'éducation les a conduits à aborder des problématiques qui recoupent d'autres champs de la discipline économique : une grande partie des recherches entreprises aujourd'hui en économie de l'éducation se situent à la frontière avec l'économie urbaine, l'économie de la santé ou encore l'économie de la criminalité.

Les thèmes que nous abordons dans cette thèse se rattachent tout autant aux

problématiques traditionnelles de l'économie de l'éducation qu'à l'exploration des nouveaux horizons ouverts par ce champ de recherche. Notre évaluation de l'impact de l'allongement de l'obligation scolaire en France s'inscrit dans le prolongement de la littérature consacrée à l'estimation des rendements de l'éducation et adopte une perspective centrée sur les relations privilégiées qui unissent l'école au monde du travail. L'analyse théorique et empirique que nous proposons des politiques de sectorisation se situe quant à elle à la frontière de l'économie de l'éducation et de l'économie urbaine, dans la mesure où elle s'intéresse à la manière dont ces politiques influencent les mécanismes de stratification résidentielle qui caractérisent le marché immobilier. Enfin, notre analyse des politiques d'orientation se veut une contribution à l'analyse des relations entre éducation et inégalités, qui sont abordées ici du point de vue d'un déterminant aussi surprenant que sous-estimé des trajectoires scolaires : le mois de naissance.

Compte tenu de la richesse des études historiques et sociologiques qui ont été consacrées aux politiques éducatives étudiées dans cette thèse, on peut légitimement s'interroger sur l'intérêt de l'approche économique que nous avons retenue. À nos yeux, cet intérêt est double.

Du point de vue théorique, l'analyse économique des politiques éducatives nous paraît présenter un avantage spécifique par rapport aux approches privilégiées par les autres disciplines. La modélisation des comportements individuels permet en effet de clarifier et de rendre plus rigoureuse la compréhension de l'impact de ces politiques. Ainsi, il n'est pas rare que les modèles utilisés par les économistes révèlent que plusieurs configurations sont possibles en fonction de la valeur prise par les paramètres qui déterminent l'équilibre. La modélisation des interactions entre choix scolaire et marché immobilier montre par exemple que l'effet des politiques de sectorisation sur les inégalités scolaires dépend tout autant de l'importance accordée par les parents à la qualité de l'établissement scolaire fréquenté par leur enfant, que de la corrélation entre performances scolaires et revenu parental ou encore de l'ampleur des effets de pairs à l'école (phénomène qui désigne le fait que la réussite scolaire d'un élève est influencée par celle de ses camarades). L'analyse théorique des politiques éducatives prépare ainsi le travail empirique, dans la mesure où elle indique au chercheur les

paramètres qui doivent être estimés en priorité. Une fois quantifiés, ces paramètres permettent de calibrer les modèles pour en faire des outils de simulations utiles à l'élaboration des politiques publiques. Dans le domaine des politiques éducatives, l'intérêt de ce type d'approche tient au fait que les réactions comportementales des agents sont susceptibles de modifier radicalement l'impact des réformes entreprises, ce qui rend nécessaire la prise en compte des effets d'équilibre général qu'elles pourraient engendrer.

Du point de vue empirique, l'analyse économique des politiques éducatives présente un triple avantage. Son premier atout tient à l'existence de méthodes empiriques qui permettent d'établir des relations de causalité de manière convaincante. L'économiste de l'éducation ne se contentera pas de constater l'existence d'une forte association entre salaire individuel et niveau d'éducation, ou entre prix immobiliers et performances scolaires, pour en inférer une quelconque relation causale. Son travail empirique consistera plutôt à établir ces relations en exploitant les variations exogènes induites par des « expériences naturelles ». La seconde force de l'approche économique tient à sa capacité à quantifier précisément les effets en jeu, sans se limiter aux aspects qualitatifs. L'économiste de l'éducation prendra toujours un soin particulier à multiplier les mesures d'un paramètre, à établir d'éventuelles variations des effets en fonction des sous-groupes de population considérés, afin de vérifier la robustesse de ses estimations. La force de conviction des interprétations proposées tient en grande partie à ce travail minutieux de validation des résultats. Enfin, le troisième avantage que nous paraît présenter l'approche économique tient à sa capacité à guider l'action publique à travers l'application d'un critère certes imparfait, mais néanmoins utile, pour mesurer l'efficacité des politiques éducatives : l'analyse coût-bénéfice. Cette méthode, qui consiste à mettre en balance l'ensemble des coûts et l'ensemble des bénéfices engendrés par une politique publique, vise à déterminer *ex ante* ou *ex post* si cette politique est rentable et si elle constitue la meilleure solution dans l'éventail des choix possibles.

Malgré ses mérites indéniables, l'analyse économique des politiques éducatives n'a pas pour autant vocation à supplanter les autres approches disciplinaires dans ce domaine de recherche. Bien qu'elle ait considérablement élargi son champ d'investiga-

tion au cours de la dernière décennie, l'économie de l'éducation demeure néanmoins structurée autour de la problématique de l'allocation efficace des ressources, qui ne constitue qu'une manière parmi d'autres d'aborder la question scolaire. Des aspects aussi diverses que l'histoire des transformations du système éducatif, la sociologie des enseignants ou encore la définition du contenu des programmes scolaires dépassent largement le domaine de compétence de l'économiste. Si le travail réalisé dans le cadre de cette thèse ne rend sans doute pas suffisamment justice à la contribution des autres disciplines à la compréhension des questions éducatives, il s'est néanmoins profondément enrichi à leur contact, ce qui a permis à son auteur d'éviter bien des pièges et des errements.

Comment évaluer les politiques éducatives ?

Bien que cette thèse comporte un certain nombre d'analyses théoriques, la plupart de ses contributions sont d'ordre empirique. De ce point de vue, notre travail a bénéficié de trois apports distincts : l'existence de méthodes économétriques particulièrement bien adaptées à l'évaluation de l'impact des politiques éducatives, l'existence de sources d'informations riches et variées et, enfin, la possibilité de tirer parti d'une approche comparative pour faciliter l'interprétation des résultats.

La délicate identification des relations de causalité

La mise en évidence de relations de causalité constitue le principal défi empirique rencontré dans l'évaluation des politiques éducatives. Les outils économétriques que nous avons utilisés pour répondre à ce défi sont issus des développements les plus récents de la littérature consacrée aux méthodes d'évaluation des politiques publiques. Outre le fait qu'elles donnent la possibilité de quantifier l'impact causal des réformes envisagées, ces méthodes ont l'avantage d'être à la fois transparentes et de se prêter aisément à des analyses graphiques.

Quiconque souhaite mesurer l'effet d'une réforme sur les individus qui y ont été assignés est confronté à un problème de données manquantes : à tout moment dans le temps, un individu soit participe au programme considéré, soit n'y participe pas.

L'enjeu essentiel des méthodes d'évaluation consiste donc à construire un contrefactuel pour les individus participant au programme. En l'absence de données expérimentales, c'est-à-dire de données collectées spécifiquement pour constituer un groupe de contrôle composé de personnes qui auraient participé au programme mais qui en ont été exclues de manière aléatoire, l'évaluation de l'impact des politiques publiques nécessite l'utilisation de méthodes visant à la construction indirecte d'un groupe de contrôle.

La première méthode que nous avons utilisée pour résoudre ce problème dans le cadre de l'évaluation de l'impact de l'allongement de la scolarité obligatoire sur l'emploi et les salaires consiste à utiliser la discontinuité induite par l'augmentation de l'âge de fin d'études minimum d'une génération à une autre. L'estimation par discontinuité dans la régression (*Regression discontinuity design*) permet en effet, sous la double hypothèse que le traitement (ici, l'allongement de l'obligation scolaire) est une fonction discontinue d'une variable latente continue (ici, la date de naissance) et que les caractéristiques des agents évoluent de manière continue avec cette variable latente, d'estimer l'impact du traitement en comparant les individus qui se trouvent juste en-dessous et juste au-dessus de la discontinuité. En l'occurrence, cette méthode nous donne les moyens de mesurer le rendement salarial d'une année de scolarité obligatoire supplémentaire.

Les techniques d'appariement (*matching*) constituent la seconde méthodologie employée dans ce travail. Cette technique s'avère particulièrement utile lorsque les unités économiques traitées et non traitées ne sont pas directement comparables, comme dans le cas des transactions immobilières qui diffèrent par un grand nombre d'attributs observables (nombre de pièces, étage, etc.) et inobservables (qualité du quartier, présence de transports en commun, etc.). L'appariement de ces unités sur la base d'un critère de proximité (la distance géographique, par exemple) permet sous certaines conditions d'aboutir à l'identification causale de l'effet du traitement. Dans le cadre de l'étude que nous consacrons à l'impact de la performance des établissements scolaires sur les prix immobiliers, nous utilisons cette approche pour comparer le prix de vente des logements tout en tirant parti des discontinuités spatiales créées par les frontières entre secteurs scolaires.

La troisième et dernière approche méthodologique retenue dans cette thèse est plus classique et repose sur l'utilisation de variables instrumentales. Cette méthode vise à résoudre le problème d'endogénéité potentielle d'une variable explicative en exploitant le fait qu'elle est elle-même influencée par une autre variable qui n'exerce pas d'effet autonome sur la variable dépendante. Dans le cadre de l'étude que nous consacrons aux inégalités scolaires induites par la date de naissance des individus, nous utilisons le fait que le mois de naissance des élèves influence leur âge relatif par rapport à leurs camarades de classe sans être *a priori* corrélé avec d'autres déterminants de la réussite éducative, pour estimer l'impact des différences d'âges sur les écarts de performances scolaires.

La validité de ces différentes méthodes repose sur des hypothèses fortes qui ne peuvent pas toujours être testées directement. Par ailleurs, ces techniques ne permettent de réaliser que des estimations moyennes et locales qui sont susceptibles d'être contaminées par des effets d'équilibre général. C'est la raison pour laquelle une grande partie de nos analyses est consacrée à la justification des choix méthodologiques effectués et à la discussion de la crédibilité des résultats obtenus. De ce point de vue, la vérification de la solidité de nos résultats et de la pertinence de nos interprétations a été grandement facilitée par l'abondance des sources utilisables en France pour étudier l'impact des politiques éducatives.

Des sources d'information nombreuses mais inégalement accessibles

À la différence d'autres thèmes de recherche en économie publique, tels que le logement ou les retraites, les sources statistiques mobilisables par les économistes de l'éducation sont à la fois riches et variées. Elles ne sont pas malheureusement toutes aisément accessibles aux chercheurs.

En France, les données les utilisées pour étudier les questions éducatives d'un point de vue empirique se sont longtemps limitées aux enquêtes réalisées par l'Insee. Les informations contenues dans ces enquêtes portent essentiellement sur la population des individus âgés de plus de 15 ans, ce qui en fait des outils d'analyse privilégiés lorsqu'on s'intéresse à la manière dont l'éducation influence le destin professionnel

des individus. Parmi ces sources, l'enquête Emploi, qui est réalisée tous les ans sur un échantillon d'environ 60 000 ménages, constitue une source particulièrement utile, dans la mesure où elle permet de croiser des informations sur le niveau de formation des individus enquêtés avec leur situation d'emploi et, depuis 1982, avec leur salaire. Les fichiers du recensement de la population présentent, pour leur part, l'avantage de contenir plusieurs millions d'observations, ce qui permet d'obtenir des estimations d'une grande précision, même si l'intérêt de ces sources est limité par le fait qu'elles ne contiennent qu'un nombre restreint de variables sur le niveau d'éducation des individus et ne sont disponibles que tous les dix ans environ.

Le principal inconvénient des données d'enquête que nous venons d'évoquer est qu'elles ne portent pas spécifiquement sur la population des élèves en cours de scolarité. Or, certaines politiques éducatives ne peuvent être évaluées qu'à la condition de connaître les caractéristiques socio-démographiques des élèves, leurs performances à l'école et leurs trajectoires scolaires. De ce point de vue, les enquêtes réalisées par la Direction de l'évaluation, de la prospective et de la performance³ (DEPP) pour le compte du Ministère de l'Éducation nationale constituent des sources d'information précieuses sur les performances des élèves et leur environnement scolaire. Dans le cadre de cette thèse, nous avons en particulier utilisé les données du Panel primaire de l'éducation nationale, qui permet de suivre la scolarité primaire complète d'un échantillon national représentatif de plusieurs milliers d'élèves entrés en CP en 1997. La scolarité secondaire d'élèves entrés en sixième en 1995 peut quant à elle être suivie à partir des données du Panel secondaire de l'éducation nationale.

Malgré la grande richesse des informations qu'elles contiennent, les enquêtes produites par la DEPP ne portent que sur des échantillons d'élèves relativement petits. C'est pourquoi nous avons, dans la mesure du possible, cherché à utiliser les renseignements contenus dans les fichiers administratifs qui sont gérés conjointement par les services statistiques des rectorats académiques et par les services centraux du ministère de l'éducation nationale. Nous avons ainsi pu travailler à partir des données individuelles de la base Scolarité, qui recense au niveau individuel l'ensemble des

³Anciennement Direction de la programmation et du développement (DPD), devenue Direction de l'évaluation et de la prospective (DEP) en 2003, puis Direction de l'évaluation, de la prospective et de la performance (DEPP) en 2006.

élèves scolarisés dans les établissements secondaires français, qu'ils appartiennent au secteur public (depuis 1993) ou privé (depuis 2000). Nous avons également pu réaliser un certain nombre d'estimations à partir de la base du Diplôme National du Brevet, qui permet d'observer l'ensemble des élèves qui ont passé les épreuves de cet examen en 2004. Si la nature administrative de ces fichiers se traduit par une information relativement limitée sur les caractéristiques socio-démographiques des élèves, leur caractère exhaustif en fait des instruments incomparables tant pour l'étude des trajectoires scolaires à l'échelle globale que pour l'analyse du fonctionnement du système éducatif au niveau local. L'accès à ce type de données pour des motifs de recherche est malheureusement soumis à de fortes restrictions et une grande partie des fichiers académiques d'élèves demeure inaccessible aux chercheurs⁴.

Un certain nombre de données permettent enfin d'élargir l'analyse des politiques éducatives à des problématiques qui se situent à la frontière de l'économie de l'éducation et de l'économie urbaine. Dans cette perspective, nous avons pu étudier l'interaction entre politiques de sectorisation et marché immobilier à Paris en croisant le fichier des transactions immobilières de la Chambre des Notaires de Paris et d'Île-de-France avec un certain nombre d'informations géolocalisées. Celles-ci incluent notamment le tracé de la carte scolaire depuis le milieu des années 1990, des indicateurs de performance des collèges parisiens et des données socio-démographiques issues du recensement de la population.

L'intérêt de la dimension comparative

L'économie de l'éducation est un domaine de recherche où la perspective comparative est susceptible d'apporter un éclairage utile pour aider à l'interprétation des résultats obtenus.

Les politiques éducatives étudiées dans le cadre de cette thèse sont loin, en effet, d'avoir constitué l'apanage de la France. Malgré des différences dans les modalités d'application, des réformes telles que l'allongement de la scolarité obligatoire ou la

⁴Malgré nos demandes répétées et le soutien du Haut Commissariat aux solidarités actives contre la pauvreté, l'accès aux fichiers des élèves de l'Académie de Paris nous a par exemple été systématiquement refusé par le Rectorat, justifiant le dépôt d'un recours auprès de la Commission d'accès aux documents administratifs (CADA).

mise en place d'une politique de sectorisation des établissements scolaires ont été adoptées dans la plupart des nations développées. Dans un certain nombre de pays, ces réformes ont été évaluées, offrant de précieux points de comparaison avec le cas français. Pourvu qu'on prenne les précautions qui s'imposent dans ce type d'exercice, la confrontation des politiques éducatives françaises avec les expériences étrangères peut permettre de mieux comprendre les facteurs institutionnels qui conditionnent l'efficacité de ces politiques.

Dans cette perspective, nous avons accordé une attention particulière aux réformes éducatives menées au Royaume-Uni, dans la mesure où l'unification du système éducatif et la réorganisation des cursus scolaires que ce pays a mis en place à partir du milieu des années 1960 présentent de nombreuses similitudes avec les transformations intervenues en France à la même époque. Cette comparaison s'appuie sur les résultats de l'abondante littérature économique consacrée aux réformes éducatives au Royaume-Uni, mais également sur un certain nombre de recherches originales menées dans le cadre de cette thèse.

Problématique et plan de thèse

Le plan de cette thèse s'articule autour des trois politiques éducatives que nous avons choisi d'examiner et qui constituent trois piliers de la démocratisation de l'école en France : l'allongement de l'obligation scolaire, la sectorisation des établissements scolaires et l'orientation des élèves. Quel a été l'impact de ces politiques sur les inégalités scolaires, entendues au sens large : inégalités économiques, inégalités socio-spatiales et inégalités liées à la date de naissance des individus ?

Partie I : Obligation scolaire et rendements de l'éducation

La première partie de cette thèse étudie les conséquences de la politique éducative qui amorça la modernisation et la démocratisation du système éducatif français : l'allongement de la scolarité obligatoire par la loi Berthoin qui fut votée en 1959 et porta l'âge de fin d'études minimum de 14 à 16 à partir de la rentrée 1967.

Dans le chapitre 1, nous évaluons cette réforme en la comparant à l'*Education*

Act de 1973 qui porta au Royaume-Uni l'âge de fin d'études minimum de 15 à 16 ans révolus. Ces deux réformes présentent un grand nombre de points communs : elles furent appliquées à peu près à la même époque et s'inscrivaient dans un processus d'unification du système éducatif. Nous exploitons les discontinuités législatives engendrées par ces deux réformes pour examiner l'impact d'une scolarité obligatoire plus longue sur le destin scolaire et professionnel des individus concernés. Les données utilisées pour effectuer nos estimations sont issues de l'enquête Emploi française et des *Quarterly Labour Force Surveys* britanniques.

Nous montrons que la réforme Berthoin en France et l'*Education Act* au Royaume-Uni ont eu sensiblement le même impact sur l'âge de fin d'études moyen : l'allongement de la scolarité obligatoire à 16 ans a augmenté la durée d'études moyenne de 0,3 année environ, cet effet étant presque exclusivement concentré sur le tiers inférieur de la distribution des âges de fin d'études. En revanche, nos estimations indiquent que ces deux réformes ont eu un impact salarial très différent : alors qu'au Royaume-Uni, l'allongement de l'obligation scolaire a augmenté les salaires masculins et féminins de 4 à 8 % par année d'études supplémentaire, la réforme Berthoin n'a eu aucun impact significatif sur les rémunérations des individus qui sont restés une ou deux années supplémentaires à l'école.

Nos recherches nous conduisent à écarter un certain nombre d'explications possibles à ce paradoxe apparent. Nous montrons en particulier que ni l'hétérogénéité des rendements de l'éducation, ni le rôle du salaire minimum, ni une éventuelle impréparation dans l'application de la réforme Berthoin ne sont susceptibles de rendre compte de manière convaincante de l'impact salarial très différent constaté en France et au Royaume-Uni à la suite de l'extension de l'obligation scolaire.

Nous privilégions une interprétation fondée sur le rôle du diplôme dans la détermination des rendements de la scolarité obligatoire. Nous montrons en effet que la réforme Berthoin et l'*Education Act* n'ont pas affecté de la même manière le niveau de qualification des élèves français et britanniques. En France, l'allongement de la scolarité obligatoire n'a nullement contribué à diminuer la proportion d'individus quittant l'école sans diplôme. Au contraire, le relèvement de l'âge de fin d'études minimum au Royaume-Uni a entraîné une réduction très sensible du nombre d'élèves sans qualifi-

cation, en particulier dans la population féminine, et a accru le nombre de titulaires du GCSE, équivalent de notre BEPC. Les estimations que nous avons réalisées à partir des données d'enquêtes internationales sur la littératie des adultes semblent également indiquer que l'augmentation du niveau de qualification de la population britannique s'est accompagnée d'une amélioration qualitative des compétences cognitives des cohortes touchées par la réforme de 1973. Davantage qu'à un pur effet de signal, l'impact positif de l'allongement de l'âge minimum de fin d'études sur les salaires au Royaume-Uni semble donc pouvoir être interprété comme la conséquence des efforts accrus consacrés par les élèves britanniques dans le cadre de la préparation des diplômes qui sanctionnent la fin de la scolarité obligatoire.

Malgré la grande similitude des cursus scolaires dans les deux pays, l'impact différent de la réforme Berthoin et de l'*Education Act* sur la structure des qualifications est la conséquence directe d'un certain nombre de spécificités institutionnelles propres au système éducatif français de la fin des années 1960 : l'existence du redoublement, l'orientation précoce des élèves en filière professionnelle et le statut mal défini du diplôme du BEPC sont trois facteurs qui contribuent à expliquer que la réforme Berthoin ne soit pas parvenue à améliorer significativement le niveau de qualification des individus ayant un faible niveau d'études.

Si les résultats de cette étude indiquent que l'allongement de la scolarité obligatoire en France ne s'est pas traduit à court terme par une réduction des inégalités scolaires, la comparaison avec le cas britannique montre qu'il n'y avait pas là de fatalité. Nos recherches suggèrent en effet que dans les pays dont les systèmes éducatifs sont structurés autour de l'obtention de diplômes, l'allongement de la scolarité obligatoire n'est susceptible d'améliorer les perspectives professionnelles des individus qui entrent précocement sur le marché du travail que si elle leur donne la possibilité de quitter l'école avec une qualification sanctionnée par un diplôme. De ce point de vue, il existe en France deux obstacles majeurs à la réduction du nombre considérable d'élèves qui abandonnent leurs études sans aucune qualification : la pratique massive du redoublement, qui conduit de nombreux élèves à quitter l'école avant d'avoir pu passer le moindre diplôme et le fait que l'obtention d'un diplôme professionnel exige la poursuite d'études au-delà de l'obligation scolaire.

Partie II : Politiques de sectorisation, marché immobilier et inégalités scolaires

La croissance très rapide des effectifs scolarisés dans le premier cycle de l'enseignement secondaire et son accélération après l'entrée en vigueur de la scolarité obligatoire à 16 ans, ont nécessité la mise en place d'un mécanisme d'allocation des élèves aux différents établissements scolaires : le principe de sectorisation, qui fut instauré en 1963, stipule qu'un élève est scolarisé dans l'établissement scolaire public qui correspond à son lieu de résidence. À l'origine, la sectorisation n'avait pour autre vocation que celle d'être un instrument prévisionnel de gestion des ressources éducatives : elle permettait de planifier les ouvertures et fermetures de classes en fonction de l'évolution de la démographie locale. Ce n'est que progressivement que ce qui n'était qu'un moyen d'affectation des élèves parmi d'autres s'est transformé en un instrument de promotion de la mixité sociale. La seconde partie de cette thèse est consacrée à l'analyse des conséquences de ce mode d'allocation d'élèves, à travers la prise en compte des interactions entre choix scolaires et choix résidentiels⁵.

Près de 40 ans après sa mise en place, force est de constater que la sectorisation n'est pas parvenue à homogénéiser le recrutement social des établissements scolaires, au point que certains y voient le symbole suprême de l'« hypocrisie » du système éducatif français (Dubet et Duru-Bellat, 2000). Le problème se situe pour l'essentiel dans les grandes agglomérations urbaines, au niveau du collège, et s'explique par deux séries de facteurs : l'existence d'une forte ségrégation résidentielle, d'une part, qui implique que les zones de recrutement des établissements scolaires se superposent à un tissu urbain fortement différencié socialement ; l'existence d'un secteur privé largement subventionné, d'autre part, qui permet d'échapper à moindre coût aux contraintes de la carte scolaire.

Le chapitre 2 propose une analyse théorique du lien unissant politiques de sectorisation, stratification résidentielle et inégalités scolaires. L'objectif de cette contribution est d'étudier les propriétés de modalités alternatives d'allocation des élèves aux établissements scolaires (sectorisation stricte, redécoupage de la carte scolaire ou libre choix de son école) afin d'éclairer les enjeux du débat qui s'est ouvert en France

⁵Cette partie a été co-écrite avec Gabrielle Fack.

au sujet de la réforme de la sectorisation. Pour prendre en compte les spécificités institutionnelles du système éducatif français, nous développons un modèle spatial de choix résidentiel et scolaire qui s'inspire de modèles existants (Bénabou 1993 ; Nechyba, 1997 ; Epple et Romano, 2003 ; Rothstein, 2006) mais où coexistent un secteur public financé au niveau national et un secteur privé subventionné. Parmi les ingrédients essentiels de ce modèle figurent la répartition spatiale non aléatoire de la qualité intrinsèque des quartiers résidentiels, la corrélation positive entre revenu des parents et niveau scolaire des enfants et l'existence d'effets de pairs à l'école. Ce cadre théorique permet une identification précise des mécanismes par lesquels les choix résidentiels et scolaires sont affectés par le mode de recrutement des établissements scolaires et modifient de manière endogène la composition sociale des écoles et les performances scolaires des élèves. Cet exercice permet également de quantifier les gains et les pertes subies par les élèves dans les différents systèmes, en fonction de leur niveau scolaire et du revenu parental. Bien que les solutions de ce modèle ne soient pas calculables analytiquement, elles peuvent être étudiées au moyen de simulations numériques que nous calibrons sur les données parisiennes en utilisant les estimations empiriques réalisées dans le chapitre suivant. Nos simulations suggèrent qu'en l'absence d'établissements privés, le redécoupage des secteurs scolaires est beaucoup plus profitable aux élèves issus de milieux défavorisés que la suppression de la carte scolaire, qui entraîne dans notre modèle une stratification totale des écoles selon le niveau des élèves. Cependant, l'impact de cette politique sur les choix résidentiels réduit de façon sensible le niveau de mixité sociale initialement visé. Nous montrons également que la présence du secteur privé modifie profondément la capacité d'une telle politique à réduire les inégalités scolaires. En effet, tout redécoupage de la carte scolaire entraîne une fuite des élèves issus des milieux aisés vers le privé, ce qui conduit à une diminution du niveau des écoles publiques et annule les effets bénéfiques de cette politique pour les enfants les plus défavorisés.

Le chapitre 3 est consacré à l'estimation de l'impact du niveau des collèges sur le prix des logements à Paris. Plus précisément, nous estimons la disponibilité à payer des ménages pour un meilleur établissement, qui est l'un des paramètres nécessaires à la calibration du modèle développé dans le chapitre précédent. Pour y parvenir, nous

croisons les informations issues de sources diverses, qui incluent le fichier exhaustif des transactions immobilières réalisées à Paris au cours de la période 1997-2003, le découpage des secteurs scolaires en vigueur chaque année depuis le milieu des années 1990, des indicateurs de performances des collèges publics et privés et les caractéristiques socio-démographiques des quartiers résidentiels parisiens. L'estimation de l'impact de la performance des collèges publics sur les prix immobiliers est compliquée par le fait que la qualité des établissements est une variable endogène, les bons collèges étant en général situés dans les quartiers les plus cotés. Pour résoudre ce problème, nous proposons une version améliorée de la stratégie proposée initialement par Black (1999), qui consiste à comparer les prix des transactions immobilières localisées de part et d'autre des frontières entre secteurs scolaires. Nos résultats indiquent qu'une augmentation d'un écart-type du niveau d'un collège entraîne une augmentation des prix immobiliers comprise entre 1,5 et 2,5 % selon l'indicateur de performance retenu. Les différences de niveaux entre collèges expliquent environ 5 % des différences de prix constatées entre les secteurs scolaires à Paris. Ces résultats sont du même ordre de grandeur que les estimations réalisées sur données américaines et anglaises pour les écoles primaires (Black, 1999 ; Gibbons et Machin, 2003 ; Figlio et Lucas, 2004). Dans un second temps, nous exploitons la forte présence du secteur privé sous contrat à Paris, qui scolarise près d'un tiers des collégiens, pour analyser la manière dont cette offre éducative alternative affecte la disponibilité des ménages à payer pour la qualité des établissements publics. Nous montrons que conformément aux prédictions théoriques des modèles de choix scolaire, une plus forte densité de collèges privés dans un quartier résidentiel réduit localement la capitalisation de la performances des collèges publics dans les prix immobiliers.

Partie III : Redoublement, orientation et inégalités selon le mois de naissance

L'unification du système éducatif français, initiée à la fin des années 1950, s'est rapidement heurtée à la difficulté de scolariser tous les élèves au sein d'un cursus scolaire indifférencié du primaire à la fin du lycée. C'est pourquoi les réformes qui ont succédé à la réforme Berthoin de 1959 ont instauré un cycle d'observation de deux,

puis quatre ans au collège, qui débouche sur des orientations différenciées. En France, l'orientation scolaire peut difficilement être dissociée de la pratique redoublement, dans la mesure où le retard scolaire accumulé par les élèves constitue l'un des critères les plus communément utilisés par les équipes pédagogiques pour déterminer si un élève doit poursuivre ses études dans la voie générale ou dans la voie professionnelle. Dans la troisième partie de cette thèse, nous étudions les conséquences de l'une et l'autre de ces pratiques éducatives sur le destin scolaire et professionnel des individus, à travers le prisme d'un facteur d'inégalité aussi surprenant que sous-estimé : la date de naissance.

Le chapitre 4 propose d'évaluer l'impact de la date de naissance des individus sur leurs trajectoires scolaires et professionnelles. Pour réaliser ce travail, nous avons mobilisé différentes sources qui permettent de mesurer les effets du mois de naissance tout au long du parcours éducatif et de la vie active des individus. Celles-ci incluent des données scolaires issues des panels primaire et secondaire de l'Éducation nationale, de la base Scolarité et de la Base du Diplôme national du Brevet, ainsi que des données sur la population adulte qui proviennent de l'enquête Emploi et des recensements de la population.

Nous commençons par montrer que bien que le mois de naissance des individus ne puisse pas être considéré comme totalement indépendant de leurs caractéristiques socio-démographiques, la composante sociologique de cette variable demeure relativement modeste et présente un certain nombre de propriétés qui permettent d'en neutraliser l'influence dans l'estimation des effets du mois de naissance sur les trajectoires scolaires et professionnelles.

Le premier enseignement de notre étude est que le handicap scolaire subi par les élèves nés en fin d'année par rapport à leurs camarades nés en début d'année, lié au fait qu'ils sont les plus jeunes au sein de leur cohorte scolaire, est plus important et persiste davantage au cours de la scolarité que ne le suggèrent les études menées en France sur ce thème (Ferrier, 2003 ; Florin et al., 2004 ; Caille et Rosenwald, 2006). Ces dernières sous-estiment en effet l'impact de l'âge relatif des élèves sur leur résultats scolaires, parce qu'elles ne prennent pas suffisamment en compte les biais de sélection induits par le plus fort taux de redoublement des élèves nés en fin d'année.

En nous inspirant de la stratégie développée par Bedard et Dhuey (2006), nous évaluons l'impact de l'âge de passage des évaluations sur les performances scolaires en utilisant le mois de naissance comme variable instrumentale. Le mois de naissance d'un élève est en effet directement corrélé à son âge absolu au moment des évaluations (cette corrélation étant égale à un pour les élèves qui n'ont ni redoublé, ni sauté une classe), mais n'exerce pas *a priori* d'influence autonome sur sa réussite scolaire. Les résultats obtenus en appliquant cette stratégie d'estimation indiquent que les écarts de performances liés aux différences d'âge sont particulièrement forts à l'entrée en primaire (tout particulièrement pour les élèves issus de milieux sociaux défavorisés) et tendent à diminuer ensuite, sans pour autant disparaître complètement : en CP, les élèves nés en fin d'année obtiennent en moyenne des scores inférieurs de 70 % d'un écart-type à ceux de leurs camarades nés en janvier ; lorsqu'ils passent les épreuves du Brevet, l'écart reste significatif et égal à 17 % d'un écart-type.

Le second enseignement de ce travail empirique est que le mois de naissance exerce une forte influence tant sur la probabilité de redoublement que sur l'orientation des élèves. Nos estimations indiquent que l'essentiel des effets du mois de naissance sur le redoublement s'exercent pendant la scolarité primaire : à l'âge de 11 ans, un élève né en décembre sur trois a redoublé contre seulement un sur six parmi les élèves nés en janvier. Nous montrons également que le mois de naissance influence significativement l'orientation scolaire : en fin de troisième, le fait d'être né en décembre plutôt qu'en janvier augmente de trois points la probabilité d'être orienté en lycée professionnel.

Le troisième enseignement de cette étude est qu'en dépit de cette différenciation des trajectoires scolaires en fonction du mois de naissance, les individus nés en fin d'année ne paraissent pas fortement pénalisés dans leur vie professionnelle par rapport aux individus nés en début d'année. Nous montrons en effet que leur situation d'emploi n'est pas significativement affectée et que l'écart de rémunération séparant les salariés nés en décembre des salariés nés en janvier est faible puisque compris entre 1,5 et 3 %. Nous interprétons ces faibles écarts salariaux, qui proviennent en grande partie de l'entrée plus tardive sur le marché du travail des personnes nées en fin d'année, comme la conséquence du fait que le mois de naissance influence moins le niveau de diplôme que le type de formation suivie (générale ou professionnelle).

Nos estimations révèlent néanmoins qu'il existe une pénalité salariale spécifiquement liée au plus faible niveau de qualification des hommes nés en fin d'année. Celle-ci se traduit par un écart de salaire égal à 1,5 % entre ceux qui sont nés en décembre et ceux qui sont nés en janvier.

Ces résultats suggèrent que si le mois de naissance des individus n'exerce pas une influence déterminante sur leur destin professionnel, il pénalise néanmoins le parcours scolaire de ceux qui sont nés en fin d'année. Dans la mesure où l'on ne choisit pas sa date de naissance, il paraîtrait souhaitable que le système éducatif prévoie un certain nombre de mécanismes institutionnels pour corriger les inégalités scolaires liés à ce facteur.

Nous concluons cette thèse en rappelant les apports de nos travaux, leurs limites et les perspectives de recherche qu'ils ouvrent.

Première partie

OBLIGATION SCOLAIRE ET RENDEMENTS DE L'ÉDUCATION

La première partie de cette thèse étudie les conséquences de la politique éducative qui amorça la modernisation et la démocratisation du système éducatif français : l'allongement de la scolarité obligatoire par la loi Berthoin qui fut votée en 1959 et porta l'âge de fin d'études minimum de 14 à 16 à partir de la rentrée 1967.

Dans le chapitre 1, nous évaluons cette réforme en la comparant à l'*Education Act* de 1973 qui porta au Royaume-Uni l'âge de fin d'études minimum de 15 à 16 ans révolus. Ces deux réformes présentent un grand nombre de points communs : elles furent appliquées à peu près à la même époque et s'inscrivaient dans un processus d'unification du système éducatif. Nous exploitons les discontinuités législatives engendrées par ces deux réformes pour examiner l'impact d'une scolarité obligatoire plus longue sur le destin scolaire et professionnel des individus concernés. Les données utilisées pour effectuer nos estimations sont issues de l'enquête Emploi française et des *Quarterly Labour Force Surveys* britanniques.

Nous montrons que la réforme Berthoin en France et l'*Education Act* au Royaume-Uni ont eu sensiblement le même impact sur l'âge de fin d'études moyen : l'allongement de la scolarité obligatoire à 16 ans a augmenté la durée d'études moyenne de 0,3 année environ, cet effet étant presque exclusivement concentré sur le tiers inférieur de la distribution des âges de fin d'études. En revanche, nos estimations indiquent que ces deux réformes ont eu un impact salarial très différent : alors qu'au Royaume-Uni, l'allongement de l'obligation scolaire a augmenté les salaires masculins et féminins de 4 à 8 % par année d'études supplémentaire, la réforme Berthoin n'a eu aucun impact significatif sur les rémunérations des individus qui sont restés une ou deux années supplémentaires à l'école.

Nos recherches nous conduisent à écarter un certain nombre d'explications possibles à ce paradoxe apparent. Nous montrons en particulier que ni l'hétérogénéité des rendements de l'éducation, ni le rôle du salaire minimum, ni une éventuelle impréparation dans l'application de la réforme Berthoin ne sont susceptibles de rendre compte de manière convaincante de l'impact salarial très différent constaté en France et au Royaume-Uni à la suite de l'extension de l'obligation scolaire.

Nous privilégions une interprétation fondée sur le rôle du diplôme dans la détermination des rendements de la scolarité obligatoire. Nous montrons en effet que la

réforme Berthoin et l'*Education Act* n'ont pas affecté de la même manière le niveau de qualification des élèves français et britanniques. En France, l'allongement de la scolarité obligatoire n'a nullement contribué à diminuer la proportion d'individus quittant l'école sans diplôme. Au contraire, le relèvement de l'âge de fin d'études minimum au Royaume-Uni a entraîné une réduction très sensible du nombre d'élèves sans qualification, en particulier dans la population féminine, et a accru le nombre de titulaires du GCSE, équivalent de notre BEPC. Les estimations que nous avons réalisées à partir des données d'enquêtes internationales sur la littératie des adultes semblent également indiquer que l'augmentation du niveau de qualification de la population britannique s'est accompagnée d'une amélioration qualitative des compétences cognitives des cohortes touchées par la réforme de 1973. Davantage qu'à un pur effet de signal, l'impact positif de l'allongement de l'âge minimum de fin d'études sur les salaires au Royaume-Uni semble donc pouvoir être interprété comme la conséquence des efforts accrus consacrés par les élèves britanniques dans le cadre de la préparation des diplômes qui sanctionnent la fin de la scolarité obligatoire.

Malgré la grande similitude des cursus scolaires dans les deux pays, l'impact différent de la réforme Berthoin et de l'*Education Act* sur la structure des qualifications est la conséquence directe d'un certain nombre de spécificités institutionnelles propres au système éducatif français de la fin des années 1960 : l'existence du redoublement, l'orientation précoce des élèves en filière professionnelle et le statut mal défini du diplôme du BEPC sont trois facteurs qui contribuent à expliquer que la réforme Berthoin ne soit pas parvenue à améliorer significativement le niveau de qualification des individus ayant un faible niveau d'études.

Si les résultats de cette étude indiquent que l'allongement de la scolarité obligatoire en France ne s'est pas traduit à court terme par une réduction des inégalités scolaires, la comparaison avec le cas britannique montre qu'il n'y avait pas là de fatalité. Nos recherches suggèrent en effet que dans les pays dont les systèmes éducatifs sont structurés autour de l'obtention de diplômes, l'allongement de la scolarité obligatoire n'est susceptible d'améliorer les perspectives professionnelles des individus qui entrent précocement sur le marché du travail que si elle leur donne la possibilité de quitter l'école avec une qualification sanctionnée par un diplôme. De ce point de

vue, il existe en France deux obstacles majeurs à la réduction du nombre considérable d'élèves qui abandonnent leurs études sans aucune qualification : la pratique massive du redoublement, qui conduit de nombreux élèves à quitter l'école avant d'avoir pu passer le moindre diplôme et le fait que l'obtention d'un diplôme professionnel exige la poursuite d'études au-delà de l'obligation scolaire.

Chapitre 1

Suffit-il d’allonger la scolarité obligatoire pour augmenter les salaires ? Une comparaison France – Royaume-Uni

1.1 Introduction

Motivation. Compulsory schooling laws have been extensively used in the literature as an exogenous source of variation in educational attainment to estimate the causal impact of education on earnings. Most papers find that instrumental variable estimates of the economic return to compulsory schooling are higher than their naive OLS counterparts and are usually in the range of 10 to 15% per additional year of education. However, a number of recent studies have found the rather puzzling result that these returns are small and even close to zero in certain countries.

Early compulsory schooling papers usually follow the instrumental variable strategy first proposed by Harmon and Walker (1995) to estimate the returns to education in the United Kingdom. This approach consists in estimating a typical Mincerian earnings equation of the form:

$$\ln w_i = \alpha + \beta s_i + x_i' \gamma + \epsilon_i$$

where $\ln w_i$ is the log of hourly wage, s_i is years of schooling and x_i is a vector of observable characteristics such as age, potential work experience, region or gender. This equation of interest is estimated based on a Two-stage least square (TSLS) approach in which schooling is instrumented by the minimum school leaving age faced by the individual. Provided that changes in the minimum school leaving age were implemented at the same time throughout the country of interest, the identification of the impact of increased educational attainment can be achieved through the comparison of average schooling and earnings across cohorts. Using this strategy and exploiting two reforms that raised the minimum school leaving age in England and Wales to 15 in 1947 and to 16 in 1973, Harmon and Walker find that the TSLS estimate of the returns to schooling are higher than the OLS naive estimates (15% versus 6%). Most papers that followed this approach using data from other countries reached similar conclusions¹. The estimation strategy proposed by Harmon and Walker has nevertheless been called into question because it fails to neutralize cohorts fixed effects by not involving a proper control group. Consequently, the entire growth in average schooling across cohorts is attributed to changes in compulsory attendance, ignoring the existence of a positive trend in average educational attainment in most developed countries over the postwar period².

In order to overcome some of the limitations of the Harmon and Walker approach, subsequent papers exploited variations in the timing of compulsory schooling laws across states or regions to control for cohort fixed effects. Acemoglu and Angrist (2000) use changes in child labor laws and school leaving age regulations across US states to estimate both the private and social returns to education. The authors' TSLS estimates of the private returns are close to 10% and are higher than the corresponding OLS estimates (7%). In a series of papers, Oreopoulos (2006b,a, 2007) follows a similar strategy to assess the impact of changes in compulsory schooling laws in Canada and the United Kingdom. Using differences in the timing of the

¹See in particular Callan and Harmon (1999) for Ireland, Brunello and Miniaci (1999) for Italy, Levine and Plug (1999) for the Netherlands, Vieira (1999) for Portugal and Pons and Gonzalo (2002) for Spain.

²As noted by Card (1999) and Oreopoulos (2006a), one can infer the presence of important cohort effects in Harmon and Walker's study from the fact that in their regression outputs, survey year effects show a 0.5 rise in average educational attainment between the 1976 and 1986 surveys, controlling for age and school leaving age dummies.

reforms across Canadian states and British nations, Oreopoulos finds that students compelled to take an extra year of school experienced an average increase of 12% in annual earnings. A number of studies that have adopted this approach suggest that the returns to compulsory education are not restricted to the labor market and may influence outcomes as different as health, criminal behavior, political involvement, fertility or teenage childbearing³. Moreover, Oreopoulos et al. (2006) find evidence that compulsory schooling laws could have contributed to reduce the intergenerational transmission of inequality in the US. Their estimates indicate that a 1-year increase in the education of either parent lowers the probability that a child repeats a grade by between 2 and 4 percentage points.

A potential concern with this approach is that to consistently estimate the returns to schooling, the timing of compulsory law changes must be assumed to be uncorrelated with other policy changes or regional characteristics that could relate to the outcome variables. Yet trends in the outcome variables could eventually differ between states or regions. This problem is usually dealt with by adding state-specific trends to the set of regressors but this raises the question of how plausible are the functional form assumptions that are imposed. Another limitation of this identification strategy is that it is inapplicable to schooling reforms that were implemented at the same time in all parts of a particular country. This is especially true in the European context, where few regional differences can be exploited within countries to estimate the returns to compulsory schooling. To evaluate the impact of such reforms, Regression Discontinuity Design (RDD) appears as a powerful and promising substitute to the difference-in-difference approach and is becoming increasingly popular in the compulsory schooling literature, as in many other fields of applied research. This methodology, which involves both parametric and non-parametric

³Using US and UK data, Milligan et al. (2004) find that compulsory schooling laws improved political involvement by attenuating the drop in voter turnout from 1964 to 2000 in the US (but not in the UK). Using US Census and FBI data, Lochner and Moretti (2004) show that compulsory attendance laws significantly reduced the probability of incarceration and arrest. Lleras-Muney (2005) estimates that one additional year of compulsory schooling in the US decreased mortality after the age 35 by about 3%. Using the same data, León (2004) finds that compulsory schooling laws had a significant negative impact on fertility in the US. Oreopoulos (2007) identifies significant gains from education to health and subjective measures of well-being, using changes in minimum school leaving ages in Great Britain and Ireland. Using US and Norwegian data, Black et al. (2008) find evidence that increased educational attainment reduced the incidence of teenage childbearing in both countries.

techniques, consists in measuring the discontinuity induced by compulsory schooling laws in otherwise continuous educational and economic individual outcomes, to credibly estimate the returns to compulsory education. Oreopoulos (2006a) was the first to apply the RDD methodology to estimate the effect of school leaving age increases in Britain (in 1947) and Northern Ireland (in 1957), which redirected almost half the population of early school dropouts to stay in school for one more year. In both cases, the monetary returns to compulsory schooling are shown to have been very large, ranging from an annual gain in earnings between 10 to 14%.

The consensus view that compulsory schooling laws have large and positive effects on labor market outcomes has nevertheless been challenged by a number of recent studies carried out on European data, which use both difference-in-difference and RDD estimation strategies. Meghir and Pålme (2005) estimate the impact of a major Swedish educational reform which, among other institutional changes, raised compulsory schooling from seven or eight years to nine years in the late 1940s. The fact that this reform was implemented gradually across municipalities allows the authors to perform a difference-in-difference estimation. They find that the overall effect of the reform on earnings was small and not significantly different from zero. Positive monetary gains did however exist for individuals with unskilled fathers, who enjoyed an average earnings increase of 3.4%. Pischke and von Wachter (2006, 2008) take advantage of differences in the timing of changes in compulsory education across German States after World War II to estimate the returns to compulsory schooling in Germany. Their estimates indicate that these returns were basically zero. Oosterbeek and Webbink (2007) evaluate the impact of the extension from 3 to 4 years in the length of vocational training programs in the Netherlands and find no beneficial effect from the change. Devereux and Hart (2008) analyze the 1947 British compulsory schooling law previously studied by Harmon and Walker (1995) and Oreopoulos (2006a). Using the New Earnings Survey Panel Data-set (NESPD), which they argue has superior earnings information compared to the data used in earlier works, the authors apply RDD and find a zero return for women and a modest return for men in the 4-7% range. These estimates are much lower than the very high instrumental variables returns to schooling (10-15%) found in previous studies of the same reform.

At the minimum, these results suggest that compulsory schooling laws do not systematically improve the labor market prospects or early school dropouts. To explain why the estimated returns to compulsory schooling in several European countries appear much lower than most estimates reported on US or Canadian data, two series of arguments have been put forward in the literature. Devereux and Hart (2008) argue that US and Canadian compulsory schooling laws raised the schooling of much fewer people than their European counterparts, so that their stronger apparent impact may simply reflect the higher returns to schooling of a smaller number of compliers. Following this interpretation, the small returns in Britain, where a large proportion of the population was impacted by the law changes, could be viewed as picking up something closer to the average treatment effect. Another candidate explanation for this puzzle is that the benefits of compulsory schooling laws may depend on the institutional context in which they take place. To explain the zero returns to compulsory schooling in Germany, Pischke and von Wachter (2006, 2008) argue that the German school system, which tracks students early, might provide low-achieving students with the skills that matter ultimately in the labor market earlier than Anglo-Saxon school systems. Such a phenomenon could explain why individuals belonging to the lower end of the ability distribution derived little benefit from the introduction of the ninth grade in Germany.

Contribution. With respect to this vast literature, the aim of this study is to throw light into the black box of compulsory schooling laws and to try and identify some of the institutional features that are likely to drive the benefits of compulsory education. More precisely, we argue that in the European context especially, academic credentials are an important element to take into account when assessing the impact of increases in the minimum school leaving age – a dimension that has been neglected in previous studies.

To test this hypothesis, we carry out a comparative analysis of two important compulsory schooling reforms: the 1967 Berthoin edict which raised the minimum school leaving age from 14 to 16 in France and the 1973 Education Act which extended compulsory education from the age of 15 to the age of 16 in the United Kingdom.

We justify this comparison on several grounds. First, we argue that when these compulsory schooling laws came into effect, French and British educational systems shared many parallel institutional features, including a similar structure of academic curriculum and credentials. Second, the French Berthoin reform and the British Education Act both came after a series of reforms aimed at unifying the schooling system and widening the access to secondary education. Finally, as shown in our analysis, these reforms impacted a similar fraction of the population and raised the average educational attainment by almost the same amount.

Our estimation strategy is based on Regression Discontinuity Design, which we believe has the greatest internal validity among applicable evaluation methods. To our knowledge, this study is the first to assess the impact of the French Berthoin reform and the first to apply RDD to the 1973 British compulsory schooling law, which was previously examined together with other law changes and in a different estimation framework by Harmon and Walker (1995) and Oreopoulos (2007).

Using data from national labor force surveys, we find that the 1967 French Berthoin reform and the British 1973 Education Act raised the average years of education by about 0.3. However, our estimates indicate that the raising of the minimum school leaving age in France had no sizable impact on the wages of women and men, whereas the monetary returns to compulsory schooling in Britain are positive and significant for both. We argue that factors such as the heterogeneity of the returns to schooling, the role of wage-setting institutions or a hypothetical defective implementation of the Berthoin reform cannot credibly account for this discrepancy. Instead, we offer an explanation based on the role of academic credentials. Indeed, we find that a major difference between the two reforms is that the fraction of individuals holding no qualifications dropped sharply after the introduction of the new minimum school leaving age in Britain, but remained unchanged in France. We link this observation to a number of institutional features of French and British educational systems and discuss the reasons why in both countries, the actual quantity of education may matter less than credentials in determining the returns to compulsory schooling. We also present some evidence supporting the idea that the benefits derived by British early school dropouts from better credentials were not entirely driven

by signalling mechanisms, but also reflected a substantial improvement in their level of skills. We interpret our findings as suggesting that in countries where sub-degree qualifications play an important role in certifying the knowledge and skills a person has achieved through study, compulsory schooling laws may be effective only if they induce a significant fraction of pupils to become enrolled in grades that lead to academic or vocational certification examinations.

Outline. The remainder of this chapter is as follows: section 1.2 briefly presents the 1967 and 1973 compulsory schooling laws and gives some institutional background on French and British educational systems; section 1.3 explains the estimation strategy; section 1.4 gives a description of the data and provides summary statistics; the main results are discussed in section 1.5; section 1.6 discusses possible interpretations of our findings and section 1.7 concludes.

1.2 Compulsory schooling laws and educational systems in France and Britain

Although implemented in different institutional contexts, the French 1967 Berthoin reform and the 1973 British Education Act share a number of characteristics which support and justify the comparison of their respective effects. Both compulsory schooling laws were designed to accelerate the democratization of secondary education and were part of a number of structural reforms aimed at establishing a comprehensive school system in place of the preexisting compartmentalized education. They were enacted in the same period and raised the minimum schooling age to 16. The Berthoin reform and the Education Act were also carefully prepared to make sure that the school system was ready to absorb the subsequent increase in total enrollment.

1.2.1 France

The 1967 Berthoin Reform. In France, compulsory schooling was extended twice during the 20th century. Primary schooling was made compulsory for female and male

children from the age of 6 to the age of 13 in 1882, with some exceptions. In 1936, compulsory schooling was extended to the age of 14. This measure was implemented at the start of the 1936 school year, for pupils born after January 1, 1923. In 1959, the Berthoin edict added two years to compulsory schooling. From then on, the minimum school leaving age was set at 16. The edict stated that the measure would be applicable for “*children over six years of age on January 1, 1959*”, in order to have enough time to be prepared for this change. The Berthoin reform came into effect in 1967 for cohorts born after January 1, 1953 (individuals born in 1953 reaching the age of 14 in 1967). This measure was thus in no way precipitated, since a waiting period of eight years was planned between the passing and the implementation of the law.

The French schooling system. In the French system, the school year starts in September and finishes in June. School cohorts are calendar cohorts, since children are supposed to start school the year they turn six. For instance, the pupils who started primary schooling during school year 1956-1957 were in principle born between January 1 and December 31, 1950.

The extension of compulsory schooling in 1967 was envisaged as an important building block for the unification of the French school system. In the postwar period, a number of structural reforms aimed at widening the access to secondary education. Until the late 1950s, the system was divided into two completely hermetic tracks, based essentially on social status. A vast majority of children (about 90% of the total in 1945) attended a basic track formed of 8 years of extended primary schooling⁴, at the end of which they took the primary schooling certification examination (*Certificat d'Études Primaires*) that only half of them passed successfully at the age of 14. Those who succeeded in passing this examination were admitted into junior secondary schools and would end their studies after 3 or 4 years of post-elementary education. The other pupils either dropped out of school at the age of 14 or entered basic vocational training. Outside this basic track, a minority of children from higher socioeconomic backgrounds followed a longer academic curriculum consisting

⁴Primary schooling included five basic grades from the age of 6 to the age of 11 and three additional grades (called *Classes de fin d'études*) from the age of 11 to the age of 14.

of 12 grades (5 years of primary schooling and 7 years of secondary schooling) at the end of which they would take the *Baccalauréat* examination, a diploma required to pursue university studies.

Between 1945 and 1975, the French school system was unified step by step and was transformed according to a new structuring principle. It was decided that schooling would consist of three educational stages: primary schooling (5 years between age 6 and age 11), junior secondary schooling (4 years between age 11 and age 15) and senior secondary schooling (3 years between age 15 and age 18), the first two stages being common to all children. The unification of primary schooling was achieved in the early 1960s. The unification of secondary schooling started in 1959 and was not yet achieved when the raising of the minimum school leaving age came into effect in 1967. At that time, about half the pupils were still enrolled in the former system of extended primary schooling while the other half would follow the integrated curriculum in one of the newly created junior high schools. The Berthoin reform accelerated this process, which was completed in 1977 when all remaining institutional differences between existing junior secondary schools were abolished.

An important feature of the French primary and secondary school curriculum is that it is structured by a set of academic and vocational credentials. When the Berthoin reform was initiated, the certificate of primary schooling examination had lost its importance since it was no longer required to enter junior secondary schools. The two main academic credentials that could be obtained by the end of secondary schooling were the *Baccalauréat*, for the most academically inclined pupils, and the CAP or BEP⁵ vocational qualifications, for the less able pupils. These qualifications were gained through examinations that students took when they turned 17 or 18. Alongside these two major qualifications, pupils had the opportunity of obtaining a lower qualification called the BEPC (*Brevet d'Études du Premier Cycle*) at the end of junior secondary schooling (grade 9). However, this credential was never explicitly designed to certify the completion of compulsory education and very few individuals would hold it as their highest qualification. The relatively low status of this junior secondary schooling credential will be discussed in more detail in section 1.6.

⁵CAP: *Certificat d'Aptitude Professionnelle*. BEP: *Brevet d'Études Professionnelles*.

1.2.2 Britain

The 1973 Education Act. In Britain, compulsory schooling was increased three times during the 20th century. In 1918, the Fisher Act extended compulsory education from the age of 12 to the age of 14. In 1944, the minimum school leaving age was further raised to 15 by the Butler Act, which did not come into effect before 1947 because of the effects of World War II. The third extension of compulsory education, from the age of 15 to the age of 16, was decided upon in 1964 and enacted in 1973.

The British schooling system. In the United Kingdom, the school year starts in September and ends in June, as it does in France. However, both the definition of school cohorts and of the school starting age are different. A school cohort is formed of all children born between September 1 of year t and August 31 of year $t + 1$. The starting age was set at five years by the 1870 Education Act and has remained stable since.

In much a similar way as the French Berthoin edict, the 1973 British Education Act came with a series of reforms aimed at widening the access to secondary education. In the early 1950s, secondary education in Britain was organized according to a “Tripartite” system, in which schools were divided into three categories: Grammar schools, Technical schools and Modern schools. Pupils were assigned to each of these three types of establishments through the so-called “Eleven Plus” selective examination. At the national level, the fraction of students being admitted into the selective Grammar schools was approximately 25%. Increasingly criticized as promoting an “elitist” form of education, this system was gradually replaced by the current Comprehensive system after 1965 and was formally abolished in 1976. When the 1973 Education Act was implemented, most Local Education Authorities had switched to the new comprehensive system, although a number of counties continued to operate the former Tripartite system (Pischke and Manning, 2006).

Academic and vocational credentials play an important role in the British education system, as they do in France. When the new 1973 Education Act was enacted, two types of qualifications could be obtained at the end of compulsory education. The General Certificate of Education (GCE) was designed for academically oriented

pupils (mostly enrolled in Grammar schools). The examinations were graded into ordinary level (O-Level) for 16-year-olds and advanced level (A-Level) for 18-year-olds. The Certificate of Secondary Education (CSE) qualification was intended for less able pupils in mainstream secondary education and examinations took place at the age of 16. There was an overlap between these two types of certificates in that a CSE grade 1 result was regarded as equivalent to an O-level. Depending on their grades on the O-level and CSE examinations, British pupils could choose either to follow the academic track leading to the A-Level that would qualify them for higher education, or undertake a variety of basic or more advanced vocational qualifications which were usually provided by private institutions⁶.

1.3 Estimation strategy: a Regression Discontinuity Design

Regression Discontinuity Design appears as the most appropriate estimation strategy to evaluate the impact of French and British compulsory schooling laws on educational and labor market outcomes.

1.3.1 A “sharp” discontinuity design

Since both these reforms were implemented uniformly in France and Britain, a difference-in-difference approach of the type adopted in existing papers (Acemoglu and Angrist, 2000; Oreopoulos, 2006b, 2007; Pischke and von Wachter, 2008) is not easily applicable in the absence of a readily available control group. In this context, a more appealing approach is to use regression discontinuity. This quasi-experimental evaluation methodology exploits the discontinuity in the probability of being treated to evaluate the impact of treatment, by comparing the average outcomes of observations located just left and right of a cutoff point. In the case of compulsory schooling laws, the discontinuity is created by the raising of the minimum school leaving age for cohorts born after a certain cutoff date. Following the distinction introduced by the

⁶These include City and Guilds crafts, RSA diplomas, SCOTVEC and BTEC diplomas, YT qualifications and Trade apprenticeship.

RDD literature, the design here is “sharp” in the sense that the assignment to treatment (the new compulsory schooling regime here) switched from 0 to 1 for cohorts born after the cutoff date. Under the assumption that average potential outcomes when receiving and when not receiving the treatment are a continuous function of a “forcing variable” (here, the school cohort an individual belongs to), the discontinuity in the compulsory schooling regime can be exploited to measure the impact of one or two additional years of compulsory education on the average educational attainment and labor market outcomes of individuals born around the cutoff date. Taking the ratio of the reform’s impact on labor market outcomes over its impact on educational attainment will enable us to retrieve the monetary and non-monetary returns to compulsory education.

We explain below how the reduced form impact of changes in compulsory schooling on educational and labor market outcomes can be estimated within the RDD framework. We then describe the way in which the implied returns to compulsory education can be derived from these reduced form estimates. Our approach essentially follows the guidelines of Imbens and Lemieux (2007) for regression discontinuity design.

1.3.2 Reduced form estimation

We are interested in evaluating the impact of minimum school leaving age increases on a series of outcomes (wages, employment status, etc.). The effect of treatment is potentially heterogenous across individuals. Let $T_i \in \{0, 1\}$ denote the treatment received by individual i , and $Y_i(0)$ and $Y_i(1)$ the pair of potential outcomes for this individual. The assignment T_i is a deterministic function of the school cohort the individual belongs to (the “forcing” variable), which we denote X_i :

$$T_i = \mathbb{1}\{X_i \geq c\}$$

All individuals belonging to school cohorts posterior or equal to cohort c receive the treatment (i.e. face the new minimum school leaving age) whereas those belonging to earlier cohorts are not eligible for the treatment (i.e. face the previous minimum

school leaving age). In the case of the French Berthoin reform, the cutoff point c is 1953 since the raising of the minimum school leaving age to 16 came into effect for individuals born in 1953 and after. The corresponding cutoff point for the 1973 British Education Act is 1958 (i.e. individuals born after September 1, 1958).

Our parameter of interest is the average causal effect of the treatment at the discontinuity point:

$$ATE = E[Y_i(1) - Y_i(0)|X_i = c]$$

Under the assumption that the conditional regression functions $E[Y_i(0)|X_i = x]$ and $E[Y_i(1)|X_i = x]$ are continuous in x , this parameter can be estimated by comparing the outcomes of individuals located just below and just above the cutoff point c :

$$\begin{aligned} & \lim_{x \downarrow c} E[Y_i(1)|T_i = 1, X_i = x] - \lim_{x \uparrow c} E[Y_i(0)|T_i = 0, X_i = x] \\ &= \lim_{x \downarrow c} E[Y_i(1)|X_i = x] - \lim_{x \uparrow c} E[Y_i(0)|X_i = x] \\ &= E[Y_i(1) - Y_i(0)|X_i = c] \end{aligned}$$

The effect of changes in compulsory schooling can thus be measured by the value of the discontinuity in the expected value of the outcome of interest at the cutoff point. This can be done first graphically by plotting the average value of the outcome for each school cohort and by inspecting whether there is any evidence of a jump in the conditional mean of the outcome around the cutoff point. In our analysis, we will add smooth fitted lines based on polynomial regressions to make the graphs more easily readable.

Several methods can be found in the literature to perform the statistical evaluation of the “jump” in the outcome variable at the threshold c . In this study, we present the results of the two most common RDD estimation techniques: global polynomial regression and local linear regression.

Parametric estimation. Following the few existing studies that have applied RDD to compulsory schooling laws (Oreopoulos, 2006b; Devereux and Hart, 2008), our first approach consists in estimating the discontinuity parametrically by the means

of a global polynomial approximation⁷. This methodology involves using the whole sample and choosing a flexible high-order polynomial to fit the relationship between the outcome Y_i and the forcing variable X_i , allowing for an intercept shift at the cutoff point:

$$Y_i = \alpha_0 + \alpha_1 T_i + g(X_i - c) + \epsilon_i \quad (1.1)$$

where the function g is a polynomial of order p :

$$g(x) = \gamma_1 x + \gamma_2 x^2 + \dots + \gamma_p x^p \quad (1.2)$$

To allow for enough flexibility in the parametrization of the control function, our empirical specifications use a 4th order polynomial in school cohorts. In equation (1.1), the estimated coefficient $\hat{\alpha}_1$ on the treatment variable measures the reform's average causal effect on the outcome Y at the assignment threshold c .

Non-parametric estimation. To assess the stability of our results, we complement the parametric estimation with a non-parametric evaluation of the discontinuity which uses local linear regression. This method consists in fitting linear regression functions to observations located within a given distance h of the cutoff point and separately on each side. As shown by Fan and Gijbels (1996) and Hahn et al. (2001) among others, local linear regression displays a number of attractive features. In addition to producing estimates which are not too sensitive to outcome values for observations far away from the threshold, local linear regression minimizes the edge effects usually associated with kernel-type nonparametric estimation. Furthermore, this specification does not force the slope coefficients to be the same on both side of the cutoff point, as in the global polynomial setting⁸.

From a practical point of view, the local linear regression estimation of the impact of a compulsory schooling law on an outcome Y can be estimated by restricting the

⁷We use a global rather than a split polynomial approximation (Lee et al., 2004; Lee, 2008) because the former approach performs better when the forcing variable is discrete and takes a small number of distinct values, as in the case of school cohorts.

⁸As emphasized by Imbens and Lemieux (2007), the main benefit of local linear regression is that the estimates of $E[Y(0)|X = x]$ rely exclusively on observations on $Y(0)$ from the left of the discontinuity point and do not depend on observations on $Y(1)$, and vice versa for the estimates of $E[Y(1)|X = x]$.

sample to individuals belonging to school cohorts in the interval $[c-h, c+h]$ (where h denotes the selected bandwidth) and by estimating a kernel-weighted linear regression based on the following equation⁹:

$$Y_i = \alpha_0 + \alpha_1 T_i + \gamma_0(X_i - c) + \gamma_1 T_i(X_i - c) + \epsilon_i, \quad X_i \in [c-h, c+h] \quad (1.3)$$

where the treatment variable T_i takes the value 1 for individuals facing the new minimum school leaving age and the kernel weights $w_c = K_h(X_i - c)$ for a chosen bandwidth h . We use a triangular kernel $K_h(u) = (1 - \frac{|u|}{h})\mathbf{1}(|u| < h)$ which is boundary optimal (Cheng et al., 1997). As in the global polynomial regression setting, the estimated coefficient $\hat{\alpha}_1$ on the treatment variable in equation (1.3) measures the average causal effect of the minimum school leaving age increase on the outcome Y at the assignment threshold c . To select the bandwidth h , we implement the cross-validation procedure developed by Ludwig and Miller (2005, 2007).

Equation (1.3) can be augmented to control for covariates Z_i in addition to the forcing variable X_i . In the RDD setting, the inclusion of such covariates essentially serves the purpose of improving the precision of the estimation.

Standard errors. An important feature of our empirical setting is that the forcing variable X_i (school cohort) is discrete by nature. As emphasized by Lee and Card (2008), this implies that the typical high-order polynomial specification of the regression function used in the parametric RDD estimation will induce deviations from the true underlying function $E(Y|X = x)$, which results in random specification errors with a group structure (since all observations with the same value for the forcing variable X_i will mechanically share the same specification error). Lee and Card demonstrate that conventional standard errors which ignore this group structure tend to overstate the precision of the estimated treatment effect. The authors propose a simple procedure to incorporate this feature into the standard errors. Un-

⁹This version of the local linear estimation is equivalent to finding:

$$\min_{\alpha_0, \alpha_1, \gamma_0, \gamma_1} \sum_{i=1}^N [Y_i - \alpha_0 - \alpha_1 T_i - \gamma_0(X_i - c) - \gamma_1 T_i(X_i - c)]^2 K_h(X_i - c)$$

where $K_h(u)$ is the kernel, and h is the bandwidth.

der the assumption that specification errors are random and that they are the same in $E[Y_i(0)|X_i = c]$ and $E[Y_i(1)|X_i = c]$, it can be shown that cluster-consistent standard errors on the distinct values of the forcing variable X_i should be used for inference instead of the conventional heteroskedasticity-consistent standard errors. Following Imbens and Lemieux (2007), we also use Lee and Card's approach in the nonparametric framework (local linear regression) by clustering the standard errors at the school cohort level¹⁰.

1.3.3 IV estimation of the returns to compulsory schooling

The reduced form estimation procedure presented above allows us to evaluate the causal impact of raising the minimum school leaving age on educational and labor market outcomes. In this section, we show how these estimates can be combined to evaluate the returns to compulsory schooling.

Let S_i denote individual i 's educational attainment, measured by the age she completed full-time education and Y_i some labor market outcome of interest. Under the hypothesis that an increase in educational attainment has a causal impact on this labor market outcome, the ratio of the jump induced by compulsory schooling laws in the labor market outcome to the corresponding jump in educational attainment measures the return to an additional year of compulsory schooling, which we denote β :

$$\beta = \frac{\lim_{x \downarrow c} E[Y|X = x] - \lim_{x \uparrow c} E[Y|X = x]}{\lim_{x \downarrow c} E[S|X = x] - \lim_{x \uparrow c} E[S|X = x]}$$

If the returns to compulsory education are heterogenous, β can be interpreted as measuring the returns to compulsory education of those individuals who were compelled to increase their educational attainment as a result of the new compulsory schooling requirement (i.e. “compliers”) and who belong to the first cohort facing the new minimum school leaving age (the estimated returns to compulsory education are therefore local by nature). This will be true as long as the following monotonicity condition holds¹¹: raising compulsory schooling will induce some individuals to in-

¹⁰specifically, we use the HC1 estimator from MacKinnon and White (1985) which applies a finite sample correction to the Eicker-Huber-White estimator of clustered standard errors (Eicker, 1963; Huber, 1967; White, 1980).

¹¹See Imbens and Angrist (1994) and Imbens and Lemieux (2007).

crease their own educational attainment but no one to drop out earlier from school¹², which seems a fairly reasonable assumption in the present case.

Parametric estimation. The returns to compulsory schooling can be evaluated in a parametric fashion by estimating on the whole sample the following equation using TSLS:

$$Y_i = \alpha_0 + \beta S_i + g(X_i - c) + \epsilon_i$$

where S_i is the endogenous regressor, the constant α_0 and the p -order polynomial $g(X_i - c)$ (see equation (1.2)) are the exogenous variables and the assignment variable $T_i = \mathbb{1}\{X_i \geq c\}$ is the excluded instrument.

Non-parametric estimation. The local linear regression approach can also be applied to estimate the returns to compulsory schooling. We then need to restrict the sample to individuals belonging to school cohorts in the interval $[c - h, c + h]$ (with h the selected bandwidth) and estimate the parameter β by TSLS methods based on the following triangular kernel weighted regression function:

$$Y_i = \alpha_0 + \beta S_i + \gamma_0(X_i - c) + \gamma T_i(X_i - c) + \epsilon_i, \quad X_i \in [c - h, c + h]$$

where the schooling variable S_i is the endogenous regressor, the constant α_0 and the regressors $(X_i - c)$ and $T_i(X_i - c)$ are the exogenous variables and the assignment variable T_i is the excluded instrument. The regression weights w_c are computed from a triangular kernel for the chosen bandwidth h .

1.4 Data and summary statistics

To assess the impact of increased compulsory schooling on educational attainment and labor market outcomes in France and Britain, we construct our samples from national labor force surveys.

¹²This would be the case for instance if some individuals responded to the raising of the minimum school leaving age to 16 by dropping out of school at the age of 17 instead of 18 under the previous less stringent compulsory schooling regime.

1.4.1 France

We perform our analysis of the French Berthoin compulsory schooling using the *Enquête Emploi* (EE) survey which is carried out every year by the *Institut National de la Statistique et de l'Administration Économique* (Insee) on a sample of some 60,000 households. The EE is a repeated cross section, but has a panel dimension, since a given household is surveyed three years in a row. Working on the 13 surveys carried out between 1990 and 2002 has the advantage of allowing the constitution of a fairly large sample. Another attractive feature of the EE is that it combines precise information on schooling, earnings and employment. An individual's school cohort is simply given by her year of birth, since school cohorts and calendar years coincide in the French educational system. Information on individual educational attainment is given by two variables: the age completed full-time education and the highest held qualification, which we grouped into six categories: no qualification; primary schooling certificate; low vocational qualification (CAP or BEP without the BEPC); junior secondary schooling certificate (BEPC); intermediate vocational certificate (CAP or BEP with the BEPC); senior secondary schooling certificate or above (which includes the *Baccalauréat*, advanced vocational training, higher education degrees, etc.).

Individual gross monthly wage is recorded as a continuous variable only since 1990¹³. Hourly earnings are derived from monthly wages by dividing the gross monthly wage (converted into 2005 euros) by the usual number of weekly hours worked. We use information on an individual's employment status to create two binary variables for employment and self-employment. To ensure that all included school cohorts are surveyed each year between 1990 and 2002, we restricted the sample to both female and male respondents who were born in France between 1944 and 1962, left full-time education between the age of 6 and the age of 25 and were aged between 25 and 60 when surveyed. To avoid using several observations for the same individual¹⁴, we restrict the sample to individuals who are interviewed in the first of the three survey waves. We are left with a sample of 188,903 observations.

¹³In previous surveys, respondents were asked to indicate their monthly wage in 19 brackets. Given the statistical pitfalls associated with earning brackets and midpoint imputation, we decided not to use these surveys.

¹⁴using the answers from several waves of the survey would induce strong serial correlation in individual outcomes.

Table 1.1: *Summary statistics (France). Source: Enquête Emploi (1990-2002).*

FRANCE – ENQUÊTE EMPLOI (1990-2002)					
Variable	Mean	s.d.	Min	Max	Number of obs
<u>INDIVIDUAL CHARACTERISTICS:</u>					
School cohort	1953.3	(5.33)	1944	1962	188,903
% female	0.48	(0.50)			188,903
Age	42.7	(6.45)	28	58	188,903
% in employment (ILO definition)	0.80	(0.40)			188,903
% self-employed (ILO definition)	0.14	(0.34)			150,597
Log of hourly wage (in 2005 euros)	2.22	(0.42)	1.08	7.75	114,977
<u>SCHOOLING:</u>					
Schooled after the 1967 Berthoin reform	0.54	(0.50)			188,903
Age left school	17.58	(2.82)	6	25	188,903
Left school by age 14	0.17	(0.38)			188,903
Left school by age 15	0.22	(0.41)			188,903
Left school by age 16	0.38	(0.48)			188,903
Left school by age 18	0.71	(0.46)			188,903
<u>ACADEMIC CREDENTIALS:</u>					
No qualifications	0.31	(0.46)			188,903
Primary schooling certificate	0.31	(0.46)			188,903
Low vocational qualification	0.22	(0.41)			188,903
Junior secondary schooling certificate	0.09	(0.29)			188,903
Intermediate vocational qualification	0.10	(0.30)			188,903
Senior secondary schooling certificate or above	0.28	(0.45)			188,903

Notes: The sample is constructed from the 13 *Enquête Emploi* surveys that were conducted in May each year by the French national statistical institute (Insee) between 1990 and 2002. It includes all respondents who were born in France, belong to the 1944 to 1962 school cohorts, left full-time education between the age of 6 and the age of 25 and were aged between 25 and 60 when surveyed. Individuals born between January 1 of year t and the December 31 of the same year are assigned to school cohort t . Each individual is only included once in the sample, as we keep only the information collected during the first of the three successive years during which a respondent is interviewed. Details on the grouping of academic and vocational credentials are given in section 1.4.1.

Descriptive statistics for this sample are displayed in table 1.1.

1.4.2 Britain

To study the consequences of the raising of the minimum school leaving age to 16 in the United Kingdom, we use the British Quarterly Labor Force Survey (QLFS) which provides the same type of information as its French equivalent. The QLFS is a continuous sample survey and collects information on about 60,000 households every quarter. We combine the 44 surveys conducted between Fall 1993 and Summer 2004 to construct a large dataset that contains information on individual education and labor market outcomes. School cohorts can be retrieved using the year and month of birth: individuals born between September 1 of year t and August 31 of year $t + 1$ are assigned to school cohort t , so an individual born on April 5, 1960 is considered to belong to school cohort 1959. In the QLFS, educational attainment is measured by the age at which the respondent completed full-time education as well as by her highest held academic qualification, which we group into five categories: no qualification; low vocational qualification (NVQ level 1 and 2 or equivalent¹⁵); junior secondary schooling qualification (GCE O-Level or CSE); intermediate vocational qualification (NVQ level 3 or equivalent¹⁶); senior secondary schooling certificate or above (includes the GCE A-Level, NVQ level 3 and 4, higher education degrees, etc.).

Information on individual hourly pay is available only for individuals interviewed in the first and fifth wave of the survey. We convert the corresponding wages in 2005 British pounds using the quarterly consumer price index series. Finally, the LFS can be used to create binary variables for employment and self-employment status. Imposing similar restrictions to those used in the French case, the final sample includes 292,934 males and females who were born in the UK between 1949 and 1967, left full-time education between the age of 6 and the age of 25 and were aged between 25 and 60 when surveyed and interviewed in the fifth of the five survey waves¹⁷. Descriptive statistics for this sample are displayed in table 1.2.

¹⁵These include the BTEC first or general certificates, YT/YTP certificates, SCOTVEC modules, RSA and low City and Guilds crafts.

¹⁶These include GNVQ intermediate certificates, the RSA diploma, intermediate City and Guilds crafts and the BTEC/SCOTVEC first or general diploma.

¹⁷We used the fifth rather than the first wave because of its smaller average non-response rate.

Table 1.2: *Summary statistics (United Kingdom). Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)					
Variable	Mean	s.d.	Min	Max	Number of obs
<u>INDIVIDUAL CHARACTERISTICS:</u>					
School cohort	1958.4	(5.49)	1949	1967	292,934
% female	0.52	(0.50)			292,934
Age	39.86	(6.32)	26	55	292,934
% in employment (ILO definition)	0.81	(0.39)			292,934
% self-employed (ILO definition)	0.13	(0.34)			237,935
Log of hourly wage (in 2005 pounds)	2.13	(0.56)	0.85	7.11	168,790
<u>SCHOOLING:</u>					
Schooled after the 1973 Education Act	0.56	(0.50)			292,934
Age left school	17.05	(2.27)	6	25	292,934
Left school by age 14	0.01	(0.10)			292,934
Left school by age 15	0.20	(0.40)			292,934
Left school by age 16	0.61	(0.49)			292,934
Left school by age 18	0.82	(0.38)			292,934
<u>ACADEMIC CREDENTIALS:</u>					
No qualifications	0.21	(0.41)			292,934
Low vocational qualification	0.05	(0.21)			292,934
Junior secondary schooling certificate	0.25	(0.43)			292,934
Intermediate vocational qualification	0.18	(0.38)			292,934
Senior secondary schooling certificate or above	0.31	(0.46)			292,934

Notes: The sample is constructed from the 44 Quarterly Labour Force Surveys that were conducted by the UK Office of National Statistics between 1993 (Autumn) and 2004 (Summer). It includes all respondents who were born in the United Kingdom, belong to the 1949 to 1967 school cohorts, left full-time education between the age of 6 and the age of 25 and were aged between 25 and 60 when surveyed. Individuals born between September 1 of year t and the August 31 of year $t + 1$ are assigned to school cohort t . Each individual is only included once in the sample, as we keep only the information collected during the fifth of the five successive quarters during which a respondent is interviewed. Details on the grouping of academic and vocational credentials are given in section 1.4.2.

1.5 Results

In this section, we use the RDD methodology to estimate the impact of the raising of the minimum school leaving age to 16 in France and Britain on schooling and earnings, separately for men and for women.

1.5.1 OLS estimates of the returns to schooling

As a starting point, our constructed samples can be used to compute the OLS estimates of the returns to education in both countries. The values of these estimates provide a natural benchmark to compare the causal returns to compulsory schooling that will be derived from the 1967 Berthoin reform and the 1973 Education Act.

Table 1.3: *Naive OLS estimates of the returns to schooling in France and Britain, calculated separately for female and male workers. Sources: Enquête Emploi (1990-2002) and Quarterly Labour Force Survey (1993-2004).*

	FRANCE (dependent var.: log of hourly wage) OLS (1)	UK (dependent var.: log of hourly wage) OLS (2)
<u>PANEL A: WOMEN</u>	0.083*** (0.001)	0.119*** (0.001)
Number of observations	56,396	86,588
<u>PANEL B: MEN</u>	0.072*** (0.001)	0.093*** (0.001)
Number of observations	58,581	82,202

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. Each coefficients comes from a separate regression of log of hourly wages on age left full-time education. Regressions include a quartic in age and a full set of survey year and school cohort dummies. The French sample (column 1) is constructed from the 1990 to 2002 French *Enquête Emploi* surveys and includes workers with a positive wage, born in France between 1944 and 1962 and aged between 25 and 60 when surveyed. The British sample (column 2) is constructed from the 1993 to 2004 Quarterly Labour Force Surveys and includes workers with a positive wage, born in the UK between 1949 and 1967 and aged between 25 and 60 when surveyed. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

The standard OLS specification regresses the log of hourly wage on the age left

full-time education and includes a quartic in age and a full-set of survey year and school cohort dummies. We perform separate regressions for males and females in France and Britain. The naive estimates of the returns to education are reported in table 1.3 with the coefficients for France in column 1 and the coefficients for Britain in column 2.

The naive estimates of the returns to education are close to 10% in both countries, which is the typical value found in the literature. The returns appear higher for women than for men (0.083 versus 0.072 in France and 0.119 versus 0.093 in Britain) and the estimates for Britain are slightly larger than the corresponding estimates for France.

Of course, this comparison tells us nothing about the causal impact of education on earnings in France and Britain, because the number of years of schooling is endogenous in the wage equation. Compulsory schooling laws are needed to circumvent this problem, as they provide us with an exogenous source of variation in educational attainment.

1.5.2 Impact of compulsory schooling laws on school leaving age

The assessment of how changes in compulsory schooling affected educational attainment raises two distinct issues.

The first question that needs to be addressed is the extent to which the French and British reforms can be considered as “adequate” natural experiments. The extension of compulsory schooling in both countries can be thought of as a source of exogenous increase in the average years of education only if some individuals who faced the new minimum school leaving age would have left school earlier if the reform had not been implemented. In other words, for changes in compulsory schooling laws to be exploited as natural experiments, it requires the new minimum school leaving age to be *binding* for a substantial fraction of the population¹⁸. We show below that this

¹⁸Compulsory schooling laws are not necessarily binding. For instance, Oreopoulos (2007) shows that when the minimum school leaving age was extended from 14 to 15 in the Republic of Ireland (1972), only 10% of the children were leaving school before the age of 15, which explains that the reform had hardly any impact on average educational attainment.

was true in France and the UK when the 1967 and 1973 compulsory schooling laws came into effect.

The second issue that deserves attention is the impact of compulsory schooling laws on the *distribution* of school-leaving ages. Lang and Kropp (1986) first noted that the positive effect of such laws on individual educational attainment could well extend beyond the subpopulation of early school dropouts. Indeed, if the age at which an individual left full-time education partially acts as a signal of her productivity on the labor market, raising the minimum-school leaving age should give all students the incentive to increase their own schooling, in order to maintain their relative educational premium over lower achievers. In such a scenario, one would expect the whole distribution of school-leaving ages to shift to the right in response to changes in compulsory attendance¹⁹. Our empirical analysis shows that while the existence of such “spillover” effects cannot be completely ruled out in France and Britain, these effects were small and restricted to individuals with a school leaving age of less than 18.

France. Figure 1.1 shows how school leaving ages changed across the 1944-1962 school cohorts. Separate graphs for women and men are displayed in figure 1.2. Each curve in these graphs indicates the fraction of individuals who left school by a given age; the first curve starting from the bottom indicates the proportion of individuals who left school by the age of 14, the second, the proportion of individuals who left school by the age of 15, and so on up to the age of 18. Three lessons can be drawn from this graph. First, it appears that the new minimum school leaving age introduced by the Berthoin law was binding for a significant fraction of the population: about a third of the pupils born in 1952 left school before the age of 16. The second observation is that the Berthoin reform did force some – but not all – individuals to stay in school for one or two years more than they would have if the reform had not been implemented.

¹⁹Note that this “spillover” phenomenon could invalidate a certain number of fairly straightforward difference-in-difference strategies. For instance, comparing the evolution of schooling and earnings of individuals belonging to the age group “caught up” by the new minimum school leaving age (treatment group) with the corresponding evolution of individuals belonging to the age group immediately above (control group) would certainly yield unreliable estimates if the composition of both these groups changes discontinuously at the cutoff point (some individuals shifting from the “treatment” to the “control” group because of the new minimum school leaving age).

Figure 1.1: Evolution of the distribution of school leaving ages in France (women and men). Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).

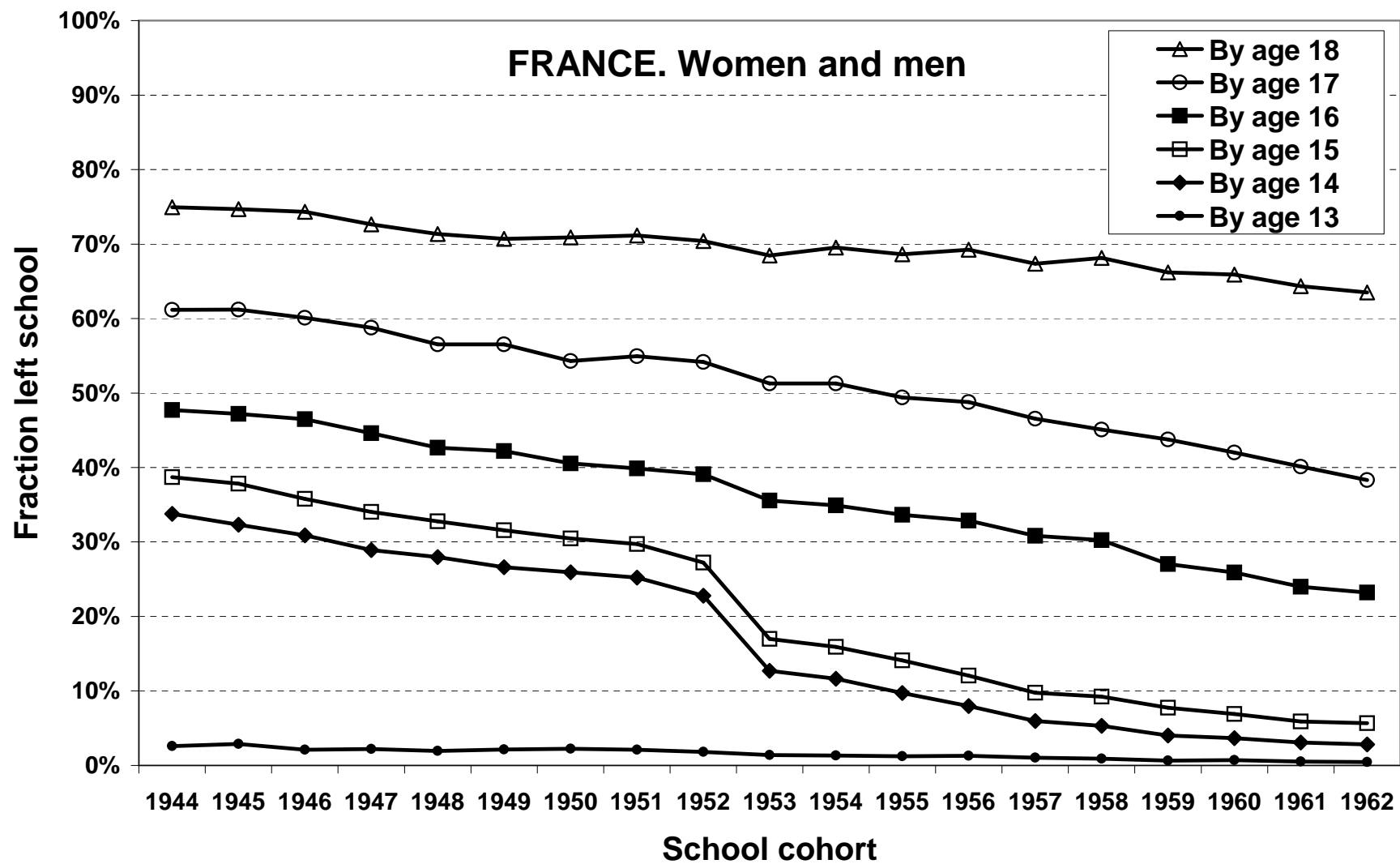
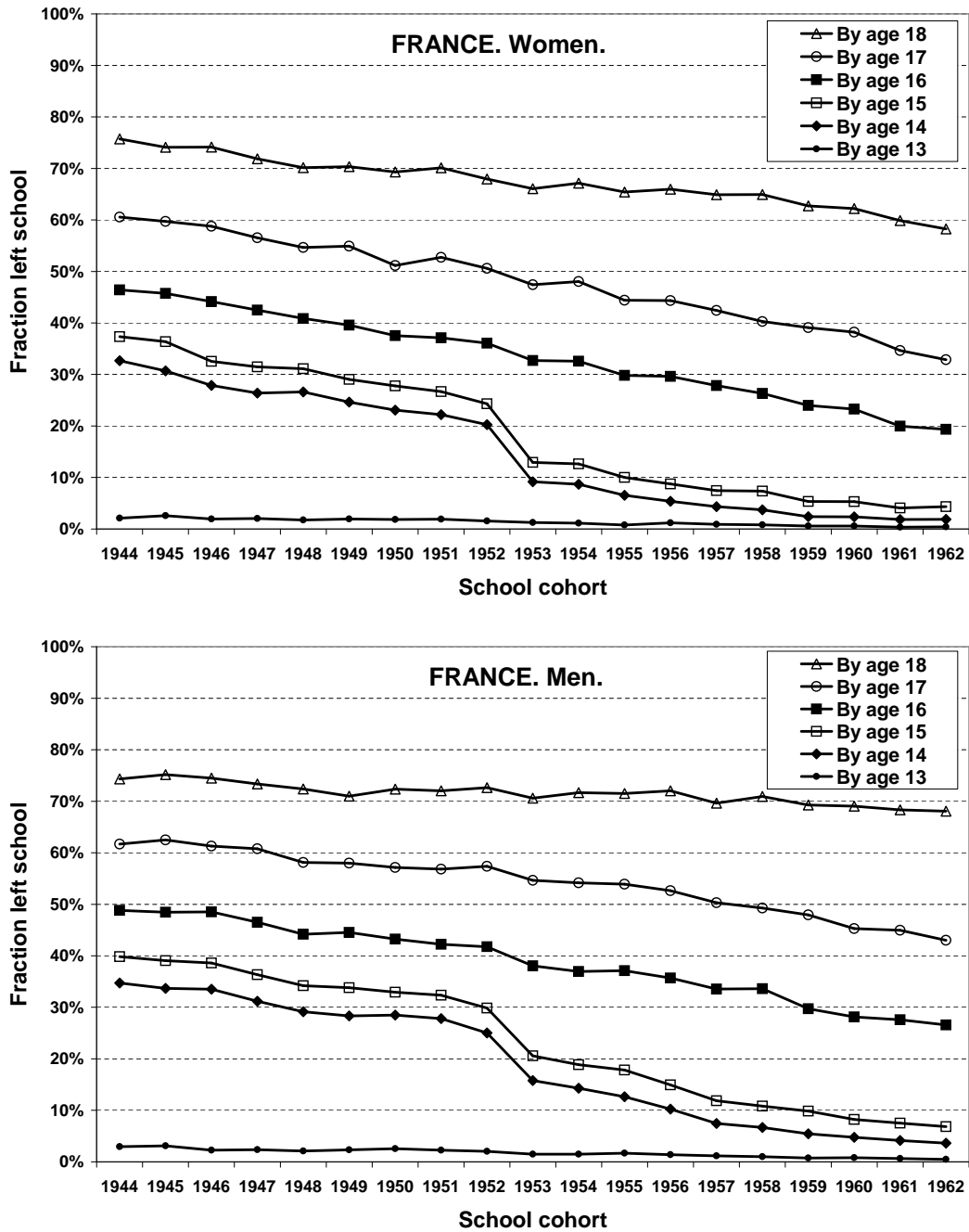


Figure 1.2: Evolution of the distribution of school leaving ages in France, calculated for women and men separately. Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).



It can be inferred from the graph that about half the individuals “caught up” by the new minimum school leaving age complied with the new mandatory requirement, while the other half did not. This pattern can be observed in most countries studied in the compulsory schooling literature and results from both imperfect enforcement of compulsory education and measurement error in the school leaving age variable. It should be noted however that these factors are unlikely to present a serious threat to identification, as long as a clear discontinuity in average educational attainment can be observed across school cohorts²⁰. Finally, the third lesson that can be drawn from these graphs is that the Berthoin edict may have triggered a limited “spillover” effect on educational attainment. Indeed, the reform seems to have slightly affected the distribution of school leaving ages *beyond* the age of 16, as the fraction of individuals who left school by the age of 16 and 17 exhibits a slight discontinuity at the cutoff point.

We take a closer look at this “spillover” phenomenon by using the RDD framework exposed above to estimate both parametrically and non-parametrically the size of the discontinuity of each school leaving age curve at the cutoff point. The results are reported in table 1.4 for three distinct samples: women and men (panel A), women only (panel B) and men only (panel C). Each column displays the estimates for a particular school leaving age. Column 1 reports the results of the global 4th order polynomial regression whereas column 2 refers to the triangular kernel weighted local linear regression (using the cross validation criterion for optimal bandwidth selection). Unsurprisingly, our estimates indicate that the 1967 Berthoin reform had a strong negative and significant impact on the fraction of individuals who were “caught up” by the new minimum school leaving age. Considering the global polynomial regression coefficients, the fraction of pupils who dropped out by the age of 14 and by the age 15 fell respectively by 9.4 and 9.2 percentage points, the effect being slightly larger for women (-10.4 and -10.5 percentage points) than for men (-8.3 and -7.8 percentage points). Our results also show that while most of the reform’s impact was concentrated on the 14-16 school leaving age group, it also induced a small fraction of individuals to prolong their studies *beyond* the new compulsory requirement: the frac-

²⁰Note that the instrumental variable strategy adopted in this study controls for potential biases arising from measurement error in the schooling variable.

Table 1.4: *Estimated impact of the French Berthoin reform (1967) on the distribution of school leaving ages of women and men. Source: Enquête Emploi (1990-2002).*

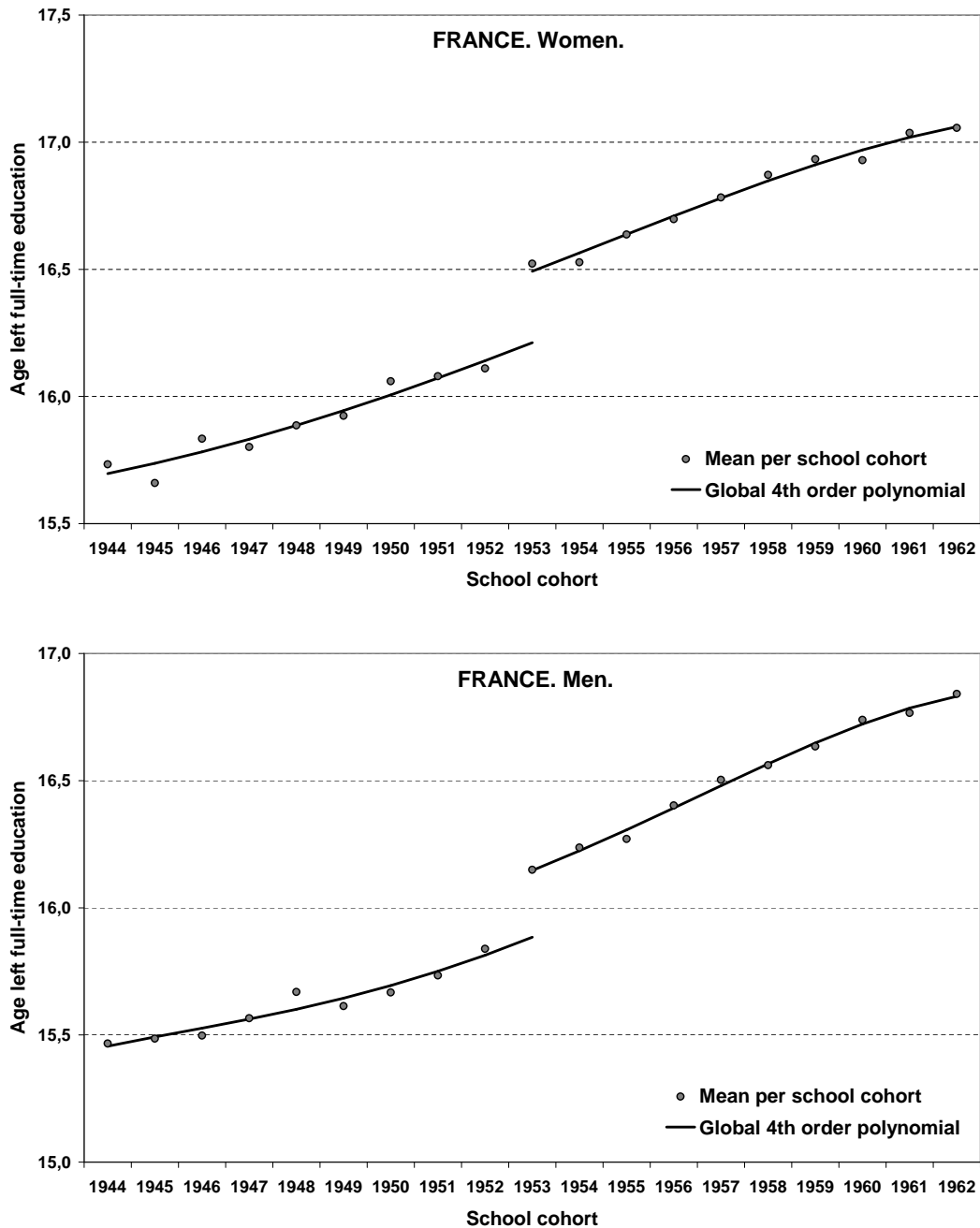
FRANCE – ENQUÊTE EMPLOI (1990-2002)		
	Global 4 th order Polynomial Regression (1)	Local Linear Regression (triangular kernel) (2)
<u>PANEL A: WOMEN AND MEN</u>		
% left school by age 13	-0.005** (0.002)	-0.004** (0.002)
% left school by age 14	-0.094*** (0.005)	-0.093*** (0.006)
% left school by age 15	-0.092*** (0.005)	-0.092*** (0.006)
% left school by age 16	-0.013*** (0.004)	-0.019*** (0.001)
% left school by age 17	-0.006 (0.007)	-0.015*** (0.004)
% left school by age 18	-0.009 (0.005)	-0.009 (0.005)
Number of observations	188,903	188,903
<u>PANEL B: WOMEN</u>		
% left school by age 13	-0.004* (0.002)	-0.004* (0.002)
% left school by age 14	-0.104*** (0.005)	-0.100*** (0.003)
% left school by age 15	-0.105*** (0.006)	-0.101*** (0.004)
% left school by age 16	-0.008 (0.006)	-0.013*** (0.003)
% left school by age 17	-0.007 (0.009)	-0.013** (0.006)
% left school by age 18	-0.012 (0.008)	-0.009 (0.008)
Number of observations	97,548	97,548
<u>PANEL C: MEN</u>		
% left school by age 13	-0.006** (0.002)	-0.005** (0.002)
% left school by age 14	-0.083*** (0.008)	-0.087*** (0.010)
% left school by age 15	-0.078*** (0.006)	-0.083*** (0.008)
% left school by age 16	-0.018*** (0.007)	-0.026*** (0.002)
% left school by age 17	-0.005 (0.008)	-0.017*** (0.004)
% left school by age 18	-0.007 (0.007)	-0.005 (0.006)
Number of observations	91,355	91,355

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are the fraction of individuals who left school by a given age. Each coefficient comes from a separate regression. Coefficients reported in column 1 are based on 4th order global polynomial regressions allowing for an intercept shift at the 1953 school cohort, whereas coefficients reported in column 2 are based on local linear regressions using a triangular kernel and an optimally selected bandwidth (using the cross validation criterion). All regressions are performed on a sample constructed from the 1990 to 2002 French *Enquête Emploi* surveys which includes individuals who were born in France, belong to school cohorts 1944 to 1962 and were aged between 25 and 60 when surveyed. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

tion of pupils dropping out by the age of 16 fell by a significant 1.3 percentage point, this phenomenon being driven essentially by the male population (the coefficient is small and insignificant for women). Interestingly, our estimates seem to indicate that the Berthoin reform's "spillover" effect did not extend beyond secondary schooling: the coefficients become close to zero when we consider school leaving ages of 18 or more.

In light of these results and in order to maximize the statistical power of our estimation strategy, we choose to focus on the group of individuals whose schooling was effectively impacted by the raising of the minimum school leaving age, i.e. those who left school by the age of 18, representing 71% of the initial sample. Because the returns to compulsory schooling are what we are ultimately interested in, we further restrict the sample to workers with a positive wage. To assess the reform's impact on the average years of education within this group, we start by providing some graphical evidence on the "jump" induced by the addition of two years to the minimum school leaving age. In figure 1.3, the dots show the average age left full-time education for each school cohort between 1944 and 1962, computed separately for women (top graph) and men (bottom graph). The solid lines show the fitted values from a global 4th order polynomial allowing for an intercept shift at the 1953 school cohort. Visual inspection suggests that the average school leaving age of men and women follows an almost linear trend and exhibits a substantial "jump" around the 1953 cohort. To precisely quantify this discontinuity, we present in table 1.5 the results of parametric and non-parametric RDD stimulations, performed separately by gender. Column 1 reports the coefficient based on a 4th order global polynomial regression, whereas column 4 shows the coefficient from a local linear regression. The table indicates that the raising of the minimum school leaving age from 14 to 16 had a clear positive impact on average educational attainment. Our estimates are fairly stable across specifications and are similar for women and men. Using the polynomial fit, the Berthoin reform is estimated to have increased the average number of years of education by 0.281 (0.039) for women and by 0.262 (0.026) for men. The corresponding local linear regression estimates are 0.317 (0.029) and 0.267 (0.034) respectively.

Figure 1.3: *Impact of the French 1967 Berthoin reform on the average age left full-time education, calculated for women and men separately. Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).*



Notes: The dots show the average age left full-time education grouped at the school cohort cell for the subsample of French female (top graph) and male (bottom graph) workers born between 1944 and 1962, who left school between by the age of 18 and have a positive hourly wage. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1953 school cohort.

Table 1.5: *Estimated impact of the French Berthoin reform (1967) on schooling and hourly wages (in 2005 euros) of female and male workers. Source: Enquête Emploi (1990-2002).*

FRANCE – ENQUÊTE EMPLOI (1990-2002)							
	Global 4 th order Polynomial Regression			Local Linear Regression (triangular kernel)			Number of observations
	First Stage (dependent var.: age left full-time education) (1)	Reduced form (dependent var: log of hourly earnings) (2)	TSLs (dependent var: log of hourly earnings) (3)	First Stage (dependent var: age left full-time education) (4)	Reduced form (dependent var: log of hourly earnings) (5)	TSLs (dependent var: log of hourly earnings) (6)	
<u>PANEL A: WOMEN</u>	0.281*** (0.039)	0.000 (0.006)	0.002 (0.022)	0.317*** (0.029)	0.001 (0.005)	0.003 (0.016)	36,427
<u>PANEL B: MEN</u>	0.262*** (0.026)	0.004 (0.007)	0.016 (0.027)	0.267*** (0.034)	-0.003 (0.007)	-0.012 (0.028)	42,070

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are age left full-time education (columns 1 and 4) and log of hourly wage (columns 2, 3, 5 and 6). Each coefficient comes from a separate regression. Regressions in columns 1 to 3 include a 4th order polynomial in school cohort and an intercept shift at the 1953 school cohort. Coefficients in columns 4 to 6 come from local linear regressions using a triangular kernel, an optimally selected bandwidth and allowing for different slopes on both sides of the 1953 cutoff point. All regressions are performed on a sample constructed from the 1990 to 2002 French *Enquête Emploi* surveys which includes individuals who were born in France, belong to school cohorts 1944 to 1962, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Britain. The same analysis can be repeated for the UK 1973 Education Act. Remarkably, our results show that the British reform had almost the same impact on average educational attainment as its French counterpart.

The evolution of the distribution of school leaving ages across the 1949-1967 school cohorts is displayed in figure 1.4 and separate graphs for women and men are shown in figure 1.5. These graphs share many similarities with their French equivalents. First, they indicate that when the raising of the minimum school took place in Britain, the previous regime of compulsory attendance was binding for a large fraction of the population: about 30% of the pupils belonging to the 1957 school cohort left full-time education before the age of 16. Second, the 1973 Education Act did force some (but not all) individuals to stay one more year in school than they would have otherwise, a fraction that represents roughly one fourth of the 1958 school cohort. Again, the fact that between 8 and 10% of respondents belonging to post-reform cohorts declare having left school before the age of 16 can be interpreted as the combined effect of imperfect enforcement of compulsory schooling and measurement error in the schooling variable. Third, the raising of the minimum school leaving age to 16 in Britain appears to have had limited “spillover” effects. Table 1.6 shows that the bulk of the reform was essentially concentrated on the population of individuals who left school by the age of 15 – with a slightly larger impact for women (-24.4 percentage points) than for men (-21.8 percentage points) – and that its effect on the distribution of post-compulsory school leaving ages was negligible.

To ensure the comparability of our results for both countries, we focus on the sample of British workers who left school by the age of 18 (representing 82% of the initial sample) and have a positive wage. What was the impact of the raising of the minimum school leaving age to 16 on the average schooling of individuals belonging to this group? Figure 1.6 shows the evolution of the average age left full-time education for school cohorts 1949 to 1967, computed separately by gender, together with the fitted values from a global 4th order polynomial allowing for an intercept shift at the 1958 school cohort. The “jump” in the average school leaving age appears very clearly and is similar in magnitude for women and men. The results of the RDD estimations are displayed in columns 1 and 4 of table 1.7. The coefficients are fairly consistent

Figure 1.4: Evolution of the distribution of school leaving ages in the United Kingdom (women and men). Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).

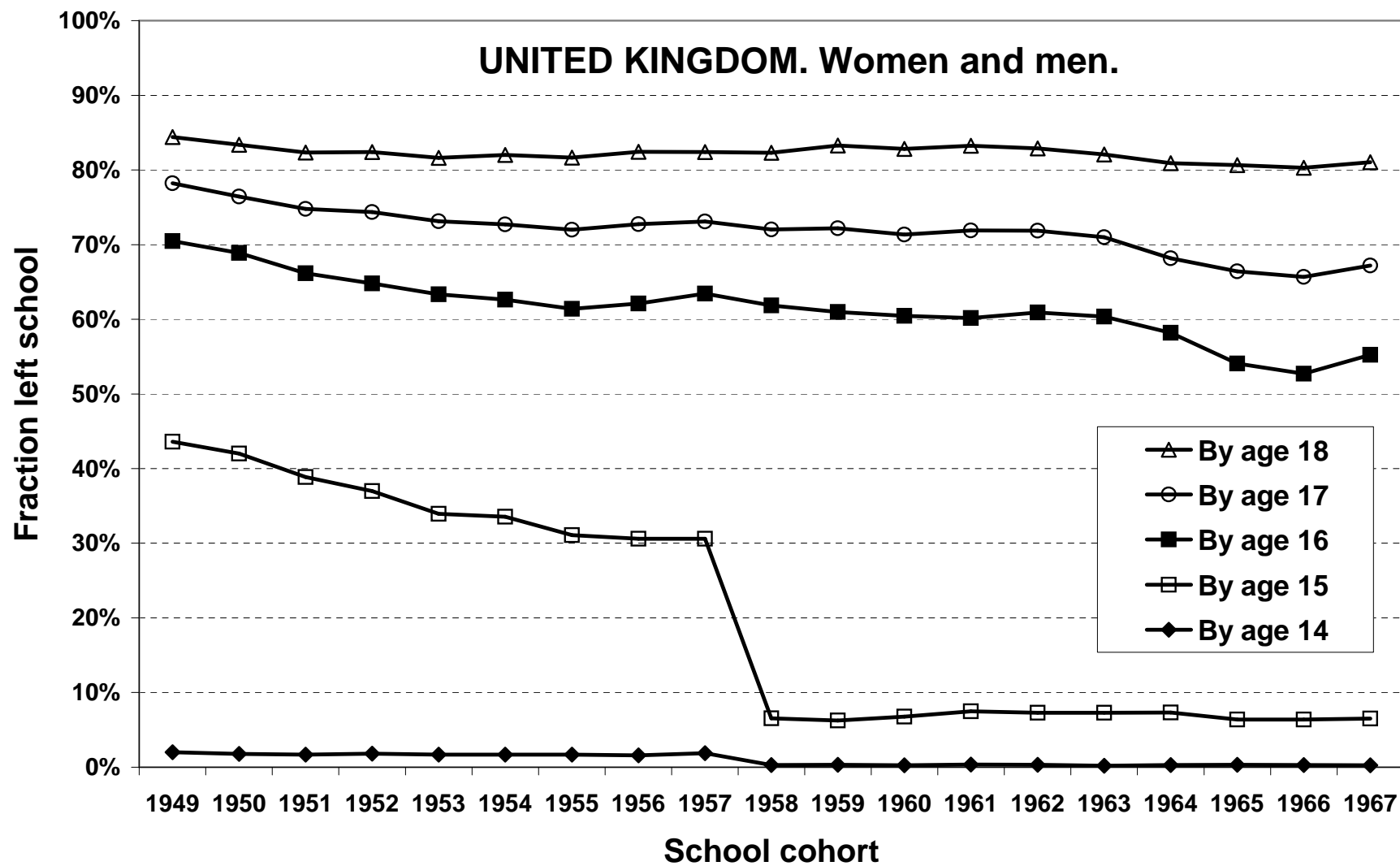


Figure 1.5: Evolution of the distribution of school leaving ages in the United Kingdom, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).

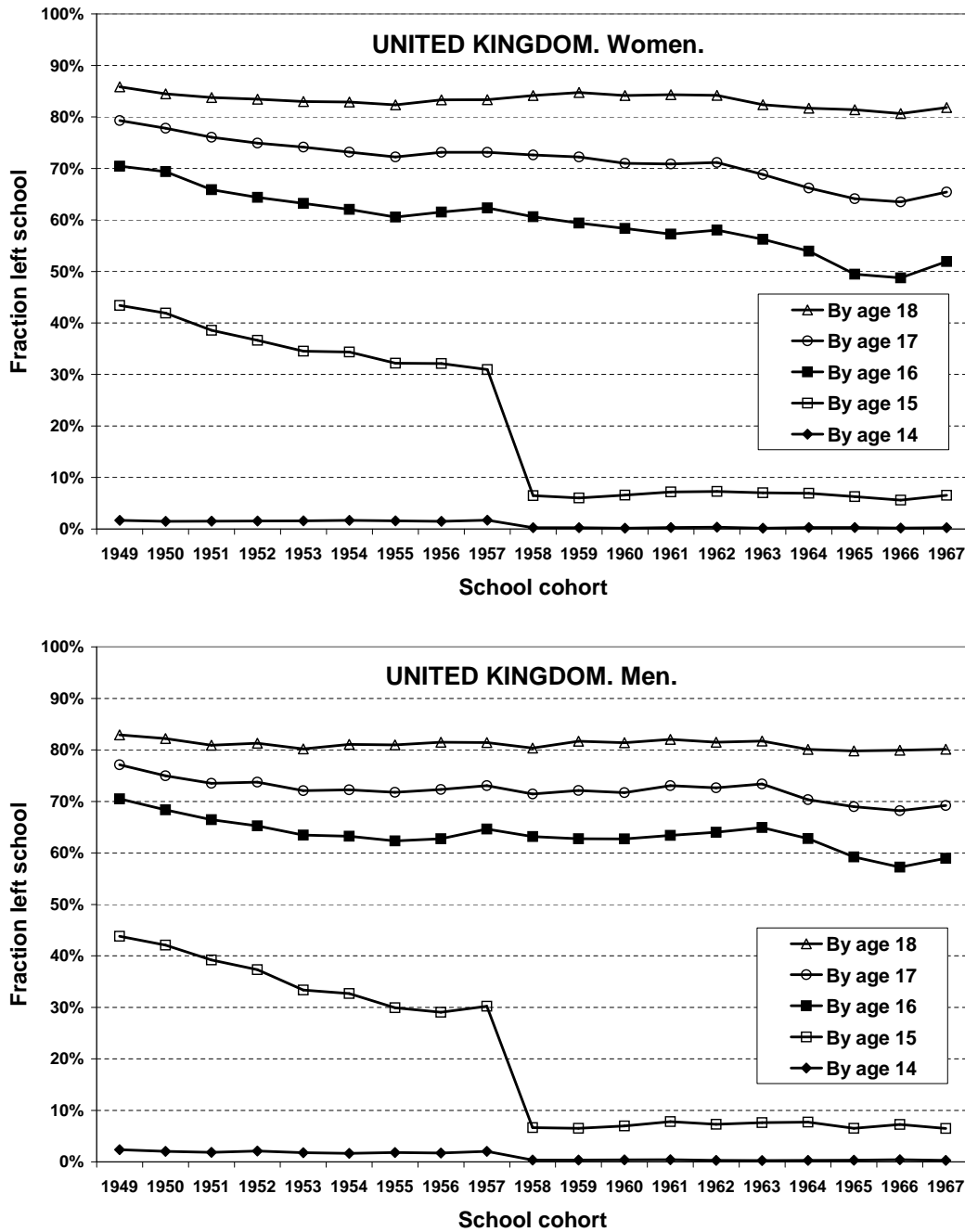
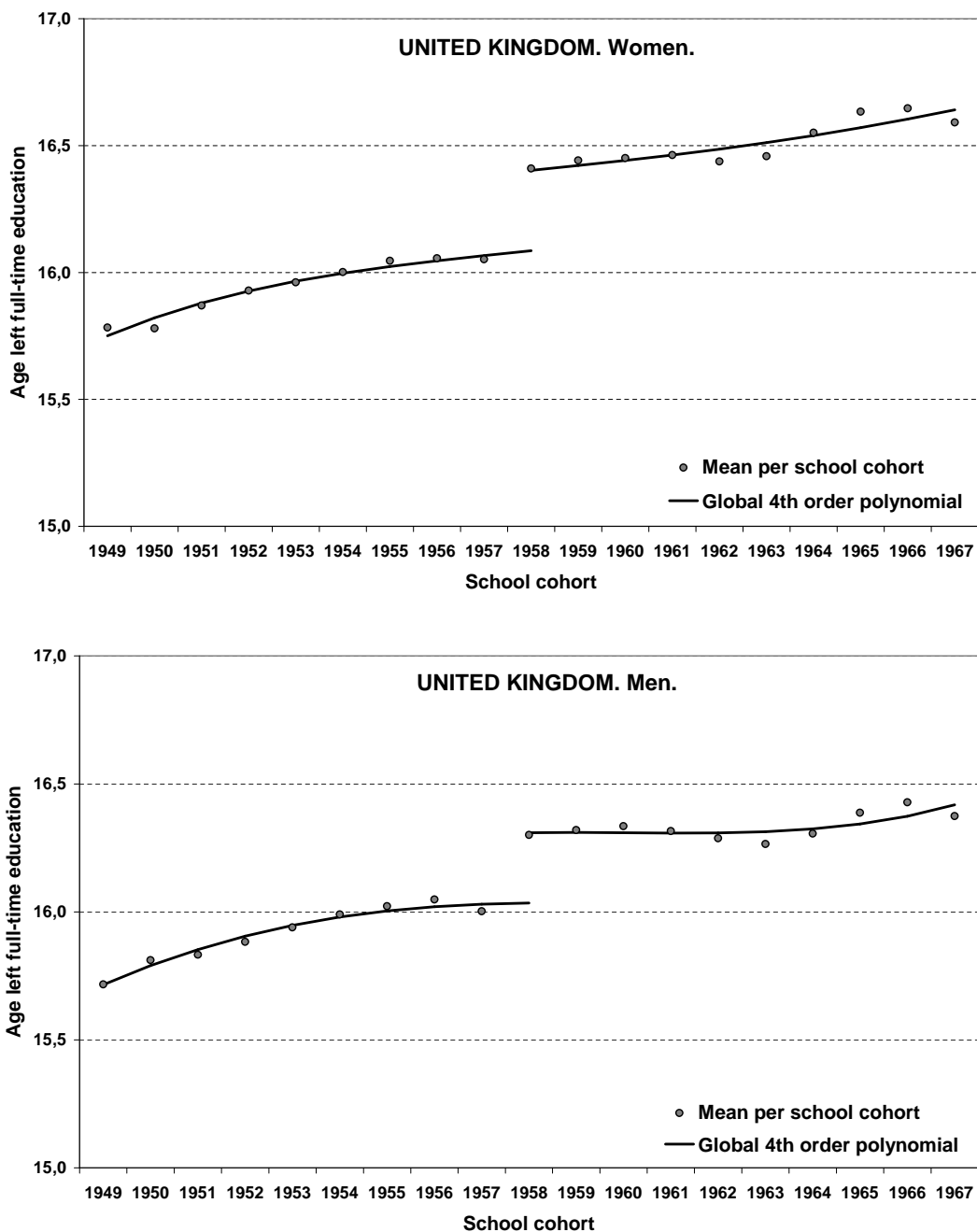


Table 1.6: *Estimated impact of the British Education Act (1973) on the distribution of school leaving ages of women and men. Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)		
	Global 4 th order Polynomial Regression (1)	Local Linear Regression (triangular kernel) (2)
<u>PANEL A: WOMEN AND MEN</u>		
% left school by age 13	-0.000 (0.000)	-0.000 (0.000)
% left school by age 14	-0.015*** (0.001)	-0.016*** (0.001)
% left school by age 15	-0.231*** (0.006)	-0.223*** (0.006)
% left school by age 16	-0.007 (0.009)	-0.017** (0.007)
% left school by age 17	-0.001 (0.006)	-0.009*** (0.002)
% left school by age 18	0.003 (0.006)	-0.000 (0.003)
Number of observations	292,934	292,934
<u>PANEL B: WOMEN</u>		
% left school by age 13	-0.000* (0.000)	-0.000*** (0.000)
% left school by age 14	-0.014*** (0.001)	-0.013*** (0.001)
% left school by age 15	-0.244*** (0.004)	-0.242*** (0.003)
% left school by age 16	-0.002 (0.010)	-0.017*** (0.006)
% left school by age 17	0.009 (0.006)	-0.002 (0.003)
% left school by age 18	0.015*** (0.004)	0.010*** (0.002)
Number of observations	151,885	151,885
<u>PANEL C: MEN</u>		
% left school by age 13	0.000 (0.000)	0.000 (0.000)
% left school by age 14	-0.017*** (0.002)	-0.017*** (0.002)
% left school by age 15	-0.218*** (0.011)	-0.223*** (0.011)
% left school by age 16	-0.012 (0.009)	-0.016* (0.009)
% left school by age 17	-0.012* (0.006)	-0.016*** (0.006)
% left school by age 18	-0.009 (0.007)	-0.011** (0.005)
Number of observations	141,049	141,049

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are the fraction of individuals who left school by a given age. Each coefficient comes from a separate regression. Coefficients reported in column 1 are based on 4th order global polynomial regressions allowing for an intercept shift at the 1958 school cohort, whereas coefficients reported in column 2 are based on local linear regressions using a triangular kernel and an optimally selected bandwidth (using the cross validation criterion). All regressions are performed on a sample constructed from the 1993 to 2004 British *Quarterly Labour Force Surveys* which includes individuals who were born in the UK, belong to school cohorts 1949 to 1967 and were aged between 25 and 60 when surveyed. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Figure 1.6: *Impact of the British 1973 Education Act on the average age left full-time education, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).*



Notes: The dots show the average age left full-time education grouped at the cohort cell for the subsample of British female (top graph) and male (bottom graph) workers born between 1949 and 1967 who left school between by the age of 18 and have a positive hourly wage. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1958 school cohort.

Table 1.7: *Estimated impact of the British Education Act (1973) on schooling and hourly wages (in 2005 pounds) of female and male workers. Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)							
	Global 4 th order Polynomial Regression			Local Linear Regression (triangular kernel)			Number of observations
	First Stage (dependent var.: age left full-time education) (1)	Reduced form (dependent var: log of hourly earnings) (2)	TSLs (dependent var: log of hourly earnings) (3)	First Stage (dependent var: age left full-time education) (4)	Reduced form (dependent var: log of hourly earnings) (5)	TSLs (dependent var: log of hourly earnings) (6)	
<u>PANEL A: WOMEN</u>	0.317*** (0.021)	0.014** (0.006)	0.044*** (0.017)	0.341*** (0.017)	0.023*** (0.003)	0.068*** (0.010)	70,360
<u>PANEL B: MEN</u>	0.274*** (0.028)	0.021*** (0.008)	0.077** (0.035)	0.283*** (0.031)	0.016*** (0.006)	0.056** (0.026)	64,799

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are age left full-time education (columns 1 and 4) and log of hourly wage (columns 2, 3, 5 and 6). Each coefficient comes from a separate regression. Regressions in columns 1 to 3 include a 4th order polynomial in school cohort and an intercept shift at the 1958 school cohort. Coefficients in columns 4 to 6 come from local linear regressions using a triangular kernel, an optimally selected bandwidth and allowing for different slopes on both sides of the 1958 cutoff point. All regressions are performed on a sample constructed from the 1993 to 2004 British *Quarterly Labour Force Surveys* which includes individuals who were born in the UK, belong to school cohorts 1949 to 1967, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

across the various specifications and differ little by gender, with a slightly larger effect for women. Using a polynomial fit, the 1973 Education Act is estimated to have increased the average number of years of education by 0.317 (0.021) for women and 0.274 (0.028) for men. The corresponding local linear regression estimates are 0.341 (0.017) and 0.283 (0.031) respectively.

Taken together, these results support the view that the Berthoin edict and the Education Act impacted the average schooling of French and British pupils in a very similar fashion. Both reforms were targeted towards the population of 14 to 15 year-old dropouts, had little spillover effects on post-compulsory schooling and raised by 3 to 4 months the average educational attainment of women and men who left school by the age of 18.

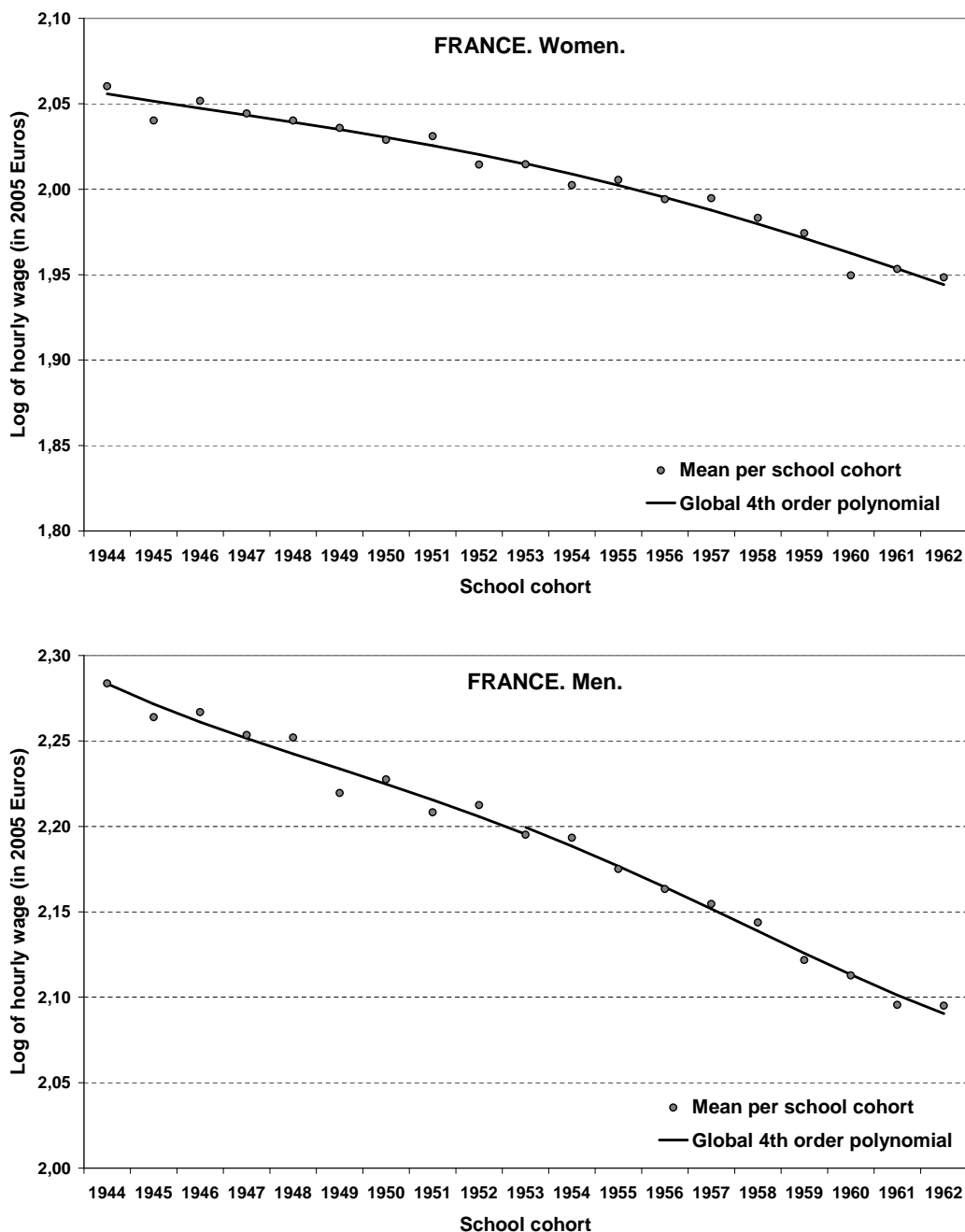
1.5.3 Impact on wages

Despite their similar effects on schooling, the French and British compulsory laws had a very different impact on hourly wages. While the wages of British women and men reacted positively to the introduction of the new minimum school leaving age, those of French men and women were left unchanged by the Berthoin reform.

France. Figure 1.7 shows the evolution of hourly wages across the 1944 to 1962 French school cohorts, separately for female and male workers. The dots indicate the average hourly wage (in 2005 euros) for each school cohort and the solid lines show the fitted values from a global 4th order polynomial allowing for an intercept shift at the 1953 school cohort. Both graphs share a very similar smooth decreasing pattern (due essentially to the effect of work experience) and show no obvious break in the series in 1953. This graphical evidence suggests that the monetary returns to the Berthoin reform were small and possibly zero.

The econometric analysis confirms this visual impression. Table 1.5 presents in columns 2 and 5 the reduced form estimates of the reform's impact on hourly wages. Both the parametric and non-parametric RDD estimates are statistically insignificant, which suggests very small effects from the change in compulsory schooling. Using a global polynomial fit, the coefficients on wages are 0.000 (0.006) for women and 0.004

Figure 1.7: *Impact of the French 1967 Berthoin reform on hourly wages (in 2005 euros), calculated for women and men separately. Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).*



Notes: The dots show the average log of hourly wage grouped at the school cohort cell for the subsample of French female (top graph) and male (bottom graph) workers born between 1944 and 1962 who left school by the age of 18 and have a positive hourly wage. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1953 school cohort.

(0.007) for men. The corresponding local linear regression estimates are 0.001 (0.005) and -0.003 (0.007).

Britain. The analysis of the evolution wages of British men and women offers quite a different picture. Figure 1.8 reveals that contrary to the French case, the average hourly wage (in 2005 British pounds) of female and male workers was positively impacted by the raising of the minimum school leaving age, the discontinuity being particularly obvious for women.

The estimates reported in columns 2 and 5 of table 1.7 are all positive and significant at the 1% level. Using a global polynomial fit, the coefficients on wages are 0.014 (0.006) for women and 0.021 (0.008) for men. The corresponding local linear regression estimates are 0.023 (0.003) and 0.016 (0.006).

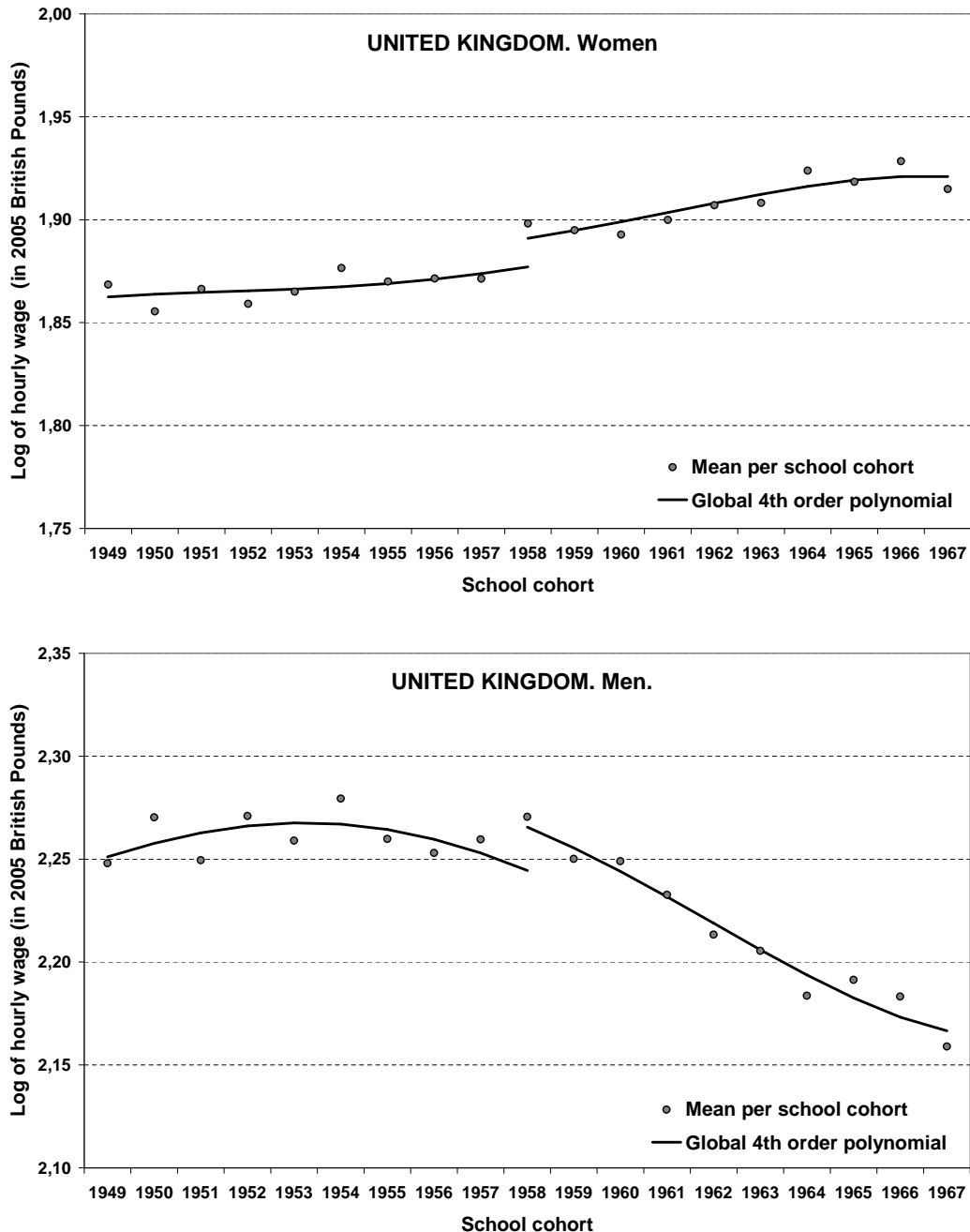
1.5.4 IV estimates of the returns to compulsory schooling

The first stage and reduced form regressions can be combined to estimate the returns to compulsory education in France and Britain within the TSLS approach described in section 1.3.3.

France. Columns 3 and 6 of table 1.5 report the TSLS estimates when schooling is instrumented by the minimum school leaving age faced by French respondents. In line with the reduced form estimates, the economic return to compulsory schooling is insignificant and close to zero for women as well as for men. The estimated returns from the global polynomial specification are 0.002 (0.022) for women and 0.016 (0.027) for men. They are respectively 0.003 (0.016) and -0.012 (0.028) when using local linear regression. These values are much lower than the previously reported naive OLS estimates of the returns to schooling in France (see table 1.3), respectively 0.083 and 0.072 for female and male workers.

Britain. The TSLS estimates of the returns to compulsory schooling in the United Kingdom are displayed in columns 3 and 6 of table 1.7. They are all positive and significant. The coefficients, in the range of 4 to 8%, are fairly similar for men and women, although the former are more precisely estimated (they are significant at

Figure 1.8: Impact of the British 1973 Education Act on hourly wages (in 2005 pounds), calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).



Notes: The dots show the average log of hourly wage grouped at the school cohort cell for the subsample of British female (top graph) and male (bottom graph) workers born between 1949 and 1967 who left school by the age of 18 and have a positive hourly wage. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1958 school cohort.

the 1% level for women but only at the 5% level for men). The estimated returns from the global polynomial specification are 0.044 (0.017) for female workers and 0.077 (0.035) for male workers. The corresponding local linear regression estimates are 0.068 (0.010) and 0.056 (0.026) respectively. These values are about half the size of the OLS estimates reported in table 1.3 (0.119 for women and 0.093 for men). Our estimates of the returns to compulsory schooling in the UK are lower than those found by Harmon and Walker (1995) and Oreopoulos (2006b, 2007), in the range of 12 to 15%, but are fairly similar to those reported by Devereux and Hart (2008) for the 1947 compulsory schooling law (in the 4 to 7% range).

Before turning to a more detailed discussion of how our results fit in the compulsory schooling literature, we perform a number of robustness checks to assess the stability of our estimates to alternative specifications.

1.5.5 Robustness checks

In this section, we show that our results are robust to the inclusion of age effects, the choice of bandwidth and to jumps at non-discontinuity points.

Inclusion of age effects. Because earnings typically follow a life-cycle pattern, one might wonder whether including age controls would affect our results. To address this issue, we add a quartic in age to the previously specified global polynomial and local linear regressions. We are able to disentangle age effects from school cohorts effects because we have several years of survey data in both the French and British case.

The estimated coefficients are reported in table 1.8 for French workers and British workers. This table shows that the reduced-form and TSLS estimates are barely impacted by the inclusion of age effects, as they appear very close to those previously reported in tables 1.5 and 1.7. The estimated returns to compulsory schooling for French female and male workers remain close to zero whereas they are significant and in the range of 4 to 8% in Britain.

Sensitivity to the choice of bandwidth. A second potential concern is that our local linear regression estimates could be sensitive to the choice of bandwidth. To

Table 1.8: *Inclusion of age effects in the estimation of the impact of the French Berthoin reform (1967) and of the British Education Act (1973) on schooling and hourly wages of female and male workers. Sources: Enquête Emploi (1990-2002) and Quarterly Labour Force Survey (1993-2004).*

	Global 4 th order Polynomial Regression			Local Linear Regression (triangular kernel)			Number of observations
	First Stage (dependent var.: age left full-time education) (1)	Reduced form (dependent var: log of hourly earnings) (2)	TSLs (dependent var: log of hourly earnings) (3)	First Stage (dependent var: age left full-time education) (4)	Reduced form (dependent var: log of hourly earnings) (5)	TSLs (dependent var: log of hourly earnings) (6)	
FRANCE – ENQUÊTE EMPLOI (1990-2002)							
<u>PANEL A: WOMEN</u>	0.283*** (0.038)	-0.001 (0.006)	-0.002 (0.023)	0.315*** (0.029)	-0.001 (0.006)	-0.003 (0.019)	36,427
<u>PANEL B: MEN</u>	0.261*** (0.026)	0.002 (0.007)	0.009 (0.025)	0.261*** (0.031)	-0.004 (0.006)	-0.013 (0.023)	42,070
UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)							
<u>PANEL A: WOMEN</u>	0.317*** (0.021)	0.014*** (0.005)	0.043*** (0.016)	0.341*** (0.018)	0.022*** (0.002)	0.065*** (0.008)	70,360
<u>PANEL B: MEN</u>	0.274*** (0.028)	0.021** (0.009)	0.078** (0.039)	0.283*** (0.031)	0.017** (0.007)	0.061** (0.031)	64,799
Age controls: Quartic in age	Yes	Yes	Yes	Yes	Yes	Yes	

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are age left full-time education (columns 1 and 4) and log of hourly wage (columns 2, 3, 5 and 6). Each coefficient comes from a separate regression. Regressions in columns 1 to 3 include a 4th order polynomial in school cohort and an intercept shift at the 1953 school cohort. Coefficients in columns 4 to 6 come from local linear regressions using a triangular kernel, an optimally selected bandwidth and allowing for different slopes on both sides of the 1953 cutoff point. All regressions include a quartic in age. Regressions in the top panel of the table are performed on a sample constructed from the 1990 to 2002 French *Enquête Emploi* surveys which includes individuals who were born in France, belong to school cohorts 1944 to 1962, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Regressions in the bottom panel of the table are performed on a sample constructed from the 1993 to 2004 British *Quarterly Labour Force Surveys* which includes individuals who were born in the UK, belong to school cohorts 1949 to 1967, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

address this issue, we present in table 1.9 (France) and table 1.10 (Britain) the reduced form and TSLS estimates obtained when changing the value of the bandwidth used in the local linear regressions from 4 to 10. The reported coefficients indicate that although the estimated effects exhibit some variation across bandwidth values, they are fairly stable. Moreover, the returns to compulsory schooling are never significant for French female and male workers but are always positive and significant for their British counterparts (the only exception being for males when the bandwidth is equal to 4). These results suggest that our results are robust to the choice of bandwidth.

Testing for jumps at non-discontinuity points Our final robustness check involves estimating jumps at points where there should be no jumps. We follow the guidelines set by Imbens and Lemieux (2007) and test for jumps in average educational attainment and hourly wages at the median of the two subsamples of pre- and post-reform school cohorts in each country. These placebo cutoff points are 1948 and 1958 in the French case and 1953 and 1963 in the British case.

The reduced form estimates for each of these cutoff points are displayed in tables 1.11 (France) and 1.12 (Britain), separately for female and male workers. None of the reported coefficients is significantly different from zero, which indicates that there are no significant jumps at the non-discontinuity points. We are therefore confident that the RDD estimations that we performed earlier are picking up the causal impact of the Berthoin reform and the Education Act on individual schooling and earnings in France and Britain.

1.6 Interpretation

How do our results fit in the compulsory schooling literature? As explained in section 1.1, the standard view that raising the minimum school leaving age systematically improves the labor market prospects of early school dropouts has been challenged by a number of recent papers which found low and sometimes zero returns to compulsory schooling laws (Meghir and Pålme, 2005; Oosterbeek and Webbink, 2007; Pischke and von Wachter, 2008; Devereux and Hart, 2008). Hence, the finding that the returns to compulsory schooling were zero in France should not in itself be

Table 1.9: *Sensitivity to the choice of bandwidth: estimated impact of the French Berthoin reform (1967) on schooling and hourly wages (in 2005 euros) of female and male workers using local linear regression with different bandwidth values. Source: Enquête Emploi (1990-2002).*

FRANCE – ENQUÊTE EMPLOI (1990-2002)			
	Local Linear Regression (triangular kernel)		
	First Stage	Reduced form	TSLS
	(dependent var: age left full-time education)	(dependent var: log of hourly earnings)	(dependent var: log of hourly earnings)
	(1)	(2)	(3)
<u>PANEL A: WOMEN</u>			
Bandwidth=4	0.371*** (0.019)	0.006 (0.005)	0.015 (0.013)
Bandwidth=5	0.334*** (0.031)	0.002 (0.005)	0.007 (0.014)
Bandwidth=6	0.317*** (0.029)	0.001 (0.005)	0.003 (0.016)
Bandwidth=7	0.304*** (0.030)	0.000 (0.005)	0.001 (0.017)
Bandwidth=8	0.309*** (0.028)	0.001 (0.005)	0.004 (0.016)
Bandwidth=9	0.308*** (0.028)	0.001 (0.005)	0.003 (0.017)
Bandwidth=10	0.314*** (0.026)	0.001 (0.005)	0.003 (0.016)
Number of observations	36,427	36,427	36,427
<u>PANEL B: MEN</u>			
Bandwidth=4	0.225*** (0.010)	-0.007 (0.006)	-0.032 (0.029)
Bandwidth=5	0.235*** (0.016)	-0.008 (0.005)	-0.033 (0.023)
Bandwidth=6	0.267*** (0.034)	-0.003 (0.007)	-0.012 (0.028)
Bandwidth=7	0.281*** (0.033)	0.000 (0.007)	-0.001 (0.025)
Bandwidth=8	0.286*** (0.031)	0.002 (0.007)	0.007 (0.023)
Bandwidth=9	0.291*** (0.030)	0.003 (0.006)	0.010 (0.021)
Bandwidth=10	0.298*** (0.030)	0.003 (0.006)	0.011 (0.020)
Number of observations	42,070	42,070	42,070

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are age left full-time education (column 1) and log of hourly wage (columns 2 and 3). Each coefficient comes from a separate triangular kernel weighted linear regression which is estimated for different values of the bandwidth. All regressions are performed on a sample constructed from the 1990 to 2002 French *Enquête Emploi* surveys which includes individuals who were born in France, belong to school cohorts 1944 to 1962, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Table 1.10: *Sensitivity to the choice of bandwidth: estimated impact of the British Education Act (1973) on schooling and hourly wages (in 2005 pounds) of female and male workers using local linear regression with different bandwidth values. Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)			
	Local Linear Regression (triangular kernel)		
	First Stage	Reduced form	TSLS
	(dependent var: age left full-time education)	(dependent var: log of hourly earnings)	(dependent var: log of hourly earnings)
	(1)	(2)	(3)
<u>PANEL A: WOMEN</u>			
Bandwidth=4	0.360*** (0.007)	0.021*** (0.001)	0.059*** (0.004)
Bandwidth=5	0.350*** (0.014)	0.023*** (0.003)	0.067*** (0.010)
Bandwidth=6	0.341*** (0.018)	0.022*** (0.002)	0.065*** (0.008)
Bandwidth=7	0.330*** (0.018)	0.020*** (0.002)	0.061*** (0.007)
Bandwidth=8	0.318*** (0.019)	0.020*** (0.002)	0.062*** (0.007)
Bandwidth=9	0.301*** (0.022)	0.019*** (0.002)	0.064*** (0.008)
Bandwidth=10	0.292*** (0.024)	0.020*** (0.002)	0.069*** (0.009)
Number of observations	70,360	70,360	70,360
<u>PANEL B: MEN</u>			
Bandwidth=4	0.315*** (0.018)	0.007 (0.005)	0.022 (0.017)
Bandwidth=5	0.297*** (0.024)	0.017** (0.008)	0.056* (0.032)
Bandwidth=6	0.283*** (0.031)	0.017** (0.007)	0.061** (0.031)
Bandwidth=7	0.266*** (0.036)	0.019*** (0.007)	0.071** (0.033)
Bandwidth=8	0.249*** (0.038)	0.017*** (0.006)	0.068** (0.032)
Bandwidth=9	0.236*** (0.037)	0.016*** (0.006)	0.069** (0.034)
Bandwidth=10	0.223*** (0.038)	0.015** (0.006)	0.069* (0.036)
Number of observations	64,799	64,799	64,799

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are age left full-time education (column 1) and log of hourly wage (columns 2 and 3). Each coefficient comes from a separate triangular kernel weighted linear regression which is estimated for different values of the bandwidth. All regressions are performed on a sample constructed from the 1993 to 2004 British *Quarterly Labour Force Surveys* which includes individuals who were born in the UK, belong to school cohorts 1949 to 1967, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Table 1.11: *Testing for jumps in schooling and hourly wages of French female and male workers at non-discontinuity points. Source: Enquête Emploi (1990-2002).*

FRANCE – ENQUÊTE EMPLOI (1990-2002)				
	Local Linear Regression (triangular kernel)			
	First Stage		Reduced form	
	(dependent var: age left full-time education)		(dependent var: log of hourly earnings)	
	Non-discontinuity points		Non-discontinuity points	
	1948	1958	1948	1958
	(1)	(2)	(3)	(4)
<u>PANEL A: WOMEN</u>	0.018	0.010	-0.002	-0.006
	(0.049)	(0.008)	(0.004)	(0.004)
Number of observations	16,991	19,436	16,991	19,436
<u>PANEL B: MEN</u>	0.037	-0.021	-0.001	-0.001
	(0.038)	(0.013)	(0.007)	(0.004)
Number of observations	18,702	23,368	18,702	23,368
School cohorts included in the subsample	1944-1952	1953-1962	1944-1952	1953-1962

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. Columns 1 and 2 are the reduced form estimates of the jumps in age left full-time education at particular non-discontinuity points (school cohorts 1948 and 1958), estimated using a local linear regression based on a triangular kernel and an optimally selected bandwidth. Columns 3 and 4 are the reduced form estimates of the corresponding jumps in the log of hourly wages. The regressions are performed separately on the subsample of French workers who faced the pre-reform minimum school leaving age of 14 (school cohorts 1944-1952) and on the subsample of workers who faced the post-reform minimum school leaving age of 16 (school cohorts 1953-1962). These subsamples are constructed from the sample of respondents to the 1990 to 2002 French *Enquête Emploi* surveys who were born in France, belong to school cohorts 1944 to 1962, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Table 1.12: *Testing for jumps in schooling and hourly wages of British female and male workers at non-discontinuity points. Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)				
	Local Linear Regression (triangular kernel)			
	First Stage (dependent var: age left full-time education) Non-discontinuity points		Reduced form (dependent var: log of hourly earnings) Non-discontinuity points	
	1953 (1)	1963 (2)	1953 (3)	1963 (4)
<u>PANEL A: WOMEN</u>	-0.012 (0.016)	0.037 (0.025)	0.009 (0.006)	-0.001 (0.005)
Number of observations	31,414	38,946	31,414	38,946
<u>PANEL B: MEN</u>	0.022 (0.016)	-0.005 (0.005)	-0.002 (0.010)	0.002 (0.007)
Number of observations	27,754	37,045	27,754	37,045
School cohorts included in the subsample	1949-1957	1958-1967	1949-1957	1958-1967

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. Columns 1 and 2 are the reduced form estimates of the jumps in age left full-time education at particular non-discontinuity points (school cohorts 1953 and 1963), estimated using a local linear regression based on a triangular kernel and an optimally selected bandwidth. Columns 3 and 4 are the reduced form estimates of the corresponding jumps in the log of hourly wages. The regressions are performed separately on the subsample of British workers who faced the pre-reform minimum school leaving age of 15 (school cohorts 1949-1957) and on the subsample of workers who faced the post-reform minimum school leaving age of 16 (school cohorts 1958-1967). These subsamples are constructed from the sample of respondents to the 1993 to 2004 British *Quarterly Labour Force Surveys* who were born in the UK, belong to school cohorts 1949 to 1967, left school by the age of 18, were aged between 25 and 60 when surveyed and have a positive hourly wage. To avoid using the same individual several times, we include only respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

viewed as a completely atypical result, especially in the European context. What is more surprising is that the British equivalent to the French Berthoin reform had a significant positive impact on earnings, albeit sharing the same timing, being targeted towards the same population and operating in a very similar educational context.

In light of the many common features shared by the two reforms studied in this chapter, we believe that none of the existing interpretations of the zero returns to compulsory education can account in a satisfactory manner for the observed discrepancy between France and Britain. Instead, we provide evidence supporting an interpretation based on the role of academic credentials.

1.6.1 Comparison with existing estimates

Before turning to the interpretation of our findings, we examine whether our estimates are consistent with existing evidence on the returns to compulsory schooling in France and the United Kingdom.

To the best of our knowledge, the French Berthoin reform has never been used to estimate the returns to compulsory education. Our study is therefore the first to any provide empirical evidence on this matter, so our results cannot be compared with previous assessments. The British case, by contrast, has been extensively studied in the literature (Harmon and Walker, 1995; Oreopoulos, 2006a, 2007; Devereux and Hart, 2008). However, the comparison of our results with previous estimates is not straightforward, for two reasons. First, most authors focus on the 1947 Butler Act, which raised the minimum school leaving age from 14 to 15, rather than on the 1973 Education Act. Indeed, a number of characteristics of the 1947 law made it a good candidate for identification, especially before RDD methods became popular among researchers: the large fraction of the population impacted by the reform; the availability of large data sets containing information on individual earnings for the relevant cohorts; and finally, the possibility of constructing a control group using Northern Ireland or the Republic of Ireland. Second, the few studies that considered the 1973 Education Act (Harmon and Walker, 1995; Oreopoulos, 2007) did not focus on this law specifically, but rather included it in a set of instruments formed of other changes in the minimum school leaving age. The reported estimates are usually very

high. Using in a first-difference approach, Harmon and Walker (1995) find that the 1947 and 1973 minimum school leaving age laws in Great Britain yielded a return of 15% for men; using a difference-in-difference methodology that takes advantage of the different timing of compulsory schooling laws in Great Britain, Northern Ireland and the Republic of Ireland, Oreopoulos (2007) finds a an average increase of 12% in annual earnings per additional year of schooling; in another paper (Oreopoulos, 2006a), the same author applies the RDD methodology to study 1947 Butler Act alone and finds a return of 15% for all workers. These estimates are about twice the size of those we find using the 1973 law only. However, because none of these estimates is derived exclusively from this compulsory schooling law, they are not directly comparable to ours. In particular, the fact that a much larger fraction of the population was impacted by the 1947 law change than by the subsequent 1973 reform could explain part of the discrepancy. Furthermore, a recent study by Devereux and Hart (2008) argues that the returns to the 1947 Butler Act were in fact lower than previously estimated: using the New Earnings Survey Panel Data-Set which contains precise and reliable information on individual wages, the authors find no evidence of any positive return for women and a return for men in the 4 to 7% range. Despite the fact that we study a later reform and that we find positive returns for both men and women, our RDD estimates are fairly consistent with those reported by Devereux and Hart.

More generally, our findings seem to fit fairly well in the overall picture that emerges from the literature on compulsory schooling laws. While most studies that use US or Canadian data report higher returns to schooling than the naive OLS estimates, papers carried on European data usually find smaller and possibly zero returns. Yet we feel than none of the existing interpretations of zero returns to compulsory schooling can convincingly account for the fact that the Berthoin reform failed in raising earnings while the British Education Act succeeded in doing so.

1.6.2 Discussion of potential explanations

In this section, we review several possible interpretations of our results and discuss the reasons that lead us to consider them irrelevant in the present case.

Local versus average returns to schooling? The standard explanation of cross-country variations in the estimated wage impact of compulsory schooling laws is that returns to schooling are fundamentally heterogenous across individuals. According to this interpretation, the size of the local average treatment effect that is measured using changes in the minimum school leaving age is a function of the characteristics of “compliers” (i.e. individuals who were forced to stay in school longer as a result of the law change), which are likely to differ from one reform to another. The most commonly cited sources of heterogeneity in the returns to schooling include credit constraints, parental guidance and cognitive abilities. Originally, this argument was put forward to account for the fact that in many US studies, IV estimates of the returns to schooling are found to be higher than OLS estimates (see Card, 1999). This phenomenon is commonly interpreted as showing that early school dropouts enjoy higher returns to schooling than the rest of the population. A number of authors argue that heterogeneity in the returns to schooling could also explain the observed discrepancy between US and European estimates. For instance, Devereux and Hart (2008) claim that a possible reason why Canadian or US compulsory schooling laws seem to yield stronger returns than the 1947 Butler Act in the UK is that the latter impacted a much larger fraction of the population than the former. In this case, Devereux and Hart’s estimates would be picking up something closer to the average treatment effect in the population, whereas US and Canadian estimates would reflect the high returns of a small number of compliers.

In our view, heterogenous returns to schooling cannot credibly explain our findings. First, the fact that our estimates for Britain are of the same order of magnitude as those reported by Devereux and Hart (2008) for the 1947 reform seems to indicate that modest returns can arise even when the fraction of compliers is relatively small (the proportion of individuals impacted by the 1973 law change was roughly 20% versus 50% for the 1947 reform). Second, the discrepancy between French and British estimates seems unlikely to be driven by the characteristics of compliers, since as shown in section 1.5.2, both laws raised the minimum school leaving age to 16 at about the same period and impacted the schooling of pupils belonging to the bottom 30% of the distribution of school leaving ages in each country. We are therefore

confident that our estimates for France and Britain are not picking up the returns to education of very different groups of individuals.

Wage rigidity. A second potential explanation for our findings pertains to the role of wage rigidity. This aspect has been relatively neglected in previous studies, except for the paper by Pischke and von Wachter (2008), who discuss several mechanisms through which wage setting institutions could explain the zero returns to compulsory schooling in Germany. In particular, German trade unions may have attempted to reduce inter-regional wage dispersion in the process of collective bargaining agreements, thus hindering the emergence of positive state level wage differentials for students enrolled in states which raised the minimum school leaving age earlier.

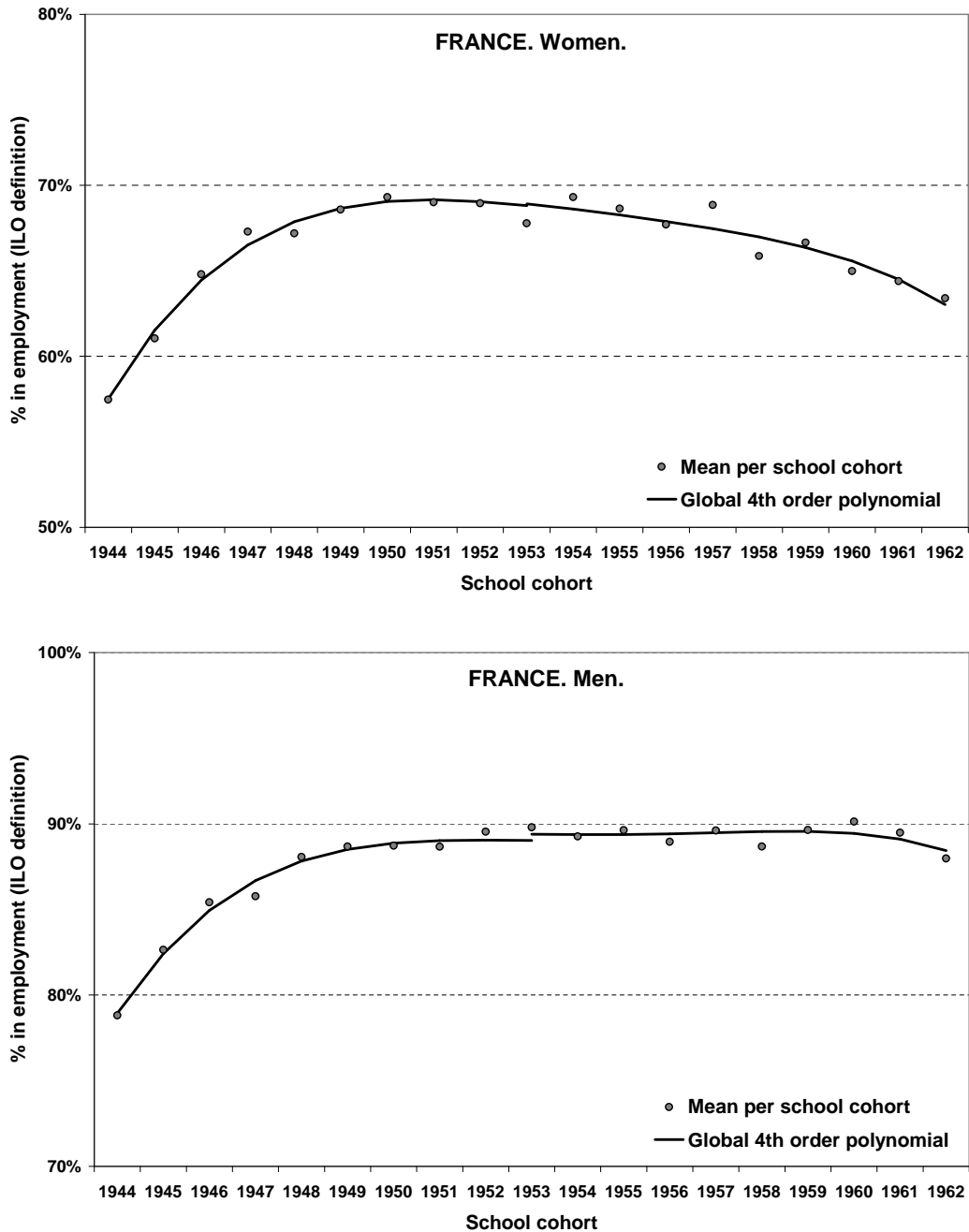
In the French and British contexts, this particular mechanism can be confidently ruled out, since compulsory schooling laws were implemented at the national level in both countries. However, the minimum wage constitutes another potentially important determinant of wage rigidity that should be seriously taken into consideration when comparing the returns to compulsory education in France and Britain. The reason why this labor market institution could play a key role is explaining our findings is that France has had a fairly high minimum wage ever since 1950²¹, whereas Britain did not introduce a national minimum wage before 1999 at a fairly low initial level, before gradually increasing it²². An explanation of the zero returns to the French Berthoin reform could therefore be that the minimum wage prevented the higher productivity of early school dropouts from being reflected in their earnings, these individuals being overrepresented at the bottom end of the wage distribution, where the minimum wage is known to have a large “bite”.

We explore this possibility by testing two implications of the wage rigidity hypothesis. First, if the minimum wage was the main explanation for the lack of monetary gains from increased compulsory education in France, we would expect the raising of the minimum school leaving age to have translated into higher employment rates for early school dropouts. Figure 1.9 investigates this issue by showing the evolution of

²¹the SMIG (*Salaires Minimum Interprofessionnel Garanti*) was created in 1950 and was replaced by the SMIC (*Salaires Minimum Interprofessionnel de Croissance*) in 1970.

²²In 2008, the total cost of the minimum wage is roughly 55% of the median wage in France versus 40% in the United Kingdom.

Figure 1.9: *Impact of the French 1967 Berthoin reform on the probability of being in employment, calculated for women and men separately. Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).*



Notes: The dots show the fraction of individuals in employment within in each school cohort cell for the subsample of French women (top graph) and men (bottom graph) born between 1944 and 1962 who left school by the age of 18. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1953 school cohort.

Table 1.13: *Estimated impact of the French Berthoin reform (1967) on the probability of being in employment and on the probability of being self-employed, for women and men separately. Source: Enquête Emploi (1990-2002).*

FRANCE – ENQUÊTE EMPLOI (1990-2002)				
	Global 4 th order Polynomial Regression		Local Linear Regression (triangular kernel)	
	Reduced form (dependent var: in employment) (1)	Reduced form (dependent var: self-employed) (2)	Reduced form (dependent var: in employment) (3)	Reduced form (dependent var: self-employed) (4)
<u>PANEL A: WOMEN</u>	0.001 (0.009)	-0.002 (0.007)	-0.009 (0.007)	-0.001 (0.007)
Number of observations	67,785	45,132	67,785	45,132
<u>PANEL B: MEN</u>	0.004 (0.005)	-0.000 (0.004)	0.000 (0.003)	0.002 (0.003)
Number of observations	65,774	57,921	65,774	57,921

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are an indicator that takes the value 1 if the individual is in employment (columns 1 and 3) and an indicator that takes the value 1 if the individual is self-employed (columns 2 and 4). Each coefficient comes from a separate regression. Regressions in columns 1 and 2 include a 4th order polynomial in school cohort and an intercept shift at the 1953 school cohort. Coefficients in columns 3 and 4 come from local linear regressions using a triangular kernel, an optimally selected bandwidth and allowing for different slopes on both sides of the 1953 cutoff point. All regressions are performed on a sample constructed from the 1990 to 2002 French *Enquête Emploi* surveys which includes individuals who were born in France, belong to school cohorts 1944 to 1962, left school by the age of 18 and were aged between 25 and 60 when surveyed. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

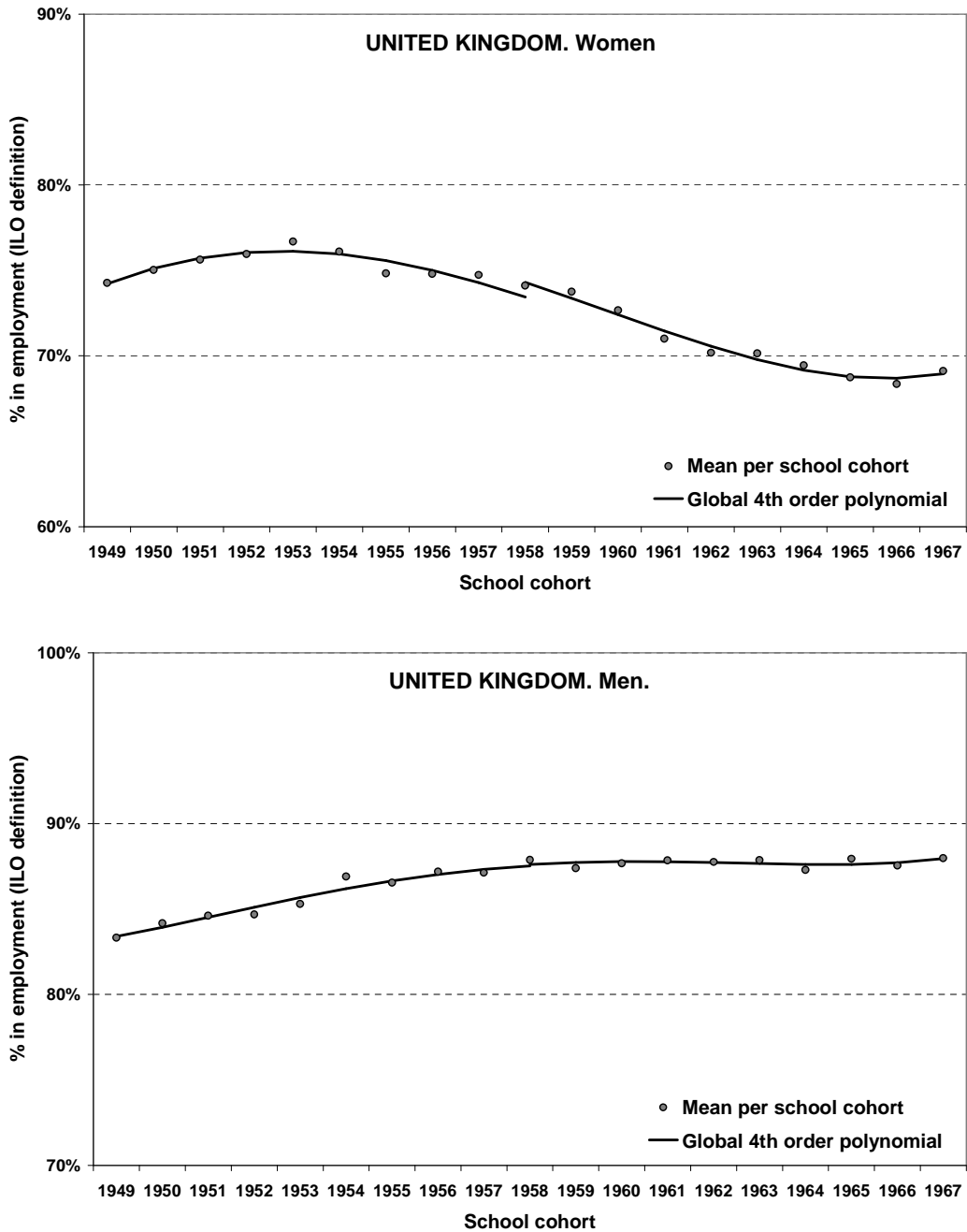
employment rates of French men and women across the 1944 to 1962 school cohorts. The two graphs show no evidence of “hidden” employment gains induced by the Berthoin reform, since the patterns of employment rates appear perfectly continuous for women men and men around the 1953 cutoff point. Columns 1 and 3 of table 1.13 display the corresponding reduced form RDD estimates, where the dependent variable is a dummy variable for whether the respondent is in employment or not. The coefficients are small and never significantly different from zero, which seems to rule out the existence of positive employment effects from the reform for both women and men. The same exercise is repeated for Britain and the estimates are displayed in figure 1.10 and table 1.14. Most coefficients are not significant and if anything, the results would tend to indicate that the British reform had a small positive impact on the probability of being in employment for women and a small negative impact on the probability of being self-employed for men. Hence, in contrast with the French situation, the raising of the minimum school leaving age in Britain could have induced marginal employment gains in addition to the previously estimated monetary gains.

A second implication of the wage rigidity argument is that the individuals whose schooling was positively affected by compulsory schooling laws would be more likely to become self-employed if their higher productivity was not rewarded in the wage and salary sector. Following Pischke and von Wachter (2008), we test this hypothesis by focusing on the evolution of the fraction of self-employed across school cohorts. Figures 1.11 and 1.12 display the graphical analysis for France and Britain and the corresponding reduced form RDD estimates are reported in columns 2 and 4 of table 1.13 and 1.14. Again, these results provide no support to the wage rigidity story, as both compulsory schooling reforms appear to have left self-employment rates unchanged.

1.6.3 Institutional features of schooling systems

The empirical evidence provided so far suggests that the explanation of the discrepancy between France and Britain lies probably outside the realms of heterogenous returns to schooling and wage-setting institutions. A rather natural question that arises then is whether our findings could be related to specific institutional features

Figure 1.10: *Impact of the British 1973 Education Act on the probability of being in employment, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).*



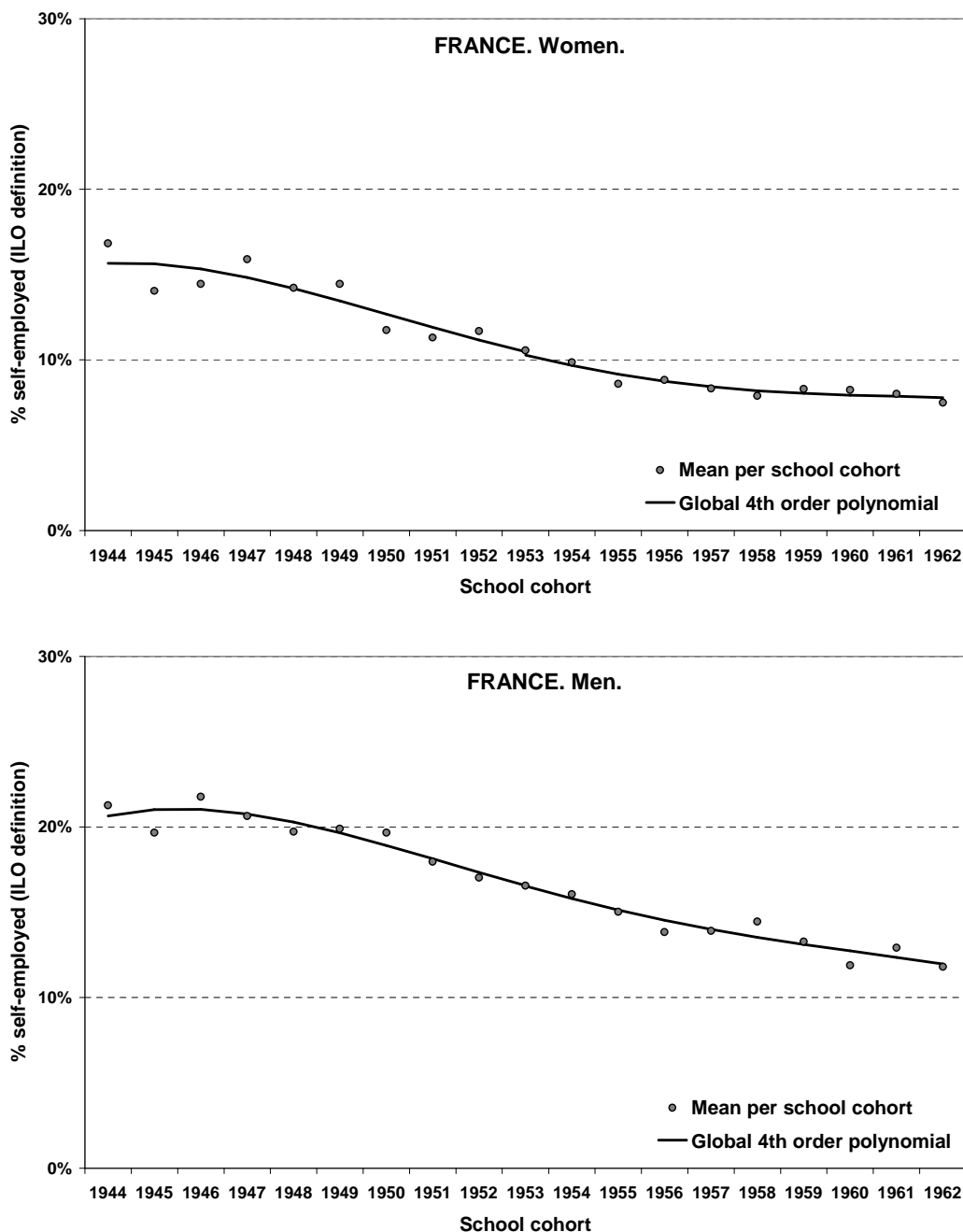
Notes: The dots show the fraction of individuals in employment within each school cohort cell for the subsample of British women (top graph) and men (bottom graph) born between 1949 and 1967 who left school by the age of 18. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1958 school cohort.

Table 1.14: *Estimated impact of the British Education Act (1973) on the probability of being in employment and on the probability of being self-employed, for women and men separately. Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)				
	Global 4 th order Polynomial Regression		Local Linear Regression (triangular kernel)	
	Reduced form (dependent var: in employment) (1)	Reduced form (dependent var: self-employed) (2)	Reduced form (dependent var: in employment) (3)	Reduced form (dependent var: self-employed) (4)
<u>PANEL A: WOMEN</u>	0.008* (0.004)	0.004 (0.003)	0.003 (0.003)	0.002 (0.003)
Number of observations	126,531	91,997	126,531	91,997
<u>PANEL B: MEN</u>	0.001 (0.003)	-0.006 (0.005)	0.001 (0.003)	-0.013*** (0.002)
Number of observations	114,630	99,929	114,630	99,429

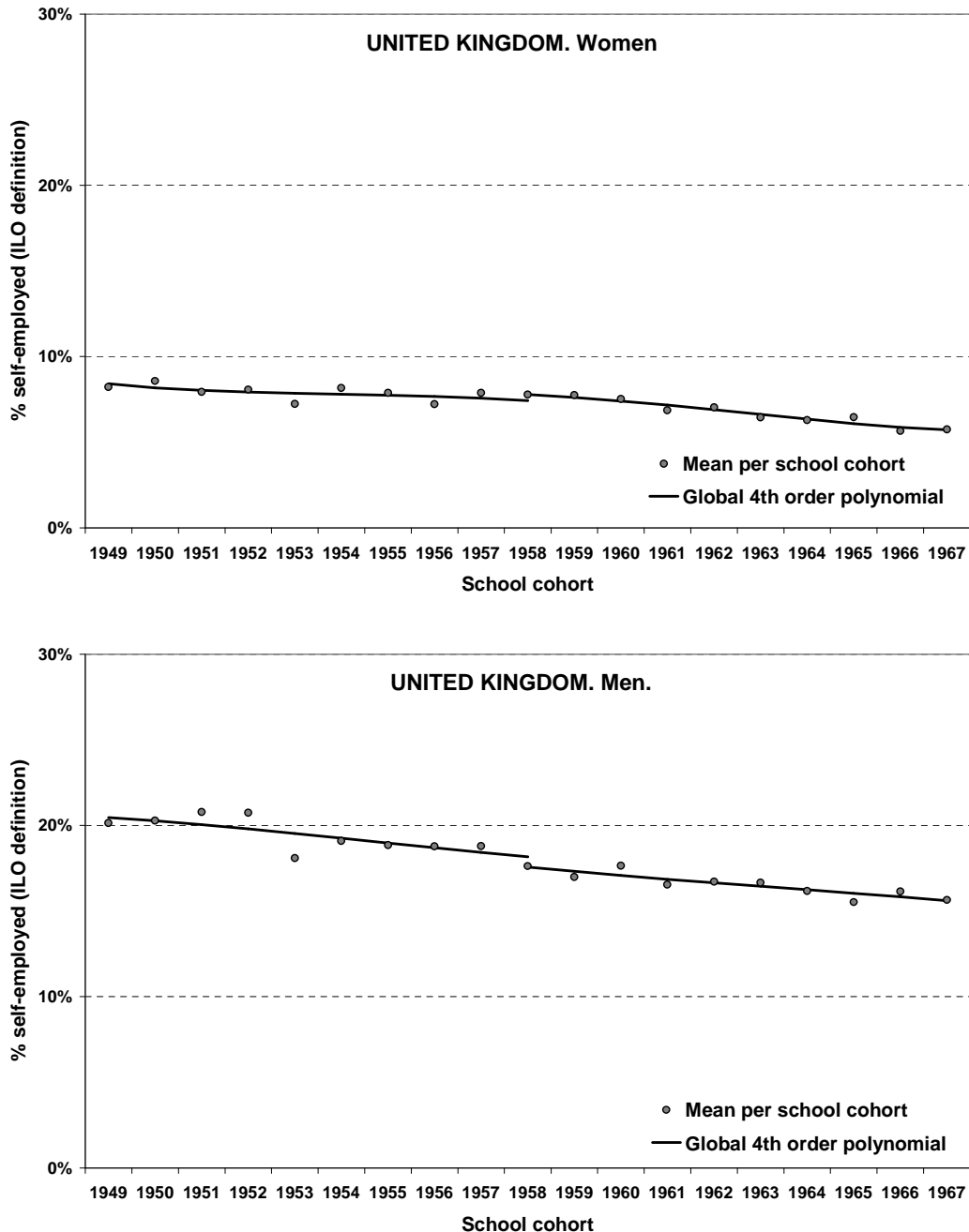
Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are an indicator that takes the value 1 if the individual is in employment (columns 1 and 3) and an indicator that takes the value 1 if the individual is self-employed (columns 2 and 4). Each coefficient comes from a separate regression. Regressions in columns 1 and 2 include a 4th order polynomial in school cohort and an intercept shift at the 1958 school cohort. Coefficients in columns 3 and 4 come from local linear regressions using a triangular kernel, an optimally selected bandwidth and allowing for different slopes on both sides of the 1958 cutoff point. All regressions are performed on a sample constructed from the 1993 to 2004 British *Quarterly Labour Force Surveys* which includes individuals who were born in the UK, belong to school cohorts 1949 to 1967, left school by the age of 18 and were aged between 25 and 60 when surveyed. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Figure 1.11: *Impact of the French 1967 Berthoin reform on the probability of being self-employed, calculated for women and men separately. Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).*



Notes: The dots show the fraction of self-employed individuals within each school cohort cell for the subsample of French male (top graph) and female (bottom graph) workers born between 1944 and 1962 who left school by the age of 18. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1953 school cohort.

Figure 1.12: *Impact of the British 1973 Education Act on the probability of being self-employed, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).*



Notes: The dots show the fraction of self-employed individuals within each school cohort cell for the subsample of British male (top graph) and female (bottom graph) workers between 1949 and 1967 who left school by the age of 18. The solid line represents the fitted values from a global 4th order polynomial regression with an intercept shift at the 1958 school cohort.

of French and British educational systems.

Institutional explanations of zero returns. In their discussion of the reasons why the returns to compulsory education were zero in Germany, Pischke and von Wachter (2008) present a few pieces of evidence supporting an explanation based on the particular features of the German school system. A first potential interpretation discussed by the authors relates to the organization of apprenticeship. In Germany, compulsory schooling does not necessarily end once secondary education is completed, since school leavers usually enter an apprenticeship. The lack of returns to the introduction of the 9th grade in secondary school may therefore simply reflect the fact that pupils who completed an apprenticeship received the same credential whether having left school at grade 8 or at grade 9. Pischke and von Wachter rule out this hypothesis by showing that the returns were also zero for individuals who did not complete an apprenticeship, and propose an alternative interpretation based on the content of German basic education. More precisely, they argue that because German pupils are tracked very early (i.e. after four years of primary schooling) into different types of schools which distinguish themselves by the academic content of the curriculum, basic labor market relevant skills such as reading, writing, and simple arithmetics are taught earlier than in countries which track their students at a later stage. Early tracking is interpreted by the authors as the key factor explaining why German pupils belonging to the lower end of the ability distribution benefited less from general schooling at the age 14 or 15 than US or UK pupils.

In the French case, none of these two institutional arguments seems likely to explain the zero returns result. First, in contrast to the German system, vocational training in France is almost completely integrated in the school system and very few pupils would take an apprenticeship without also gaining a vocational qualification involving general education courses. This means that a respondent who reports having left school at a given age is unlikely to have subsequently completed an apprenticeship which could have dissipated the returns attached to additional years of compulsory education. Second, the comparison of French and British educational systems suggests that the argument based on the particular teaching of skills induced by early

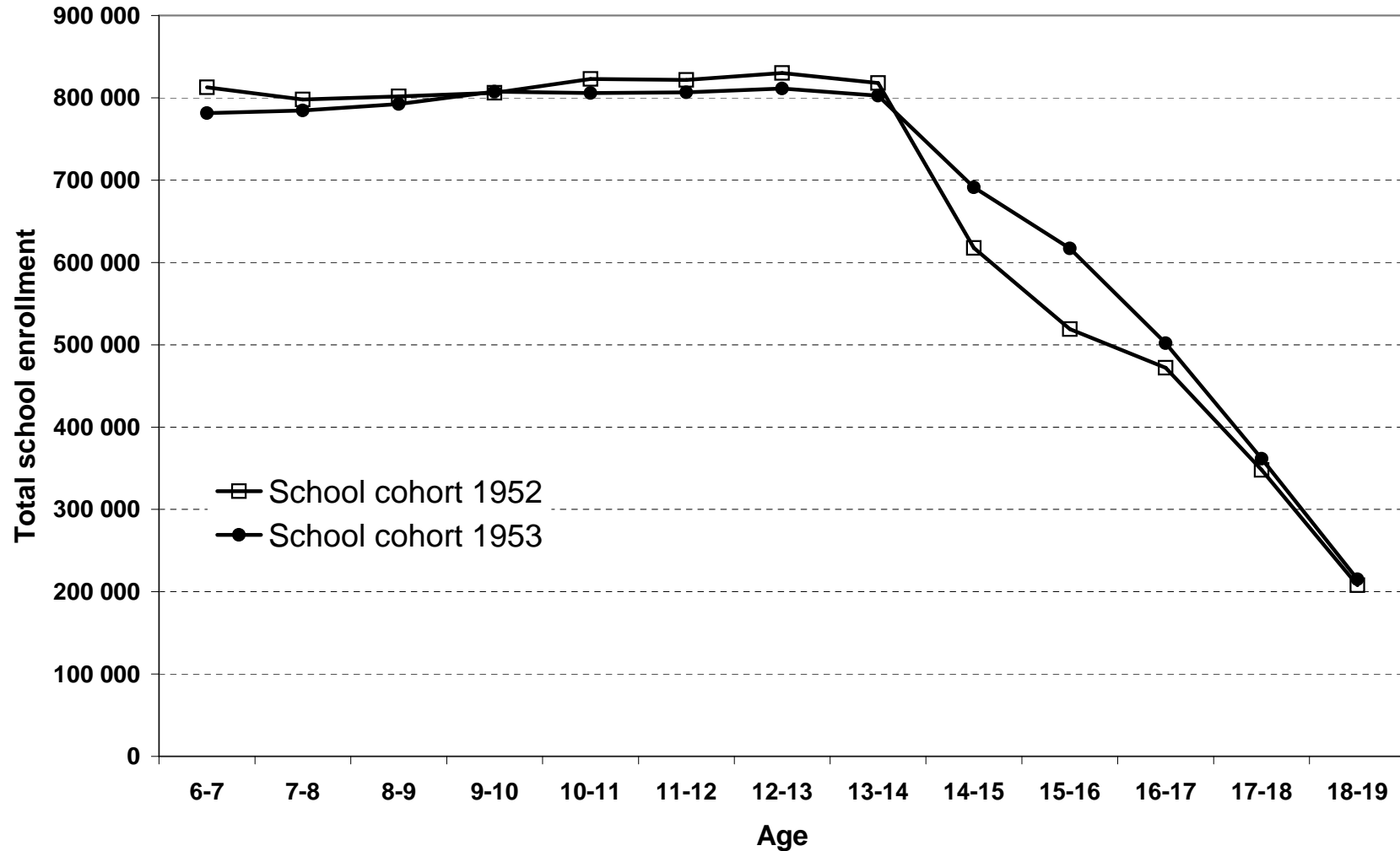
tracking in Germany cannot be easily transposed to explain the lack of positive returns to the Berthoin reform. This is because in the late 1960s, early 1970s, almost all French and British pupils followed the same school curriculum and pupil tracking into vocational training did not take place before the final years of junior secondary schooling. Furthermore, compulsory schooling laws were explicitly designed to allow for a unified and comprehensive curriculum of general education. Individuals who were induced by compulsory schooling laws to pursue their studies until the age of 16 were therefore intended to complete the acquisition of the skills considered relevant for their professional and civic lives. The comparable structure of school curriculums in France and Britain seems therefore incompatible with an explanation based on the timing of basic skills teaching.

A defective implementation of compulsory schooling? One could argue however that despite the apparent proximity of French and British compulsory schooling laws, the inability of the Berthoin reform to raise the labor market prospects of early school dropouts could simply be the result of a defective implementation. If, for instance, low achieving pupils who were compelled to stay in school longer were in fact kept in the same grade until they reached the new minimum school leaving age, there would be nothing surprising in finding no positive returns to these additional years of schooling. A careful assessment of what exactly happened to the cohorts surrounding the implementation of the Berthoin reform is therefore required to clarify the interpretation of our results.

Precise statistical information on the educational careers of cohorts born in 1952 and 1953 can be found in retrospective statistics released in 1969 and 1978 by the French Ministry of Education. Both cohorts counted a total of approximately 800,000 individuals. Figure 1.13 indicates, for each cohort, the number of pupils enrolled in school each year between the age of 6-7 and the age of 18-19. This graph shows that virtually all children from a given cohort were still in school at the age of 13-14²³. As from the age of 14-15, enrollment numbers drop sharply, falling to about 200,000 at the age of 18-19. The impact of the Berthoin reform is clearly perceptible.

²³The slight increase in the number of children at school between the age of 6-7 and the age of 13-14 is mainly a consequence of immigration.

Figure 1.13: Total school enrollment of French school cohorts 1952 and 1953 every year from the age of 6-7 to the age of 18-19. Source: Tableaux de l'Éducation Nationale, Ministère de l'Éducation nationale (1969, 1978).



Notes: This graph is constructed from the retrospective statistics released in 1969 and 1978 by the French Ministry of Education. It follows the total school enrollment between the age of 6-7 and the age of 18-19 of the two French school cohorts that surround the implementation of the Berthoin reform in 1967. Pupils born in 1952 faced a minimum school leaving age of 14 whereas those born in 1953 faced a minimum school leaving age of 16. For both cohorts, each dot represents total school enrollment (Y-axis) at a particular age (X-axis).

While 200,000 individuals born in 1952 dropped out at the age of 14, this number was reduced by one half when the new compulsory regime was implemented. Only 100,000 individuals born in 1953 dropped out of school at the age of 14, the other 100,000 being forced to wait two years before leaving full-time education at the age of 16. From then on, the two curves become very similar again.

What exactly did happen to the 100,000 pupils who were redirected to stay in school for two more years? To answer this question, table 1.15 compares the educational career of both cohorts. Total enrollment is broken up by the type of school attended, which includes preschooling or special education (for pupils with special needs), primary schooling, extended primary schooling, junior secondary schooling, vocational training and senior secondary schooling. Several conclusions can be drawn from this table. First, it appears that most pupils repeated at least once in primary school: by the end of elementary schooling, 60% were one year behind and 20% two years behind. Therefore, one should not take the theoretical educational progression by age to be a general rule: about half the students started junior secondary schooling at the age of 12, and not 11. Second, table 1.15 shows that for the last cohort facing a minimum school leaving age of 14, virtually all pupils completed primary schooling. Half of them attended extended primary schooling until they reached the age of 14 and dropped out of school or started vocational training, while the other half attended junior secondary schooling. 10% of junior secondary pupils dropped out before reaching the end of the 4-year curriculum, 50% went on to secondary education and 40% were tracked into vocational training. The corresponding figures for the subsequent cohort of pupils born in 1953, who faced the new minimum school leaving age of 16, clearly indicates that the French school system was ready to absorb the flow of about 100,000 pupils that were compelled to stay two additional years in school. The comparison between the educational situation of both cohorts at the age of 15-16 reveals that the vast majority of these 100,000 additional pupils attended junior and sometimes senior secondary schooling (about 60,000) or started vocational training (about 25,000), while the remaining 15,000 attended extended primary schooling or special education. The figures also indicate that compared to the previous cohort, pupils born in 1953 were much more likely to be enrolled in a junior secondary

Table 1.15: *Educational career of the 1952 and 1953 cohorts between the age of 6-7 and the age of 18-19. Source: Tableaux de l'Éducation Nationale, Ministère de l'Éducation nationale (1969, 1978).*

COHORT 1952													
School Year	Compulsory schooling												
	1958-59 Age 6-7	1959-60 7-8	1960-61 8-9	1961-62 9-10	1962-63 10-11	1963-64 11-12	1964-65 12-13	1965-66 13-14	1966-67 14-15	1967-68 15-16	1968-69 16-17	1969-70 17-18	1970-71 18-19
Total	812,729	797,980	801,669	806,087	822,865	821,655	830,235	817,842	617,796	519,208	472,124	347,702	207,945
Preschools or special education	39,201	5,716	8,639	10,916	13,413	14,328	14,643	13,706	7,867	5,129	8,867	2,187	
Primary schools	773,528	792,267	792,961	793,621	758,280	487,120	185,469	65,817	11,650				
Extended primary schools			69	1,550	6,261	65,993	221,731	322,233	81,837	9,085	1,589	176	
Junior secondary schools					44,911	254,214	408,392	416,086	378,298	243,822	81,885	8,838	1,096
Vocational training									117,175	155,910	181,440	116,970	43,051
Senior secondary schools									20,969	105,262	201,343	219,531	163,798

COHORT 1953													
School Year	Compulsory schooling												
	1959-60 Age 6-7	1960-61 7-8	1961-62 8-9	1962-63 9-10	1963-64 10-11	1964-65 11-12	1965-66 12-13	1966-67 13-14	1967-68 14-15	1968-69 15-16	1969-70 16-17	1970-71 17-18	1971-72 18-19
Total	781,443	784,555	792,400	807,630	805,676	806,672	811,369	802,469	691,275	617,169	502,013	361,345	215,157
Preschools or special education	20,276	5,605	8,694	11,967	14,303	15,894	16,123	15,088	10,503	10,942	6,743	2,511	
Primary schools	761,167	778,950	783,690	795,150	739,284	473,255	170,569	61,850	14,887				
Extended primary schools			16	513	5,579	52,078	183,029	273,014	92,456	17,522	1,651	50	
Junior secondary schools					46,510	265,445	441,648	452,362	424,606	296,660	84,966	7,942	830
Vocational training								155	127,231	181,001	193,832	123,088	44,954
Senior secondary schools									21,592	111,044	214,821	227,754	169,373

Notes: This table is constructed from the retrospective statistics released in 1969 and 1978 by the French Ministry of Education. It follows the educational careers of the two school cohorts that surround the implementation of the Berthoin reform in 1967. Pupils born in 1952 faced a minimum school leaving age of 14 whereas those born in 1953 faced a minimum school leaving age of 16. Total enrollment is broken up by type of school (lines) and school year (columns) between the age of 6-7 and the age of 18-19. School levels include preschooling or special education (pupils with special needs), primary schooling (*École élémentaire*, 5 grades from the age of 6 to the age of 11), extended primary schooling (*Classes de fin d'études*, 3 grades from the age of 11 to the age of 14), junior secondary schooling (*Collège*, 4 grades from the age of 11 to the age of 15), vocational training (*Lycée professionnel*, 3 grades from the age of 14 to the age of 17 or 2 grades from the age of 16 to the age of 18) and senior secondary schooling (*Lycée*, 3 grades from the age of 15 to the age of 18).

form the age of 12 onwards than to spend these years attending extended primary schooling. On the whole, this analysis suggests that when the Berthoin reform was implemented, the French schooling system had made the necessary investments to accommodate the sudden increase in total enrollment (which included the construction of new junior secondary schools) and that early school dropouts born after 1953 were not artificially kept in the same grade until they reached the new minimum school leaving age. It seems therefore difficult to attribute the lack of positive returns to compulsory schooling in France to a defective implementation of the Berthoin reform.

1.6.4 The role of academic credentials

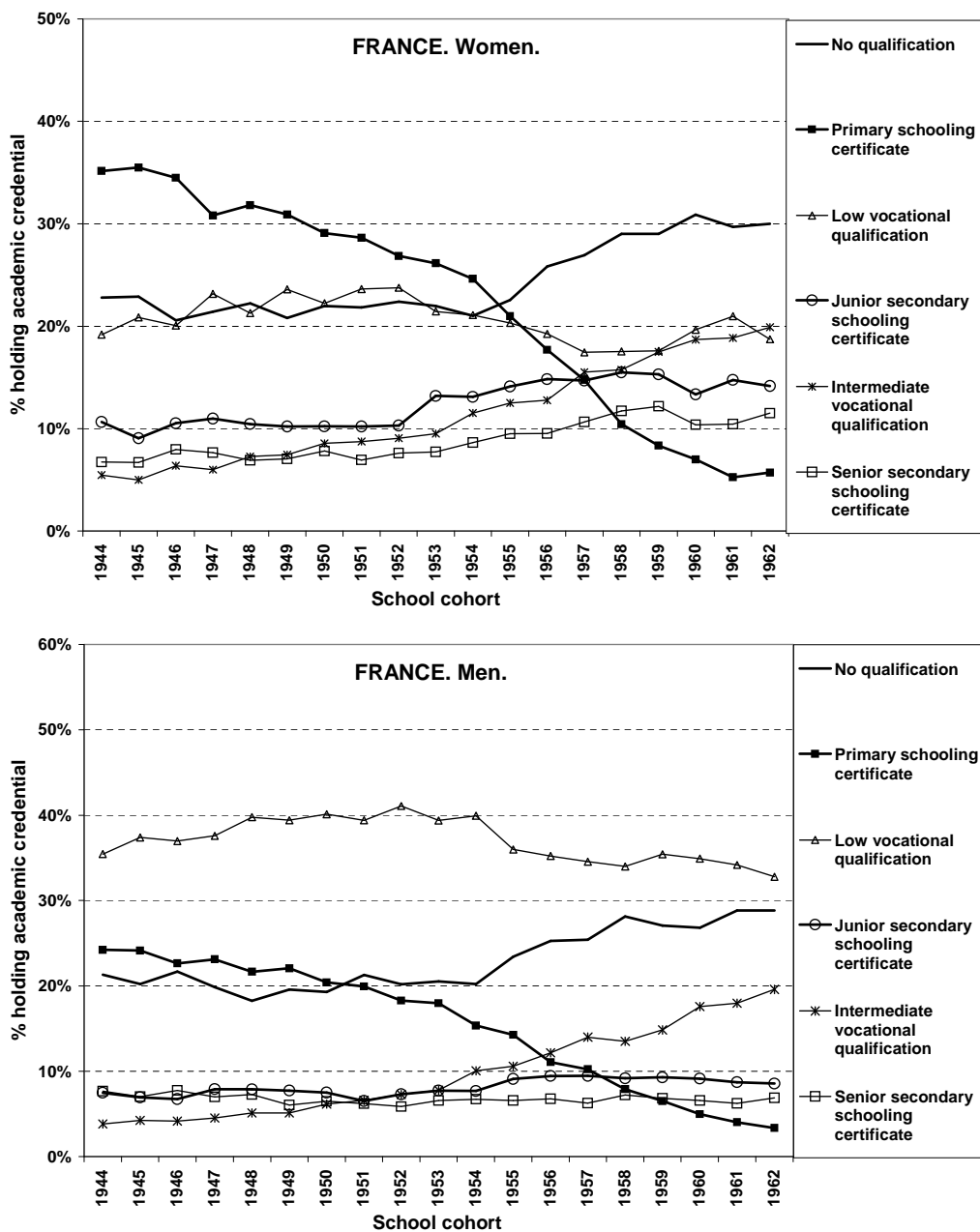
We feel that none of the explanations that we have discussed so far can satisfactorily account for the fact that the raising of the minimum schooling leaving age to 16 had a positive impact on wages in Britain, but not in France. Below, we present pieces of evidence supporting an alternative interpretation, based on the role of academic credentials.

While the actual number of years of education is certainly an important outcome to consider when assessing the effect of the minimum school leaving age increases, we believe that it is not necessarily the most adequate way of measuring the educational benefits of compulsory schooling laws. This is especially true in many European countries, where school curriculums are explicitly structured to allow each level of study to lead to a specific credential. These credentials play a significant role in the job hiring process, as they provide employers with a way of verifying that an applicant has attained a certain level of competency in a specific field. Furthermore, many professional positions are accessible only to workers who hold particular academic credentials. In France and Britain, vocational or technical certificates are often required in several trade, technical, occupational, and business fields, while certificates of secondary education would generally be needed for working as a public or a private sector employee.

Compulsory schooling laws and the distribution of academic credentials.

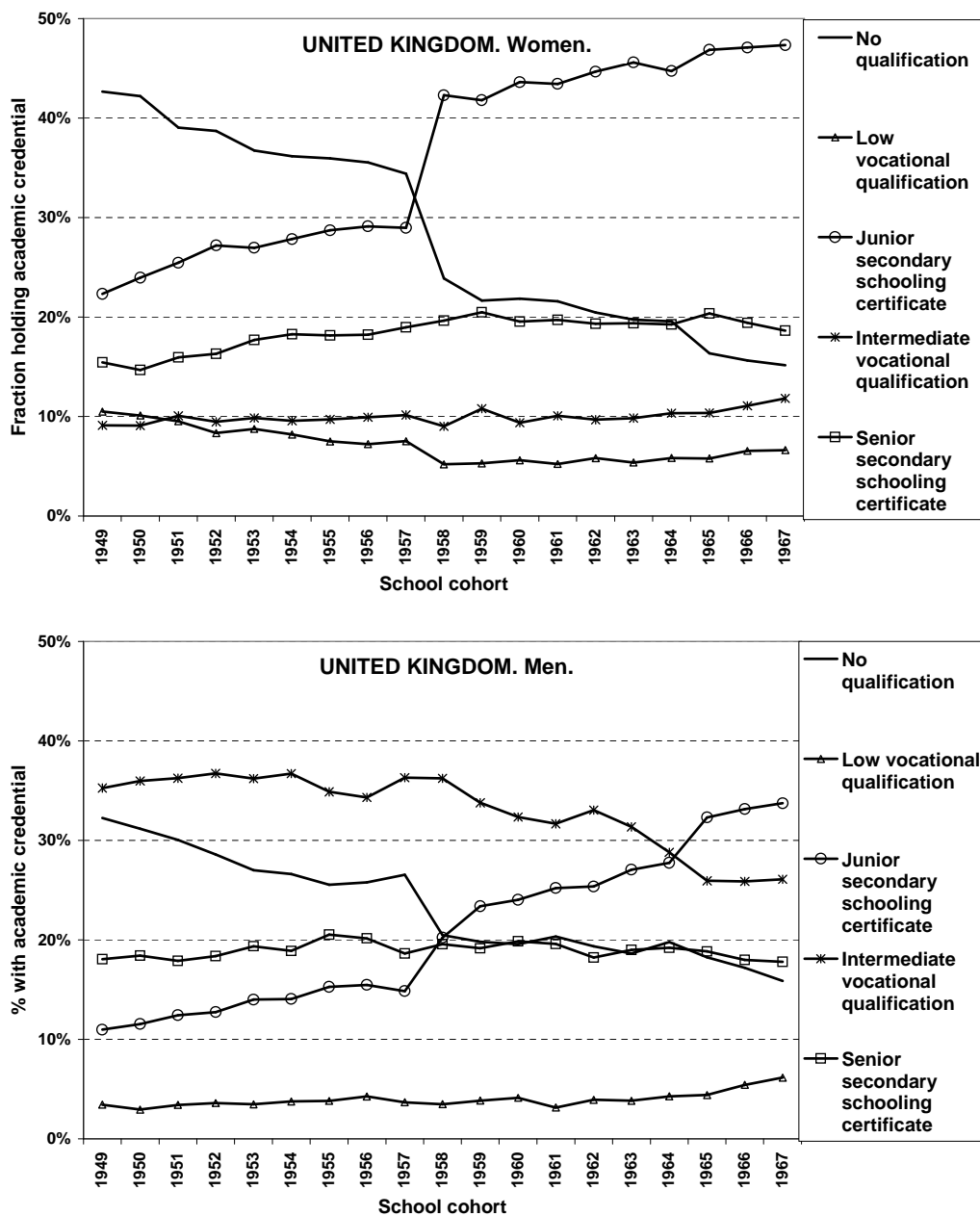
A natural question that arises then is whether, despite their very similar impact on

Figure 1.14: *Impact of the French 1967 Berthoin reform on the distribution of academic credentials, calculated for female and male workers separately. Cohorts 1944-1962. Source: Enquête Emploi (1990-2002).*



Notes: This graph displays the evolution of academic qualifications across school cohorts in France for the subsample of French women (top graph) and men (bottom graph) born between 1944 and 1962 who left school by the age of 18. Academic and vocational credentials are grouped into six categories: no qualification; primary schooling certificate (*Certificat d'Études Primaires Élémentaires*); low vocational qualification (CAP or BEP without the BEPC); junior secondary schooling certificate (BEPC); intermediate vocational qualification (CAP or BEP with the BEPC); senior secondary schooling certificate or above (includes the *Baccalauréat*, advanced vocational training, higher education degrees, etc.).

Figure 1.15: *Impact of the British 1973 Education Act on the distribution of academic credentials, calculated for female and male workers separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).*

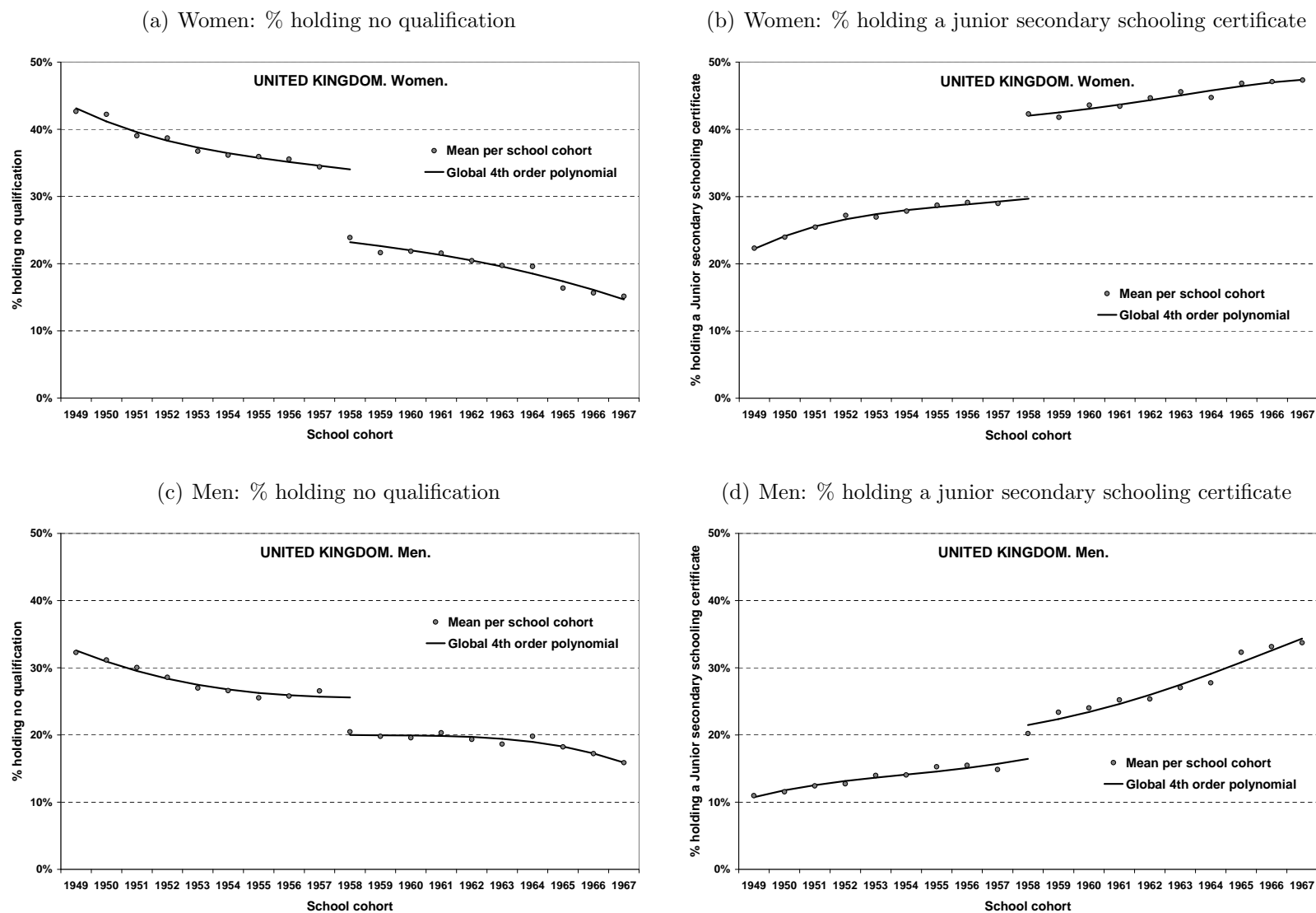


Notes: This graph displays the evolution of academic qualifications across school cohorts in the United Kingdom for the subsample of British women (top graph) and men (bottom graph) born between 1949 and 1967 who left school by the age of 18. Academic and vocational credentials are grouped into five categories: no qualification; low vocational qualification (NVQ level 1 and 2 or equivalent); junior secondary schooling certificate (GCE O-Level or CSE); intermediate vocational qualification (NVQ level 3 or equivalent); senior secondary schooling certificate or above (includes GCE A-Level, NVQ level 3 and 4, higher education degrees, etc.).

average educational attainment, the Berthoin reform in France and the Education Act in Britain had different effects on the distribution of academic credentials. Figures 1.14 and 1.15 display the evolution of the distribution of academic credentials across French and British school cohorts, separately for women and men. These graphs show striking evidence that the two reforms had a very different impact on the level of academic qualifications held by early school dropouts. In the French case, there is no sign of a clear discontinuity around the 1953 school cohort for men, and only a very small jump in the fraction of women holding a junior secondary schooling certificate which seems to compensate a drop in the fraction holding a low vocational credential. On the contrary, we observe in the British case a striking discontinuity around the 1958 school cohort: the fraction of respondents holding no qualification dropped sharply while the fraction of respondents holding a certificate of junior secondary schooling (O-Level or CSE) went up by a similar percentage. Figure 1.15 also indicates that the effect of the 1973 Education act on academic credentials was roughly two times larger for women than for men. The discontinuity is particularly obvious in figure 1.16, in which we plot separately for women and men the fraction of respondents holding no qualification against the fraction of respondents holding a junior secondary schooling certificate, the solid line showing the fitted values from a global 4th order polynomial allowing for an intercept shift at the 1958 school cohort.

Tables 1.16 and 1.17 present the corresponding RDD estimates, which are computed separately by gender and academic credential. The results confirm the graphical evidence and are robust across specifications. We find that the Berthoin reform had virtually no effect on the distribution of academic credentials in France. Most coefficients for men are not significantly different from zero, except for a small and far from obvious percentage point increase in the fraction holding a senior secondary schooling certificate or above. The situation for women is more complicated, as our estimates indicate a small drop in the fraction holding no qualification or a low vocational credential (between 1 and 3 percentage points each) and a slight increase in the fraction holding either a primary or a junior secondary schooling certificate. However, this situation cannot be interpreted as a substantial improvement in the distribution of credentials, as low vocational credentials typically rank higher than

Figure 1.16: *Impact of the British 1973 Education Act on the fraction of individuals holding no academic qualification and on the fraction holding a junior secondary schooling certificate (GCE O-Level or CSE), calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).*



Notes: These graphs show the evolution of the fraction of individuals holding no academic qualification (on the left) and of the fraction holding a junior secondary schooling certificate (on the right), separately for women (top graphs) and men (bottom graphs). The sample includes individuals born between 1949 and 1967 who left school by the age of 18. Dots indicate mean values grouped by school cohort cell. The solid lines are fitted values from a global 4th order polynomial regression with an intercept shift at the 1958 school cohort.

Table 1.16: *Estimated impact of the French Berthoin reform (1967) on the distribution of the highest held academic credential of women and men. Source: Enquête Emploi (1990-2002).*

FRANCE – ENQUÊTE EMPLOI (1990-2002)		
	Global 4 th order Polynomial Regression (1)	Local Linear Regression (triangular kernel) (2)
<u>PANEL A: WOMEN</u>		
% No qualification	-0.030*** (0.010)	-0.017* (0.010)
% Primary schooling certificate	0.024*** (0.007)	0.011* (0.006)
% Low vocational qualification	-0.017** (0.008)	-0.025*** (0.004)
% Junior secondary schooling certificate	0.027*** (0.004)	0.028*** (0.002)
% Intermediate vocational qualification	-0.005 (0.006)	0.002 (0.003)
% Senior secondary schooling certificate or above	0.002 (0.005)	0.002 (0.002)
Number of observations	67,785	67,785
<u>PANEL B: MEN</u>		
% No qualification	-0.014 (0.011)	-0.012 (0.009)
% Primary schooling certificate	0.006 (0.004)	0.002 (0.004)
% Low vocational qualification	-0.010 (0.014)	-0.012 (0.009)
% Junior secondary schooling certificate	0.006 (0.006)	0.009* (0.005)
% Intermediate vocational qualification	0.002 (0.006)	0.003 (0.004)
% Senior secondary schooling certificate or above	0.010*** (0.002)	0.009*** (0.002)
Number of observations	65,774	65,774

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are the fraction of individuals who hold a given academic or vocational credential as their highest qualification. Each coefficient comes from a separate regression. The coefficients reported in column 1 are based on 4th order global polynomial regressions allowing for an intercept shift at the 1953 school cohort, whereas coefficient reported in column 2 are based on local linear regressions using a triangular kernel and an optimally selected bandwidth (using the cross validation criterion). All regressions are performed on a sample constructed from the 1990 to 2002 French *Enquête Emploi* surveys which includes individuals who were born in France, belong to school cohorts 1944 to 1962, left school by the age of 18 and were aged between 25 and 60 when surveyed. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the first of the three waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

Table 1.17: *Estimated impact of the British Education Act (1973) on the distribution of the highest held academic credential of women and men. Source: Quarterly Labour Force Survey (1993-2004).*

UNITED KINGDOM – QUARTERLY LABOUR FORCE SURVEY (1993-2004)		
	Global 4 th order Polynomial Regression (1)	Local Linear Regression (triangular kernel) (2)
<u>PANEL A: WOMEN</u>		
% No qualification	-0.108*** (0.007)	-0.107*** (0.006)
% Low vocational qualification	-0.018*** (0.003)	-0.018*** (0.003)
% Junior secondary schooling certificate	0.124*** (0.005)	0.122*** (0.005)
% Intermediate vocational qualification	-0.003 (0.007)	-0.006 (0.006)
% Senior secondary schooling certificate or above	0.006 (0.004)	0.009** (0.004)
Number of observations	126,531	126,531
<u>PANEL B: MEN</u>		
% No qualification	-0.056*** (0.007)	-0.058*** (0.005)
% Low vocational qualification	-0.003 (0.003)	-0.003 (0.003)
% Junior secondary schooling certificate	0.050*** (0.012)	0.056*** (0.008)
% Intermediate vocational qualification	0.010 (0.014)	-0.001 (0.012)
% Senior secondary schooling certificate or above	-0.002 (0.009)	0.006 (0.007)
Number of observations	114,630	114,630

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are the fraction of individuals who hold a given academic or vocational credential as their highest qualification. Each coefficient comes from a separate regression. The coefficients reported in column 1 are based on 4th order global polynomial regressions allowing for an intercept shift at the 1958 school cohort, whereas coefficient reported in column 2 are based on local linear regressions using a triangular kernel and an optimally selected bandwidth (using the cross validation criterion). All regressions are performed on a sample constructed from the 1993 to 2004 British *Quarterly Labour Force Surveys* which includes individuals who were born in the UK, belong to school cohorts 1949 to 1967, left school by the age of 18 and were aged between 25 and 60 when surveyed. To avoid using the same individual several times, the sample is restricted to respondents who were interviewed in the fifth of the five waves of the survey. Standard errors are clustered by school cohort using the finite-sample correction by MacKinnon and White (1985).

primary schooling certificates. Moreover, the fact that the fraction of women with no qualification was reduced as a result of the Berthoin reform is not a particularly striking feature of the top graph displayed in figure 1.14. On the contrary, our estimates indicate that the distribution of academic credentials in Britain was unambiguously improved by the 1973 Education Act, especially for women. The fraction of British female workers holding no qualification or a low vocational qualification fell by respectively 11 and 2 percentage points, while the fraction holding a junior secondary schooling certificate increased by roughly 12 percentage points. Our results also indicate a positive, although smaller impact for male workers, with a 6 percentage point increase in the fraction holding a junior secondary schooling certificate which compensates a corresponding drop in the fraction holding no academic qualification.

A discrepancy driven by institutional features of school systems. Considering the very similar structure of French and British school curriculums and academic credentials, the fact that the distribution of academic credentials was impacted very differently in both countries seems rather paradoxical. Yet we show below that a number of small institutional differences between the two countries' school systems could well explain why the raising the minimum school leaving age succeeded in improving the academic credentials of early school dropouts in one case and not in the other. In our view, three factors should be considered: the status of junior secondary schooling certification, grade repetition and vocational tracking.

In France, the certificate of junior secondary schooling (BEPC) was never explicitly meant to manifest the fulfilment of compulsory education. The examinations took place at the end of the final year of junior secondary schooling (grade 9) for pupils theoretically aged between 14 and 15. However, these examinations were not mandatory and many junior secondary schooling pupils would simply not sit for them. The fairly marginal status of this academic credential in the French school system can be clearly inferred from the graphs displayed in figure 1.14, which show that among the respondents who were born before 1953 and left school by the age of 18, very few (less than 10%) declare holding the BEPC as their highest academic qualification. In this situation, the fact that this fraction remained unchanged for men

and only slightly increased for women after the introduction of the new minimum school leaving age should not be viewed as completely surprising. On the contrary, the British GCE O-Level and CSE certificates were designed to serve as a standard of qualification for pupils who reached the end of compulsory education. These credentials were exam-based and although not mandatory, they were taken at the age of 15-16 by most pupils and were prepared from the age of 13-14. Figure 1.15 shows that even before the 1973 Education Act came into effect, O-Levels and CSEs already represented a fairly large fraction (about 15% for men and 30% for women) of academic qualifications held by individuals with a school leaving age of less than 18. In this context, the most likely explanation of the sudden jump in the fraction of British respondents who report holding a junior secondary schooling certificate as their highest qualification is that the raising of the minimum school leaving age from 15 to 16 exogenously induced a number early school dropouts to sit for the O-Level or CSE examinations and, for some of them, to be awarded the credential.

In addition to the weaker status of junior secondary schooling certification in France compared with Britain, grade repetition is a second institutional factor that could explain why the impact Berthoin reform had virtually no effect on the structure of qualifications. Grade repetition has always been a very common practice in France, whereas almost no children are held back in the United Kingdom if they do not reach the academic targets. As emphasized earlier, this means that at the time the Berthoin reform, a very large fraction of early school dropouts would be one or two years behind their peers who did not repeat a grade. Many of those who had to comply with the new minimum school leaving age would therefore have reached the age of 16 well before the period of BEPC examinations, enabling them to drop out of school without even sitting for the exam. On the contrary, virtually all 15 to 16-year-old British pupils would be enrolled in year 11, at the end of which the O-Level and CSE examinations took place.

Vocational tracking is the third factor that could explain why the French Berthoin reform had such little impact on the level of academic credentials. As previously noted, about 40% of early school dropouts who were induced by the reform to spend one or two more years in school attended school-based vocational training. Most of

them would have been tracked into this curriculum at the end of the second year of junior secondary schooling, before they could actually sit for the BEPC exam. But then why did the fraction of pupils holding a vocational qualification remain stable after the introduction of the new minimum school leaving age (as can be seen in table 1.16)? The most likely explanation is that the majority of these pupils dropped out of vocational training before actually completing it. Vocational credentials would typically be awarded at the end of a 3-year curriculum, meaning that most pupils reached the minimum school leaving age before actually taking the final examinations. Table 1.15 clearly indicates that this phenomenon was operating when the Berthoin reform came into effect: whereas 181,001 of the 15 to 16-year-old pupils born in 1953 were enrolled in vocational training versus 155,910 among those born in 1952, the gap narrowed considerably at the age of 16-17 (193,832 versus 181,440). In this context, the raising of the minimum school leaving age had little chances of increasing the number of vocational qualifications awarded to early school dropouts.

The comparison of French and British school systems suggests that the status of secondary schooling certification, the practice of grade retention and the timing of vocational tracking are the three major institutional features that explain why, despite their similar effects on average schooling, minimum school leaving age increases in both countries had a very different impact on the distribution of academic credentials for the subpopulation of early school dropouts. We interpret the zero versus positive return to compulsory schooling as a direct consequence of this phenomenon. In France and Britain, the actual quantity of education is likely to be far less important in determining the monetary gains from increased schooling than the credentials held by individuals when they enter the labor market. In such a configuration, the raising of the minimum school leaving age may only increase earnings if the completion of compulsory schooling is sanctioned by an academic certification. This was the case in Britain, but not in France.

1.6.5 The nature of the monetary gains from better academic credentials

The question that remains unanswered is the exact nature of the monetary gains induced by better qualifications. Two possible interpretations of our findings can be given. A first possibility is the existence of a “sheepskin” effect of credentials, driven by a signalling mechanism. According to this hypothesis, the productivity of early school dropouts in Britain may not actually have increased after the introduction of the new minimum school leaving age, but their earnings were higher simply because the signaling value of junior secondary schooling credentials remained unchanged²⁴. A second possible interpretation is that there exists some degree of “lumpiness” in the learning process, which leads to more skill acquisition in examination years than in preceding years²⁵. In this vision, the preparation of junior secondary schooling certification in year 11 induced British pupils to work harder and to learn more skills relevant for the labor market than if there was no final examination.

Obviously, these two lines of interpretations have very different policy implications. In the first scenario, the extent to which compulsory schooling laws can improve the labor market prospects of early school dropouts appears fairly limited, since the signalling value of credentials remains unaffected only if the distribution of academic qualifications does not shift too much. In the second scenario, raising the minimum school leaving age could serve as an efficient productivity-enhancing tool if it induces more pupils to acquire the skills that are needed to succeed on certification exams and that are ultimately rewarded in the labor market.

While it is difficult with our data to distinguish between these two competing explanations, we believe that the signalling theory is too extreme to fully account for the positive returns enjoyed by British school dropouts after the introduction of the new minimum school leaving age. First, the shift in the distribution of academic credentials seems too large to have been completely unnoticed by employers when post-1953 cohorts entered the labor market. Second, we provide some empirical evi-

²⁴Such sheepskin effects would be consistent with sorting models of education (Arrow, 1973; Spence, 1973; Stiglitz, 1975; Weiss, 1983).

²⁵For discussions on this topic, see Chiswick (1973), Lange and Topel (2006) and Flores-Lagunes and Light (2007)

dence supporting the view that, at least for women, the 1973 Education Act actually improved the cognitive skills of early school dropouts. We use data from the British sample of the 1994 International Adult Literacy Survey (IALS)²⁶. This survey was designed to assess the literacy skills of adults in a number of OECD countries. The specific literacy tasks designed for IALS were scaled by difficulty from 0 to 500 points and each participant received a global score for prose, document, and quantitative literacy²⁷. While the relatively small size of the sample (about 3,000 individuals) inevitably reduces the statistical power of the regression discontinuity approach, the IALS survey nevertheless provides a insightful way of assessing whether the 1973 Education Act had any impact on the skills of British pupils.

The graphs displayed in figure 1.17 show, for women and men separately, the average score received on each of the three literacy scales by each cohort of British participants who were born in the UK between 1949 and 1967 and left full-time education by the age of 18. Note that because participants' month of birth is not included in the IALS dataset, cohorts here are calendar cohorts. This is why we decided to exclude participants born in 1958 from the analysis, as some of them (born between January 1 and August 31) faced a minimum school leaving age of 15 while the others (born between September 1 and December 31) faced a minimum school leaving age of 16. The solid lines show the fitted values from a global 4th order polynomial allowing for an intercept shift at the 1958 cohort. Despite a rather erratic pattern, these graphs seem to indicate the existence of a positive impact of the raising of the minimum school leaving age on the literacy skills of women, but not of men (except maybe for quantitative literacy). The corresponding estimates reported in table 1.18 confirm this visual impression and our results appear fairly consistent across parametric and non-parametric specifications. While the actual size of the coefficients should not be over-interpreted, we find evidence of a positive and significant improvement in

²⁶Unfortunately, data on the literacy skills of French participants to IALS are not available because France decided to withdraw from the program after the publication of the preliminary results.

²⁷The survey defines prose literacy as “the knowledge and skills needed to understand and use information from texts such as editorials and news stories”, document literacy as “the knowledge and skills required to locate and use information contained on schedules, maps, tables, graphs, indexes, and so forth” and quantitative literacy as “the knowledge and skills needed to apply arithmetic operations, either alone or sequentially, that are embedded in printed materials, such as calculations required on deposit slips, amount of elapsed time from schedules, cost of goods”.

Figure 1.17: *Literacy scores of British women and men on three scales (prose, document and quantitative literacy). Individuals born between 1947 and 1969. Source: International Adult Literacy Survey (1994).*



Notes: The graph plots the average score obtained by British participants to the 1994 International Adult Literacy Survey on each of the three literacy scales (prose, document and quantitative literacy) and separately for women and men. Scores are in the range 0-500. The sample includes IALS female and male participants who were born in the United Kingdom between 1947 and 1969 and left school between the age of 6 and the age of 18. Individuals born in 1958 are excluded from the sample because some of them (born between January 1 and August 31) faced a minimum school leaving age of 15 while the others (born between September 1 and December 31) faced a minimum school leaving age of 16. The dots show the average score within each calendar cohort cell. The solid lines are the fitted values from a global 4th order polynomial regression with an intercept shift at the 1958 cohort.

Table 1.18: *Estimated impact of the British Education Act (1973) on the literacy scores of female and male participants to the 1994 International Adult Literacy Survey.*

UNITED KINGDOM – INTERNATIONAL ADULT LITERACY SURVEY (1994)		
	Global 4 th order Polynomial Regression (1)	Local Linear Regression (triangular kernel) (2)
<u>PANEL A: WOMEN</u>		
Prose literacy	17.52*** (6.65)	14.85*** (4.08)
Document literacy	16.52** (8.45)	13.89** (6.10)
Quantitative literacy	16.55*** (6.36)	16.30*** (4.35)
Number of observations	738	738
<u>PANEL B: MEN</u>		
Prose literacy	1.25 (13.66)	4.28 (7.62)
Document literacy	4.55 (15.48)	6.17 (8.63)
Quantitative literacy	12.39 (14.85)	16.17** (7.07)
Number of observations	571	571

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The dependent variables are the scores obtained by British participants to the 1994 International Adult Literacy Survey on each of the three literacy scales (prose, document and quantitative literacy), separately for women and men. Scores are in the range 0-500. Each coefficient comes from a separate regression. The coefficients reported in column 1 are based on 4th order global polynomial regressions allowing for an intercept shift at the 1958 cohort, whereas coefficient reported in column 2 are based on local linear regressions using a triangular kernel and an optimally selected bandwidth (using the cross validation criterion). The sample includes IALS female and male participants who were born in the UK between 1947 and 1969 and left school between the age of 6 and the age of 18. Individuals born in 1958 are excluded from the sample because some of them (born between January 1 and August 31) faced a minimum school leaving age of 15 while the others (born between September 1 and December 31) faced a minimum school leaving age of 16. Standard errors are clustered by birth cohort using the finite-sample correction by MacKinnon and White (1985).

literacy skills on all three scales for women, and no significant effect for men except for a positive increase in their quantitative literacy score when using the local linear regression specification.

Given the imprecision of our estimates, the fact that we find no evidence of clear improvement in the literacy skills of men does not necessarily mean that the effect was zero, nor that the reform did not help them improve other skills relevant for the labor market. Nonetheless, a smaller impact for men than for women should not be viewed as something totally implausible. First, we saw earlier that the increase in the fraction of British women holding a junior secondary schooling certificate was two times larger than for men and that their estimated positive returns to compulsory schooling appear more robust. Second, a number of empirical studies²⁸ have found that female and male educational outcomes respond very differently to financial and non-financial incentives for academic performance, the effects being usually higher for female students than for males. Both the larger shift in the distribution of academic credentials and the higher impact on literacy skills for British women are consistent with this interpretation, which could also explain the small but positive effect of the Berthoin reform on the fraction of French women holding a junior secondary schooling credential as opposed to the zero effect for French men.

Taken together, these various observations lead us to the conclusion that the 1973 British Education Act yielded positive wage returns because it induced a number of early school dropouts, and especially female pupils, to complete junior secondary schooling and to learn the skills that are needed to succeed in the final examinations and are rewarded in the labor market. On the contrary, because the French 1967 Berthoin reform was not explicitly designed to reduce the fraction of pupils leaving school with no academic qualifications, it failed to raise the labor market outcomes of those who complied with the new mandatory requirements.

²⁸See in particular Angrist and Lavy (2002) and Angrist et al. (2007).

1.7 Conclusion

Comparing the effects of two major compulsory schooling reforms has enabled us to shed light on some of the institutional factors that may drive the returns to compulsory schooling.

The 1967 Berthoin reform in France and the 1973 Education Act in the United Kingdom both raised the minimum school leaving age to 16. Using large data sets constructed from national labor force surveys and applying a Regression discontinuity approach, we find that despite having impacted a similar fraction of the population and having increased the average educational attainment by the same amount (between 3 and 4 month of additional schooling), both laws had a very different effect on wages. Our estimates indicate that the Berthoin reform had no significant impact on the hourly wages of French women and men, whereas the Education Act yielded positive monetary returns in the range of 4 to 8% for both British women and men. While our estimates are significantly below those reported in previous studies carried out on US or Canadian data, they are fairly consistent with the results of a number of recent papers which find small and sometimes zero returns to compulsory education using European data.

We argue that interpretations based on heterogenous returns to schooling and wage-setting institutions are unlikely to explain the observed discrepancy between France and Britain. Furthermore, we find no evidence that the zero returns to compulsory schooling in France could come from a defective implementation of the Berthoin reform, since retrospective statistics show an unambiguous increase in total junior secondary school enrollment. Instead, we suggest an explanation based on the role of academic credentials. There is striking empirical evidence that French and British compulsory schooling laws had a very different impact on the distribution of academic credentials among the subpopulation of early school dropouts. Whereas the introduction of the new minimum school leaving age reduced the fraction of British school leavers holding no qualification by 12 and 5.5 percentage points for women and men respectively, the distribution of credentials held by French female and male workers was essentially unaffected by the Berthoin reform. We show that a number of institutional differences between French and British school systems, such as

the status of junior secondary schooling certification, grade retention and vocational tracking, can plausibly account for this discrepancy.

We interpret our findings as suggesting that, in education systems where sub-degree qualifications certify the knowledge and skills a person has achieved through study, the actual number of years of education matters less than credentials in determining the returns to education. In such systems, compulsory schooling laws may be effective only if they are consistent with the general organization of the school curriculum. This means that for early school dropouts to actually benefit from increased schooling, the raising of the minimum school leaving age must induce a significant fraction of pupils to become enrolled in grades in which they can take academic or vocational certification examinations. We present some evidence supporting the idea that the benefits derived by early school dropouts from better credentials are not entirely driven by signalling mechanisms, but also reflect an actual improvement in skills that are relevant for the labor market. Part of the observed “sheepskin” effects of credentials in the United Kingdom could originate from the fact that the preparation of junior secondary schooling examinations significantly enhanced the skill acquisition of Year 11 British pupils.

Naturally, our results must be appreciated in light of the limitations of any country-level assessment of educational reforms. While we believe that the comparison of the French Berthoin reform and the British Education Act can help us to better identify some of the key factors that drive the returns to compulsory schooling, our conclusions cannot be generalized to other countries without further investigations. In particular, they do not provide a straightforward explanation for the puzzle that estimated returns to compulsory schooling appear larger in the US or Canada than in many European countries. In any case, our findings show that raising the minimum school leaving age does not necessarily improve the labor market prospects of early school dropouts and that the particular organization of national school systems is an important parameter to take into account when assessing the benefits of compulsory education.

Deuxième partie

SECTORISATION, MARCHÉ IMMOBILIER ET INÉGALITÉS SCOLAIRES

La croissance très rapide des effectifs scolarisés dans le premier cycle de l'enseignement secondaire et son accélération après l'entrée en vigueur de la scolarité obligatoire à 16 ans, ont nécessité la mise en place d'un mécanisme d'allocation des élèves aux différents établissements scolaires : le principe de sectorisation, qui fut instauré en 1963, stipule qu'un élève est scolarisé dans l'établissement scolaire public qui correspond à son lieu de résidence. À l'origine, la sectorisation n'avait pour autre vocation que celle d'être un instrument prévisionnel de gestion des ressources éducatives : elle permettait de planifier les ouvertures et fermetures de classes en fonction de l'évolution de la démographie locale. Ce n'est que progressivement que ce qui n'était qu'un moyen d'affectation des élèves parmi d'autres s'est transformé en un instrument de promotion de la mixité sociale. La seconde partie de cette thèse est consacrée à l'analyse des conséquences de ce mode d'allocation d'élèves, à travers la prise en compte des interactions entre choix scolaires et choix résidentiels²⁹.

Près de 40 ans après sa mise en place, force est de constater que la sectorisation n'est pas parvenue à homogénéiser le recrutement social des établissements scolaires, au point que certains y voient le symbole suprême de l'« hypocrisie » du système éducatif français (Dubet et Duru-Bellat, 2000). Le problème se situe pour l'essentiel dans les grandes agglomérations urbaines, au niveau du collège, et s'explique par deux séries de facteurs : l'existence d'une forte ségrégation résidentielle, d'une part, qui implique que les zones de recrutement des établissements scolaires se superposent à un tissu urbain fortement différencié socialement ; l'existence d'un secteur privé largement subventionné, d'autre part, qui permet d'échapper à moindre coût aux contraintes de la carte scolaire.

Le chapitre 2 propose une analyse théorique du lien unissant politiques de sectorisation, stratification résidentielle et inégalités scolaires. L'objectif de cette contribution est d'étudier les propriétés de modalités alternatives d'allocation des élèves aux établissements scolaires (sectorisation stricte, redécoupage de la carte scolaire ou libre choix de son école) afin d'éclairer les enjeux du débat qui s'est ouvert en France au sujet de la réforme de la sectorisation. Pour prendre en compte les spécificités institutionnelles du système éducatif français, nous développons un modèle spatial

²⁹Cette partie a été co-écrite avec Gabrielle Fack.

de choix résidentiel et scolaire qui s'inspire de modèles existants (Bénabou 1993 ; Nechyba, 1997 ; Epple et Romano, 2003 ; Rothstein, 2006) mais où coexistent un secteur public financé au niveau national et un secteur privé subventionné. Parmi les ingrédients essentiels de ce modèle figurent la répartition spatiale non aléatoire de la qualité intrinsèque des quartiers résidentiels, la corrélation positive entre revenu des parents et niveau scolaire des enfants et l'existence d'effets de pairs à l'école. Ce cadre théorique permet une identification précise des mécanismes par lesquels les choix résidentiels et scolaires sont affectés par le mode de recrutement des établissements scolaires et modifient de manière endogène la composition sociale des écoles et les performances scolaires des élèves. Cet exercice permet également de quantifier les gains et les pertes subies par les élèves dans les différents systèmes, en fonction de leur niveau scolaire et du revenu parental. Bien que les solutions de ce modèle ne soient pas calculables analytiquement, elles peuvent être étudiées au moyen de simulations numériques que nous calibrons sur les données parisiennes en utilisant les estimations empiriques réalisées dans le chapitre suivant. Nos simulations suggèrent qu'en l'absence d'établissements privés, le redécoupage des secteurs scolaires est beaucoup plus profitable aux élèves issus de milieux défavorisés que la suppression de la carte scolaire, qui entraîne dans notre modèle une stratification totale des écoles selon le niveau des élèves. Cependant, l'impact de cette politique sur les choix résidentiels réduit de façon sensible le niveau de mixité sociale initialement visé. Nous montrons également que la présence du secteur privé modifie profondément la capacité d'une telle politique à réduire les inégalités scolaires. En effet, tout redécoupage de la carte scolaire entraîne une fuite des élèves issus des milieux aisés vers le privé, ce qui conduit à une diminution du niveau des écoles publiques et annule les effets bénéfiques de cette politique pour les enfants les plus défavorisés.

Le chapitre 3 est consacré à l'estimation de l'impact du niveau des collègues sur le prix des logements à Paris. Plus précisément, nous estimons la disponibilité à payer des ménages pour un meilleur établissement, qui est l'un des paramètres nécessaires à la calibration du modèle développé dans le chapitre précédent. Pour y parvenir, nous croisons les informations issues de sources diverses, qui incluent le fichier exhaustif des transactions immobilières réalisées à Paris au cours de la période 1997-2003, le

découpage des secteurs scolaires en vigueur chaque année depuis le milieu des années 1990, des indicateurs de performances des collèges publics et privés et les caractéristiques socio-démographiques des quartiers résidentiels parisiens. L'estimation de l'impact de la performance des collèges publics sur les prix immobiliers est compliquée par le fait que la qualité des établissements est une variable endogène, les bons collèges étant en général situés dans les quartiers les plus cotés. Pour résoudre ce problème, nous proposons une version améliorée de la stratégie proposée initialement par Black (1999), qui consiste à comparer les prix des transactions immobilières localisées de part et d'autre des frontières entre secteurs scolaires. Nos résultats indiquent qu'une augmentation d'un écart-type du niveau d'un collège entraîne une augmentation des prix immobiliers comprise entre 1,5 et 2,5 % selon l'indicateur de performance retenu. Les différences de niveaux entre collèges expliquent environ 5 % des différences de prix constatées entre les secteurs scolaires à Paris. Ces résultats sont du même ordre de grandeur que les estimations réalisées sur données américaines et anglaises pour les écoles primaires (Black, 1999 ; Gibbons et Machin, 2003 ; Figlio et Lucas, 2004). Dans un second temps, nous exploitons la forte présence du secteur privé sous contrat à Paris, qui scolarise près d'un tiers des collégiens, pour analyser la manière dont cette offre éducative alternative affecte la disponibilité des ménages à payer pour la qualité des établissements publics. Nous montrons que conformément aux prédictions théoriques des modèles de choix scolaire, une plus forte densité de collèges privés dans un quartier résidentiel réduit localement la capitalisation de la performances des collèges publics dans les prix immobiliers.

Chapitre 2

Politiques de sectorisation, inégalité scolaires et ségrégation résidentielle : analyse théorique et simulations sur données parisiennes

2.1 Introduction

The theoretical literature on public good provision and sorting into neighborhoods has been growing since the seminal paper by Tiebout (1956). More recently, the heated debate on school choice versus school zoning started a specific trend in the literature that studies the link between school enrollment policies, residential segregation and school stratification by pupil ability.

In this chapter, we construct a theoretical multi-community model of local public and private schools and housing markets with the objective of assessing the analytical properties of various school enrollment schemes (strict school zoning, shifting of school attendance boundaries, open enrollment) on housing prices, residential and school social segregation as well as on student performance in the presence of peer effects. This theoretical framework allows for a precise identification of the channels through which the behavioral response of parents to different school enrollment policies might affect their housing location and their decision to opt out of the public school system.

This exercise also permits a quantification of the educational gains and losses incurred by pupils in the different enrollment schemes, depending on their ability and parental income.

In order to calibrate the model and to perform school enrollment policy simulations, we use the values of the parameters that we estimate in the next chapter.

The remainder of this chapter is organized as follows. In Section 2.2, we review the theoretical literature on residential sorting and school enrollment policies¹. The next two sections are devoted to the presentation of our model. Section 2.3 presents the simple case where all schools belong to the public sector. Private schools are introduced in section 2.4. Section 2.5 presents the simulation results of alternative school enrollment schemes. Finally, section 2.6 calibrates the model using the estimated values of the key parameters in the Parisian case.

2.2 Review of literature: theoretical models of school choice

In line with canonical models of local public finance, a strand of the literature has focused on the quality of local public schools. Most models show that when schools are financed locally, social inefficiencies may arise because of a sorting phenomenon.

Fernández and Rogerson (1996) study a multi-community model in which the local public good is educational spending per pupil, which is financed through local income tax. Households are heterogenous in income. In this model, both the level of provision of school quality and the level of tax are affected by the composition of the community, a feature that may lead to multiple equilibria. In this type of models, imposing a single crossing condition on preferences is required to yield a unique stable equilibrium, characterized by the stratification of households by income².

¹We do not review the literature that studies other educational issues related to sorting, such as the papers by Bénabou (1996a,b) and Fernández and Rogerson (1998). These papers study the macroeconomic effects of educational systems in the presence of credit constraints and complementarities in productions, and compare systems of local public finance to State finance in the long-run. For a review of this literature, see Fernández (2003).

²Imposing a local stability condition ensures that the pooled equilibrium with all communities having the same composition is not a stable equilibrium.

A critical assumption: the single-crossing condition. As the single crossing condition is imposed in the vast majority of models, let us write it more formally. Typically, a simple multi-community model consists of a number of communities that provide a local good q (such as school quality) at a local price p (either a housing price or a local tax). Households choose where to reside depending on the level of q and p . The utility function $U(c, q)$ of a family is a function of a numeraire good c (equal to the income minus the price of the local good p) and the local public good q . U is assumed to be twice continuously differentiable and its partial derivatives U_1 and U_2 with respect to c and q to be positive.

The single crossing condition assumes that the slope of the indifference curves in the (q, p) space is everywhere increasing (or decreasing) in y :

$$\frac{\partial \left(\frac{dp}{dq} \Big|_{u=\bar{u}} \right)}{\partial y} = \frac{\partial \left(\frac{U_2}{U_1} \right)}{\partial y} = \frac{U_{21}U_1 - U_{11}U_2}{(U_2)^2} > 0 \quad (2.1)$$

condition (2.1) ensures that if a household with income y_0 weakly prefers the bundle (p_j, q_j) offered in community j to the bundle (p_k, q_k) offered in community k where $p_j > p_k$, then all households with income $y > y_0$ also prefer (p_j, q_j) over (p_k, q_k) , as shown in figure 2.1. Alternatively, if a household with income y_0 weakly prefers (p_k, q_k) over (p_j, q_j) , then all households with income $y < y_0$ will also prefer (p_k, q_k) over (p_j, q_j) .

One might wonder how realistic is the single crossing condition. In the most common form where households are heterogenous in income, this condition simply states that the demand for school quality must be increasing with income, i.e. that school quality is a normal good. When households are heterogenous in ability, the condition implies that the demand for school quality is increasing with ability, which might seem more controversial, as we discuss below³.

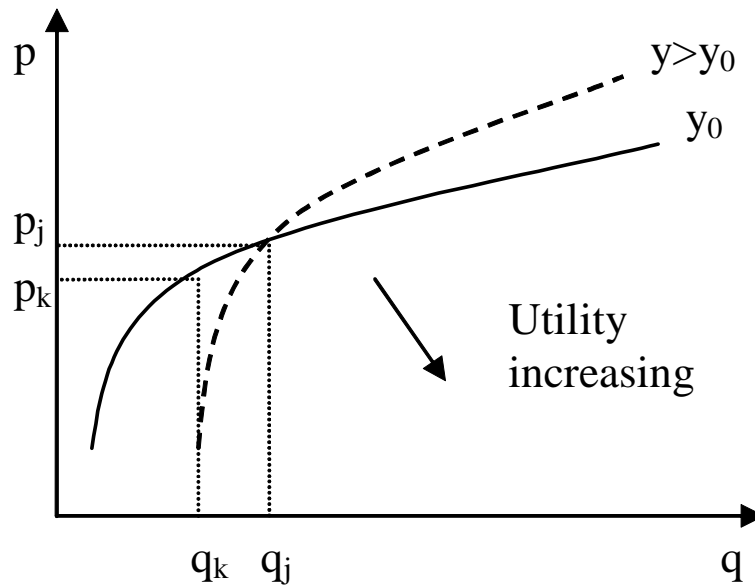
The single crossing condition is a very powerful assumption since it implies that individuals have incentives to sort, and gives rise to stratified equilibria⁴. However,

³In dynamic settings where education is modelled as an investment, it is possible to get the same sorting effects without imposing heterogeneity in preferences, by assuming imperfect credit markets. These types of models are reviewed in Fernández (2003).

⁴Fernández (2003) stresses two other implications of the single crossing condition: first, it ensures the existence of a majority voting equilibrium over the price p of the local public good for models with local taxes; second, it rules out the existence of pooling equilibria (in which all communities

Fernández and Rogerson show that because of peer-group externalities, this equilibrium is not necessarily Pareto-efficient. When a household chooses a community, it does not consider the (positive or negative) effect of its choice on the community peer group⁵.

Figure 2.1: *Single crossing condition: indifference curves in the housing quality (q) and housing price (p) space.*



The importance of peer effects. Models of local school finance in which school quality is solely determined by per pupil spending are not entirely appropriate to study school enrollment schemes within schools districts. In these models, equalizing educational spending across district would indeed suppress the stratification of schools. In order to study the effect of school enrollment policies and in particular the introduction of school choice, one might therefore wish to introduce peer effects at the school level, since this assumption will result in stratification across schools even when finance systems are centralized at the district level. With the introduction of peer effects, school quality may vary with the composition of the student body.

share the same composition) as long as a local stability condition is imposed.

⁵The inefficiency in Fernández and Rogerson (1996) comes from the the additional assumption that the local tax rate voted by the median voter decreases when the mean income in the community increases. Under this assumption, moving some individuals with the lowest income from one community to an adjacent community in which they become the richest group increases mean income in both communities, decreases taxes and increases school quality, leaving everybody better off.

Peer effects play a crucial role in several models, including Bénabou (1993, 1996a,b) and de Bartolome (1990), but we restrict our presentation to models that focus on school zoning and school choice policies.

Modelling peer effects is a difficult task, as the empirical evidence is mixed on this issue. While many empirical papers estimate a linear-in-means model and find that pupils' individual performance depends on the mean ability of their classmates, recent empirical findings by Hoxby and Weingarth (2005) call into question the linearity of peer effects⁶. Their results suggest that even if most pupils benefit from having better peers, too much heterogeneity in the classroom can be detrimental. The worst class composition seems to be the bimodal distribution, with a mix of low-ability and high-ability children and no middle ability peers in-between. According to their estimates, the most efficient class composition would be characterized by a continuum of low to high achieving children without too much variance.

In existing multi-community models, the utility function of households typically depends on the average pupil ability in the school. In one specification, Nechyba (2003) allows for curriculum targeting (a lower variance in peer quality is assumed to improve the targeting of resources). Whether such information on school composition is available to parents and should enter their utility function is disputable. If families only observe an overall measure of school quality that reflects the mean of peers abilities, they can only rely on this information to make their choices. The welfare effects, however, will critically depend on the form of peer effects.

School zoning and school choice. Epple and Romano (2003) develop a model where families derive utility $U(c, q, b)$ from a numeraire good c and their child's educational achievement which is itself a function of school quality q and the child's own ability b . School quality has two components: per student educational expenditure and the mean ability of the school peer group. In the benchmark case, school admission is residence-based and families take school catchment areas into account when choosing their home location. Education is centrally financed and a single income tax rate is determined by majority voting. Epple and Romano show that if income

⁶See Hoxby and Weingarth (2005) for a critical discussion of the linear-in-means model and a review of alternative models of peer effects.

and ability are positively correlated, assuming that the demand for school quality is normal (i.e. the single crossing condition is verified: $\frac{U_q}{U_y}$ increases with y) and independent of ability (i.e. $\frac{U_q}{U_y}$ is invariant with ability), then the model generates a fully stratified equilibrium⁷ such that the wealthiest households live in the community with the best schools and the higher housing prices⁸.

Introducing school choice radically modifies school composition, under the assumption that schools face no capacity constraints and must admit all applicants. As a result, school qualities are equalized, causing a decrease in housing prices. This phenomenon yields an income effect that triggers an increase in educational expenditures (through a higher tax rate). Compared to the school zoning case, low-income families attend better schools whereas the effect is ambiguous for higher income households, as they suffer from a decrease in the peer group quality but benefit from an increase in educational expenditure. Simulations indicate that the aggregate educational change due to a move from school zoning to choice is always negative, once the fall in housing prices is taken into account. The introduction of a fixed transportation cost for those who want to attend a school outside of their neighborhood further reinforces this negative educational effect since in this case, the poorest children are left behind in their local schools whereas the children from the same neighborhoods but with more privileged socio-economic backgrounds pay the cost of attending a better school, causing peer quality to fall everywhere.

This model is insightful but suffers from two main limitations. First, it describes school choice in a fairly unrealistic way, as schools are assumed to have no capacity constraints. Yet in reality, oversubscribed schools usually have the possibility of selecting their pupils. Second, it generates a very high level of stratification, which is not what we observe in the data. This seems problematic for our purpose, since our goal is to calibrate a model to study the effect of various school enrollment schemes on school and residential stratification.

In order to yield more realistic levels of segregation in equilibrium, several models have introduced additional fixed attributes in the community, such as housing

⁷Epple and Romano show that stratified equilibria might also arise if ability is independent of income but the demand for school quality increases with ability.

⁸Other non fully stratified equilibria might exist but they are unstable under a reasonable definition of stability.

quality (Nechyba, 1997) or school effectiveness (Rothstein, 2006). In these models, incomplete stratification arises because choices are constrained by the fixed number of communities. The added richness of such models nevertheless comes with a drawback in terms of tractability: if the existence of equilibria can be proved, they can only be characterized through numerical simulations.

Models with additional community attributes. In a series of papers, Nechyba (1997, 1999, 2000, 2003) developed and calibrated a general equilibrium model of school finance that includes multiple school districts and multiple neighborhoods within school districts. The purpose of these models is to study the effect of various educational policies, such as finance equalization, State grants, school choice and school vouchers. The features of Nechyba’s model exhibit three main differences compared with the previously described general equilibrium models of school finance:

1. Wealth is determined endogenously, as households are initially endowed with a house whose price is set in equilibrium. In addition, taxes are paid on property instead of income. As a result, the single crossing condition is no longer sufficient to ensure the existence of stratified equilibria⁹.
2. The housing stock is fixed, with an *ex ante* partition of districts into neighborhoods of different qualities, and households have to move in order to adjust their housing consumption. This assumption is crucial for the model to exhibit an equilibrium.
3. Private schools are introduced in the model. They function as clubs of parents who share the cost of the school equally and are free to choose their members. As there is no fixed cost, private schools are only composed of single income/peers types of households. On the contrary, public schools must admit all pupils (subject to a residence requirement in the school zoning benchmark case).

In the model, households care about the “quality” of the community, the quality of their house (which is considered exogenous), the quality of the school attended by

⁹As explained in Nechyba (1997), stratified equilibria are obtained by making assumptions about the equilibrium marginal willingness to pay, instead of initially ordering agents by increasing marginal rates of substitution.

their child and private consumption. As in Epple and Romano (2003), school quality is assumed to be a function of per pupil spending and of the average peer quality in the school¹⁰. Public schools are state or locally financed through income or property taxes set by majority vote.

Nechyba (1997) proves the existence of an equilibrium, but simulations are needed to characterize its properties¹¹. In the benchmark equilibrium without private schools, increased valuation of peer effects leads to greater residential stratification of households based on income and wealth, more stratified property values and more stratified school quality. When private schools are allowed to operate, they tend to appear in low-income communities. A number of middle to high-income households then decide to relocate in the high quality houses of low-income communities in order to pay lower house prices and send their children to the newly created private schools. As a result, private schools tend to lower the stratification of income, wealth and property values. The creation of private schools in low-income communities does not directly improve local public schools because they decrease peer quality in the public schools by cream-skimming the best pupils. But since the high-income parents who reside in these communities and send their children to private schools also finance public education through property taxes, spending per pupil increases in public schools. The net effect is ambiguous and varies with the type of financing scheme adopted and the value of the calibration coefficients¹².

¹⁰In the earlier paper (Nechyba, 1999), peer effects are perfectly correlated with income but later papers allow for more sophisticated versions of peers effects.

¹¹Practically, the computation of the equilibrium involves several steps, starting with the benchmark case without private schools. Households are initially endowed with income and a house. They vote over property tax rates (if schools are locally financed) taking other characteristics (housing prices, school quality) as given. Given the tax rate, each household chooses its preferred location, then prices adjust and voting takes place again and so forth until the equilibrium is found. Nechyba states that a unique equilibrium without private schools exists when preferences are identical and communities sufficiently different in their inherent desirability. In the equilibrium with private schools, parents first choose their preferred tuition levels taking other things (taxes, housing prices, school quality) as given, then choose whether they prefer to enroll their child in a public or a private school. The vote over the tax rate then takes place, with households who have chosen private schools voting for a zero tax rate. Again, people choose simultaneously their home location and public/private school. This process is repeated until the equilibrium is reached.

¹²In Nechyba (1999), the latter effect dominates and results in an increase in the overall school quality. But in the calibration that replicates the New York mix of state and local finance (Nechyba, 2000), public school quality falls because the decrease in peer quality is not offset by an increase in public school spending.

The introduction of State-financed private school vouchers yields the same qualitative results, with higher income households opting out of public schools and moving to high quality houses in lower quality neighborhoods where they send their children to private schools. As long as the public support for public school remains above the 50% threshold, per pupil public school spending in poor communities increases as peer quality falls, with a net positive effect. However, the quality of public school in the high-income neighborhoods unambiguously suffers from the introduction of vouchers, as both peer quality and per pupil spending fall.

The case of state-funded public schools where per pupil spending is equalized across districts (instead of locally financed schools), which is of great interest from our perspective, yields pretty much the same results. As long as peer quality matters to parents, state funding does not prevent substantial stratification of income to arise in the benchmark equilibrium. However, the introduction of private schools and private school vouchers yields much lower fiscal benefits for poorer districts, because the increase in per pupil spending triggered by high-income households opting out of the public school system is now spread across the entire State and no longer concentrated in the poorest communities. Moreover, the pattern of relocation of high-income households in low-income communities happens at a much lower pace when school vouchers are introduced, since there are no incentives tied to lower property tax rates when public education is state-funded.

For our purpose, the main predictions of such models are that the presence of private schools and the introduction of school choice through the introduction of vouchers decreases residence-based stratification and housing prices differentials between neighborhoods, while it increases school-based stratification. Public schools experience a decrease in peer quality that may not always be compensated by an increase in per pupil spending. Nechyba points out that these results reflect very “pessimistic” assumptions about the impact of school choice, since there are no efficiency enhancing effect from the higher level of competition between public and private schools. Introducing such effects yields an increase in public school quality as private school attendance increases, which may offset the decrease in peer quality¹³.

¹³Nechyba (2003) introduces two types of effects: *curriculum targeting*, implying that the lower the variance in peer quality, the better the targeting of resources, and *competitive efficiency gains*,

Rothstein (2006) develops a model to test the strength of these efficiency enhancing effects. More precisely, the aim of his paper is to test whether parental preferences depend primarily on school effectiveness or on school peers. This question is related to the argument that school choice is the “tide that lifts all boats”¹⁴. Advocates of school choice argue that increased school competition will give schools the incentives to improve their efficiency, since parents value school effectiveness. But if parents value peer groups more than school effectiveness (they may for example prefer poorly run schools with high quality peer over more effective ones but with a worse student body), then increasing school choice would not strengthen the incentives for school management to improve school effectiveness.

These predictions are derived from a multi-community model where household utility is a function of non-housing consumption (income minus housing rents) and school quality. School quality in community j is a function of two parameters: peer group quality (measured by \bar{x}_j , the mean income of families living in community j) and an exogenous school effectiveness factor μ_j :

$$q_j = \delta \bar{x}_j + \mu_j$$

The fact that parents care about composite school quality q_j rather than about each of its component drives the results of perfect sorting along q_j , but not necessarily perfect sorting along peer group quality and school effectiveness.

Provided that the single crossing condition holds, the rankings of communities by quality, housing rents and average income are identical in equilibrium: the n highest-income families live in the highest quality, highest rent community; the next n families in the second highest-quality, second-highest-rent community, and so on. If parents do not care about peers effects but only about school effectiveness (i.e. $\delta = 0$), there is a unique equilibrium which sorts families by effectiveness. But for higher values of δ , other equilibria might arise, as shown in Rothstein’s simulations. The model predicts that if parents do care about school effectiveness more than school peers, then

which means that the marginal product of a dollar of spending rises in public schools when they face greater competition.

¹⁴See Hoxby (2003).

we should see perfect effectiveness/income sorting. Households with higher income would therefore tend to locate in neighborhoods with better schools, both in terms of peer group and effectiveness. However, if parents value peer groups more than school effectiveness, there can be “unsorted” equilibria in which communities with ineffective schools have the wealthiest families and are the most preferred. These equilibria result from coordination failures, as no individual family in the wealthiest neighborhood is willing to move alone to a district with a lower peer group.

Simulations suggest that this problem is less severe when the number of districts is large, thus offering more choice to families. The difference in peer quality between two “adjacent” communities is smaller in high choice districts, reducing the cost (in terms of peer quality) of moving to the next lower peer group community if it is endowed with a more efficient school. One testable prediction of this model is that if parents have a moderate taste for school effectiveness, then effectiveness sorting should be more complete in “high choice” cities (*i.e.* in areas with many small districts) than in cities with little school choice. This prediction is tested empirically. The estimates show no evidence of higher effectiveness sorting in high choice districts. The most plausible explanation to this result is that either parents place a low weight on school effectiveness or they value it but lack information.

Our model builds principally on the model developed Nechyba, but simplifies it and adapts it to the specific context of the French school system. In our model as in Rothstein’s, we abstract from tax concerns and we do not look at schooling expenditures, as the French system is centrally financed. We focus therefore only on the peer effect component of school quality. But as we also want to derive predictions on residential sorting, we build on Nechyba and assume that families not only care about school quality, but also value neighborhood characteristics.

Finally, we wish to introduce private schools because they are well developed in France and represent an outside option for many parents. Apart from Nechyba, very few papers include private schools in multi-community settings. Epple and Romano (1998) model sorting into private schools but without housing markets. They show that when families differ both in terms of income and child ability, the equilibrium composition of private schools may mix high-ability children from low-income families

with higher income but lower ability children¹⁵. This result comes from the fact that private schools are able to internalize the externalities caused by peer-effects by setting different fees for different levels of abilities. Martínez Mora (2006) introduces choice between private and public schools along with residential choice, but peers do not play any role as school quality is produced with a constant return to scale production function. As we explain below, our goal is to take into account peer-effects at the private school level and to model the specific features of the French private school system.

2.3 Modelling alternative school enrollment schemes

In this section, we construct a multi-community model of school and residential choice that allows us to compare two alternative school enrollment schemes. In the first setting, the assignment of pupils to schools is purely residence-based while in the second, parents can freely choose their child's school. For now, all schools belong to the public sector. Private schools will be introduced in section 2.4.

2.3.1 Strict school zoning

We first consider a setting in which pupil enrollment is based on a strict school zoning scheme.

2.3.1.1 Assumptions

The city. As in Nechyba (1997) and Rothstein (2006), we assume that the city consists of a continuum of families. In our model, families are points of a bounded measurable subset \mathcal{C} of the plane. Henceforth, the measure $\mu(A)$ of a measurable subset A of \mathcal{C} will be the Lebesgue measure (or area) of A . The city is partitioned into K distinct neighborhoods $\{N_1, \dots, N_K\}$ that are assumed to be measurable subsets of \mathcal{C} satisfying $\mu(N_k) > 0$ for any $k = 1, \dots, n$. Each neighborhood is endowed with the same measure of identical houses, owned by absentee landlords. Neighborhood k has

¹⁵They assume that the creation of a private school entails a fixed cost, defining a minimum size for the private school.

a positive specific quality q_k , and we shall denote by q the corresponding bounded measurable function which is constant and equal to q_k on N_k . This specific quality may include all the neighborhood amenities (architectural style, transportation facilities, parks...) except schools. The geographical distribution of these qualities is not uniform, as quality is assumed to decrease with the distance to the city center¹⁶. In the simulations, we assume that the logarithm of neighborhood quality $\ln(q_k)$ follows a distribution ϕ and that its rank is correlated with the rank of its proximity to the city center (denoted $\text{rk}(\frac{1}{d_k})$, where $\frac{1}{d_k}$ is the inverse of the distance to the city center d_k) with a positive correlation equal to λ .

School catchment areas. The city is partitioned into J middle school zones $\{Z_1, \dots, Z_J\}$ that are assumed to be measurable subsets of \mathcal{C} , such that $\mu(Z_j) > 0$ for any $j = 1, \dots, J$. The boundaries of school catchment areas do not necessarily coincide with neighborhoods' borders and some school zones may overlap several neighborhoods. Initially, the assignment of pupils to schools is strictly residence-based: a pupil living in a specific catchment area must attend the local school. The size of schools is therefore equal to the number of children (one per household) living within the corresponding attendance zone.

The subsets $B_{jk} = Z_j \cap N_k$ of \mathcal{C} are called blocks. We will only consider the pairs $(j, k) \in \{1, 2, \dots, J\} \times \{1, 2, \dots, K\}$ such that $\mu(B_{jk}) > 0$.

As the French system is centrally financed, we assume that school spending does not play any role: spending per pupil is the same across schools¹⁷ and schools are equally efficient in the use of their funds. We therefore abstract from tax and voting concerns to focus our analysis on peer effects. As a result, families care about the average ability of children enrolled in the school, because it is the only factor that drives variations in school performance. We assume that a child benefits from higher ability peers in the school. Educational achievement e_i of pupil i is a linear function of her own log ability $\ln a_i$ and of the average level of log ability in her school j (S_j

¹⁶To replicate Parisian features, we suppose that quality decreases as we move away from the city center.

¹⁷We therefore do not have to include the within-neighborhood voting over public school spending, which is a central feature of many multi-community models.

denoting the subset of families whose children are assigned to school j):

$$e_{ij} = \ln(a_i) + \delta \int_{S_j} \ln(a_i) d\mu_i \quad (2.2)$$

Our specification in natural logarithm reflects the fact that even though scores are usually normally distributed, it is not necessarily the case that abilities follow the same distribution. In fact, a log-normal distribution of abilities sounds more plausible, since earnings are log-normally distributed. Our interpretation is that the grading of tests maps log-normal distributions of abilities into normal distributions of test scores¹⁸. We assume that parents only observe test scores, which themselves reflect the log of abilities. The average test score at the school level can be written:

$$\bar{e}_j = (1 + \delta) \int_{S_j} \ln(a_i) d\mu_i \quad (2.3)$$

Our specification of peer effects implies that a point increase in the average test score raises a pupil's own achievement by $\frac{\delta}{1+\delta}$. This coefficient may therefore be easily related to empirical estimates of peer effects based on test score data.

It is important to note that given our linear specification of peer effects, school choice reforms that modify the composition of schools will not yield aggregate educational gains. This specification allows us to restrict our analysis on the distribution of gains and losses induced by different school enrollment policies.

Families. Families living in the city are endowed with an exogenous level of income. We denote by $y_i = y(i)$ the income of family $i \in \mathcal{C}$ and by y the corresponding f income function. Each family must rent a house to live in the city and we assume that the number of houses is equal to the number of families. Every family i has a child of middle school age, with school ability $a_i = a(i)$ and we shall denote by a the corresponding function. Families only differ in their income y_i and in their child's ability a_i . These two variables $y, a : I \rightarrow \mathbb{R}_+$ are assumed to be bounded (and measurable). They are not independent and we will assume that their correlation coefficient ρ is strictly positive.

¹⁸For example, it is well-known that the IQ tests are designed to yield a normal distribution of scores.

The utility enjoyed by family i is a function of two key parameters that will jointly determine its residential choice: private consumption c_i and the desirability h_{jk} of its house located in school zone j and neighborhood k (i.e. block B_{jk}). The utility function $U_{ijk} = U(c_i, h_{jk})$ is assumed to be twice continuously differentiable and U_1 and U_2 (the partial derivatives with respect to the first and second argument) are both positive. Household private consumption c_i is what is left after the payment of the rent p_{jk} .

The desirability h_{jk} of a house is weighted average of two parameters:

1. the quality q_k of the neighborhood k ;
2. the perceived quality of the local middle school j , which is measured by the average logarithm of abilities in this school.

More precisely, we set:

$$h_{jk} = \beta \int_{S_j} \ln(a_i) d\mu_i + (1 - \beta) \ln(q_k) \quad (2.4)$$

with $0 < \beta < 1$.

If we assume that parents do not really observe the average logarithm of abilities in the middle school but instead only know the average test score results \bar{e}_j , then the parameter β which measures the taste for school quality includes peer effects. In this case, we can rewrite the block's desirability as a function of average achievement:

$$h_{jk} = \frac{\beta}{(1 + \delta)} (\bar{e}_j) + (1 - \beta) \ln(q_k) \quad (2.5)$$

Key assumptions on the utility function. We make two important assumptions about households' utility function.

H1: Single crossing condition First, we make the usual assumption that the utility function satisfies the single crossing condition:

$$U_{12}U_1 - U_{11}U_2 > 0$$

This condition simply states that demand for desirable housing is normal. As

we will see below, this condition ensures that any residential equilibrium will be perfectly stratified on the basis of household income (highest-income families will live in the highest desirable neighborhood, and so on). However, it does not imply perfect stratification on school or neighborhood quality, and does not rule out the possibility of multiple equilibria.

H2: Demand for desirability is invariant to its level. In addition, we make the following assumption:

$$U_{22}U_1 - U_{12}U_2 = 0$$

which ensures that:

$$\frac{\partial \left(\frac{U_2}{U_1} \right)}{\partial h} = \frac{U_{22}U_1 - U_{12}U_2}{(U_1)^2} = 0$$

This condition states that the demand for a desirable house is invariant to its level. It is needed for an equilibrium to emerge when families are free to choose the school they wish to send their child to. This condition ensures that two families with the same income y_0 but children of different abilities will have the same demand for housing desirability.

The family of functions that satisfy these two assumptions has the form:

$$U(c, h) = W\left(V(c) + h\right)$$

with V a C^2 strictly increasing concave function and W a C^2 strictly increasing function.

Let us now prove more formally, with this form of utility functions, the stratification property implied by the single crossing condition.

Lemma: Suppose that $h_{jk} > h_{j'k'}$, $p_{jk} > p_{j'k'}$ and that utility functions are of the above form. Then we have:

- (i) If a family with income y_0 (weakly) prefers block (j, k) to block (j', k') , then all families with income $y > y_0$ will strictly prefer block (j, k) to block (j', k') .

- (ii) If a family with income y_0 (weakly) prefers block (j', k') to block (j, k) , then all families with income $y < y_0$ will strictly prefer block (j', k') to block (j, k) .

Proof:

- (i) By hypothesis, we have $W\left(V(y_0 - p_{jk}) + h_{jk}\right) \geq W\left(V(y_0 - p_{j'k'}) + h_{j'k'}\right)$ and hence, since W is strictly increasing:

$$V(y_0 - p_{jk}) + h_{jk} \geq V(y_0 - p_{j'k'}) + h_{j'k'}$$

So, we get:

$$V(y_0 - p_{j'k'}) - V(y_0 - p_{jk}) = \int_{y_0 - p_{jk}}^{y_0 - p_{j'k'}} V'(t) dt \leq h_{jk} - h_{j'k'} \quad (2.6)$$

Since $V' > 0$ and $V'' < 0$, V' is a strictly decreasing positive function. We thus get, for any $y > y_0$:

$$V(y - p_{j'k'}) - V(y - p_{jk}) = \int_{y - p_{jk}}^{y - p_{j'k'}} V'(t) dt < \int_{y_0 - p_{jk}}^{y_0 - p_{j'k'}} V'(t) dt \leq h_{jk} - h_{j'k'} \quad (2.7)$$

and the proof of (i) is complete.

- (ii) The proof is analogous to the proof of (i).

Remark 1. Note that the condition (ii) still holds if we only have $p_{jk} > p_{j'k'}$, provided we assume that the family with income y_0 strongly prefers block (j', k') to block (j, k) . We shall use this remark later to show the stratification properties of all equilibria.

Housing prices. In this model where families are competing for houses, we make the assumption that the price of houses is fixed by landlords. Let us denote by $p : \mathcal{C} \rightarrow \mathbb{R}_+$ the price function which is assumed to be constant on each block B_{ij} .

Family decision problem. Families are utility maximizing agents that choose location (j, k) while taking public school quality and housing prices as given. Family i chooses its location among all feasible locations in the budget set determined by the constraint $y_i \geq p_{j,k}$.

School zoning equilibrium

An equilibrium is defined by a set of housing prices and a partition of families into bins $C_{j,k}$. Given the spatial layout of school zones and neighborhoods, the distribution of neighborhood qualities, income and pupil abilities, families have a housing preference for each block B_{jk} that must satisfy the following conditions:

1. **Each family has a house:** $\mu(B_{jk}) = \mu(C_{jk})$ for any pair (j, k) .
2. **Nash equilibrium:** at the specified housing prices, no family would prefer a house other than the one to which she has been allocated, given the set of affordable locations:

$$U(y_i - p_{jk}, h_{jk}) \geq U(y_i - p_{j'k'}, h_{j'k'}) \quad \forall i, j, k$$

where

$$h_{jk} = \beta \int_{\cup_k C_{jk}} \ln(a_i) d\mu_i + (1 - \beta) \ln(q_k)$$

Our model of strict school zoning describes an economy that is a simplified version of the economy described in Nechyba (1999). It simply abstracts from tax concerns, property wealth and private schools. Nechyba's proof of the existence of an equilibrium (Nechyba, 1999) applies to our model.

As household wealth is exogenous in our model, the single crossing condition implies that the equilibrium is stratified. We will use this property to construct an algorithm that converges rapidly to the equilibrium.

Proposition. In equilibrium, we have, for any pair $(j, k) \neq (j', k')$:

- (i) If $h_{jk} \geq h_{j'k'}$ (respectively $h_{jk} > h_{j'k'}$), then $p_{jk} \geq p_{j'k'}$ (respectively $p_{jk} > p_{j'k'}$).

- (ii) If family i lives in B_{jk} and family i' lives in $B_{j'k'}$ and if $y_i > y_{i'}$, then $h_{jk} \geq h_{j'k'}$ and $p_{jk} \geq p_{j'k'}$.

Proof. Let us prove this proposition.

- (i) Assume that $h_{jk} \geq h_{j'k'}$. To prove that $p_{jk} \geq p_{j'k'}$, we use a proof by contradiction. Assume that $p_{jk} < p_{j'k'}$. Then, any household i who can afford a house in $B_{j'k'}$ can also afford a house in B_{jk} since we have $y_i \geq p_{j'k'} > p_{jk}$. Since U is strictly increasing in both arguments, we get from the inequalities $y_i - p_{j'k'} < y_i - p_{jk}$ and $h_{j'k'} \leq h_{jk}$ the following inequality:

$$U(y_i - p_{j'k'}, h_{j'k'}) < U(y_i - p_{jk}, h_{jk})$$

But we have in equilibrium that for any family i living in $B_{j'k'}$:

$$U(y_i - p_{j'k'}, h_{j'k'}) \geq U(y_i - p_{jk}, h_{jk})$$

which leads to a contradiction. This shows that $p_{jk} \geq p_{j'k'}$. The case of strict inequality is demonstrated in a similar fashion, so (i) is proved.

- (ii) Consider a family i living in B_{jk} , a family i' living in $B_{j'k'}$ and assume that $y_i > y_{i'}$. We have $y_i \geq p_{jk}$, $y_{i'} \geq p_{j'k'}$ and since $y_i > y_{i'}$, we get $y_i > p_{j'k'}$. In particular, family i can afford to buy a house in block $B_{j'k'}$. Since we are in equilibrium, we have:

$$U(y_i - p_{jk}, h_{jk}) \geq U(y_i - p_{j'k'}, h_{j'k'}) \tag{2.8}$$

Let us prove by contradiction that we must have $h_{jk} \geq h_{j'k'}$. If $h_{jk} < h_{j'k'}$, we would get $p_{jk} \geq p_{j'k'}$ by property (i). By remark 1 (see above), we would then write (using equation (2.8)):

$$U(y_i - p_{jk}, h_{jk}) > U(y_i - p_{j'k'}, h_{j'k'})$$

for any $y < y_i$, and hence, for $y = y_i$:

$$U(y_i - p_{jk}, h_{jk}) > U(y_i - p_{j'k'}, h_{j'k'})$$

But this contradicts the relation:

$$U(y_i - p_{j'k'}, h_{j'k'}) \geq U(y_i - p_{jk}, h_{jk})$$

which follows from the fact that we are in equilibrium. Since we get a contradiction, we have proved that $h_{jk} \geq h_{j'k'}$. The relation $p_{jk} \geq p_{j'k'}$ follows from (i).

This proposition ensures that the equilibrium will be perfectly stratified by income. The equilibrium allocation of families to blocks is generally not unique because the quality of neighborhoods within a particular school zone might be heterogenous. However, the simulations that we perform in section 2.5 to compare the properties of alternative school enrollment policies are based on an algorithm that always converges to a unique equilibrium for a given initial geographical distribution of households.

2.3.2 School choice

So far, we have assumed that a strict school zoning policy is enforced by the local education authority. We now examine what happens if we introduce school choice.

Ranking of schools under school choice

We model school choice as follows. A family is free to apply to any school and to move from one school to another as long as the school principal is willing to enroll the child. Total enrollment cannot exceed a school's capacity which is assumed to be equal across schools. Any transportation cost paid to attend a school located outside the district is publicly funded so the location of middle schools no longer matters. Families have identical preferences over schools: they care about average test scores, so all of them would prefer sending their child to the school which enrolls the most able pupils. For a given distribution of average abilities across establishments, the

ranking of schools will be identical for all families.

The school choice allocation rule. We assume that the objective of school principals is to maximize their establishments' performance. The optimal strategy is therefore to enroll children with the highest abilities. For the moment, we assume that a school principal can freely choose the student body, and therefore selects the m^{th} highest ability pupils among applicants. The allocation process will typically involve several rounds, the principal being allowed to turn down an application to accept instead a pupil with higher ability.

School choice equilibrium. An equilibrium is reached when all children are enrolled in a school and when no family can find a preferred school in which its child could be enrolled instead. In equilibrium, the highest ability children will be allocated to their preferred schools and lower ability children would not be admitted to that school by the principal. Schools are therefore perfectly stratified by ability: the children that compose the student body of the j^{th} preferred school ($1 \leq j \leq J$) have an ability comprised in the interval $[\underline{a}_j, \bar{a}_j]$.

In this set-up, the quality $\Gamma(j)$ of the j^{th} preferred school is higher than the quality of all schools j' such that $j' < j$ and can be written:

$$\Gamma(j) = \overline{\ln(a)}^j = \int_{A_j} \ln(a_i) d\mu_i \quad (2.9)$$

where A_j is the set of all pupils whose ability a_i lies in the interval $[\underline{a}_j, \bar{a}_j]$.

Residential equilibrium

The residential equilibrium is defined in the same way as in the residence-based assignment setting, but the equilibrium that emerged under strict school zoning will not in general be an equilibrium under the new school choice scheme. In the strict school zoning scheme, households are sorted on the basis of the desirability of blocks, which is itself a function of both neighborhood quality and local public school quality.

When school choice is introduced, local public school quality no longer matters for family i and is replaced in its utility function by the quality of the "chosen" public

school, $\overline{\ln(a)}^j$. Note that all blocks within a given neighborhood k are now strictly equivalent from the point of view of parents, so we drop the subscript j and denote by B_k all blocks of this neighborhood and by h_k the common desirability of the houses located within these blocks.

The stratified equilibrium. The single crossing condition defined over the desirability of block k is no longer sufficient to ensure that the stratified equilibrium is a Nash equilibrium. We need to show that at the specified housing prices and given the current distribution of income across neighborhoods and the current distribution of abilities across schools, no family would prefer to live in a district other than the one in which it currently resides:

$$U(y_i - p_k, h(a_i, q_k)) \geq U(y_i - p_{k'}, h(a_i, q_{k'})) \quad \forall i, k, k'$$

If we assume that $q_k > q_{k'}$ and $p_k > p_{k'}$, the single crossing condition ensures that if a family with income y_0 and child ability a_0 weakly prefers block k to block k' , then a family with higher income but identical child ability will strictly prefer block k . The proof is similar to the one given in the strict school zoning setting.

But as abilities are not perfectly correlated with income, the single crossing condition does not guarantee that a family with the same income y_0 but different child ability ($a_1 \neq a_0$) also strictly prefers block k to block k' . Assumption H2 ensures that this holds (the proof is straightforward given the family of functions defined above).

Under these two conditions, a stratified equilibrium exists with higher income households living in the more desirable neighborhoods (in terms of intrinsic quality q). As neighborhood qualities are assumed exogenous, this equilibrium is unique. School quality is no longer capitalized in housing prices that are determined uniquely by differences in neighborhood qualities.

2.4 Introducing private schools

We have assumed so far that the only available schools are public schools. This assumption is very unrealistic, in particular in the French case, where private schools

account for about 20% of total middle school enrollment (the proportion going up to 30% in Paris). Yet modelling private school in the presence of peer effects is not a straightforward task. In the existing literature, Nechyba (followed by Ferreyra, 2007) is the only author to introduce private schools with peer effects in a housing market model. However, he ensures that peer quality is unique for every child by assuming that the creation of a private school does not require any fixed investment, so the size of a particular private school could be equal to one student. As a result, every private school enrols only pupils that share the same income and ability. In this set-up, a family willing to opt out from the public school system always has the possibility of sending its child to a private school with a peer group that matches its child's ability.

Because many features of the French education system would not be well reproduced under such assumptions, we decided to model private schools in a way that seems more appropriate.

Some stylized facts about private schools in France. The specific features of the French private school system are detailed in the next chapter of this thesis. Three main specificities are worth emphasising here:

- Private schools can freely select their pupils. As a result, the average performance of private schools is usually higher than that of public schools.
- Private school fees are much smaller than in the US or the UK, because the vast majority of public middle schools receive public funding which covers teachers' pay and most equipment expenses. For simplicity, we will assume that there are no private school fees.
- As private schools are usually State-funded, the opening of new schools is subject to very tight restrictions¹⁹. The size (and growth) of the private school sector is therefore limited. More precisely, public subsidies to private schools depend on the approval of a commission (composed of members of public administration, representatives of local communities and representatives of private

¹⁹In France, although opening a private school is relatively easy, it is much more difficult to get public funding. As a result, the number of non-subsidized private middle schools (*Hors contrat*) is very small and can be neglected.

schools) and must meet a “recognized need for education” (*besoin scolaire reconnu*)²⁰.

In a strict school zoning scheme, private schools offer an outside option to parents that are unwilling to send their child to the local public school. However, as the size of the private sector is restricted, private schools cannot completely replace public schools, as is the case in Nechyba (1999) or Epple and Romano (1998), where private schools are costless. If the size of the private sector was unrestricted, we would end up with an equilibrium similar to the school choice equilibrium, with private schools fully replacing public schools.

Private schools as local clubs. In our model, private schools function as local clubs but we assume that there is an implicit fixed cost to create a private school and a maximum public subsidy such that the size of a private school corresponds to a neighborhood (assuming that all children living in the neighborhood attend the private school). For simplicity, we do not model the cost of setting up a private school since in France, this cost is usually not incurred by parents of children enrolled in the private sector. However, if private schools were costless and if no restrictions were imposed on the number of private schools operating in a city, they would completely replace public schools and be perfectly stratified by parental income.

In order to model the restrictions that apply to private schooling in France, we need to define a rule for the creation of private schools. As publicly funded private schools can only be created if they meet a “recognized educational need”, we assume that private schools will appear in neighborhoods where parents are the most dissatisfied with the quality of the public school. We suppose that parents living in the same neighborhood can form a coalition to set up a private school and opt out from public education. This coalition is sustainable if the quality of this prospective private school is high enough compared to the quality of the assigned public school. For each neighborhood, we define a “will to secede” s_k from the public school as the difference between the quality of the private school $\overline{\ln(a)}^k$ that could be created in the neighborhood (assuming that all children in the neighborhood attend the private

²⁰see art. L442-5 to L442-11 of the *Code de l'Éducation*.

school) and the quality of the local public middle school $\overline{\ln(a)}^j$:

$$s_k = \overline{\ln(a)}^k - \overline{\ln(a)}^j \quad (2.10)$$

If a neighborhood k is included in several school zones, the will to secede is the weighted average will to secede in the neighborhood:

$$s_k = \overline{\ln(a)}^k - \sum_j \frac{\mu(B_{jk})}{\mu(N_k)} \overline{\ln(a)}^j \quad (2.11)$$

where $\mu(B_{jk})$ is the measure of the fraction of neighborhood k that is assigned to public school j and $\mu(N_k)$ is the total measure of neighborhood k .

We assume that parents with the highest will to secede s_k are willing to spend more time lobbying members of the commission for private schools. As a result, only neighborhoods with a s_k superior or equal to a specific exogenous threshold s_0 will be allowed to create a private school. We assume that the private school never turns down the application of children that live in the neighborhood. As all the children in the neighborhood prefer the private school to the public school, it is not possible for a child outside the neighborhood to enter the private school, given the capacity constraints. These assumptions ensure that private schools do not entirely cream skim public schools. Otherwise, we would end up with only the pupils belonging to the top percentiles of ability attending private schools. Our model is closer to what is observed in reality: private schools are on average better than public schools but some public schools rank higher than private schools.

Private schools in a model of school zoning. We first establish the equilibrium that emerges in the absence of private schools when the assignment of pupils to public schools is strictly residence-based, and then investigate the effect of introducing private schools in a dynamic setting. The introduction of private schools is therefore dependant on the previous equilibrium and we perform simulations to reach the new one. We do not formally prove the existence of an equilibrium with private schools but our simulations suggest that we can always find one. The creation of a private school changes the valuation of the neighborhood where it is located, since private

school quality is always better than that of the local public school. The desirability of a neighborhood in which a private school operates therefore becomes:

$$h_{jk} = \beta(\overline{\ln(a)}^k) + (1 - \beta) \ln(q_k) \quad (2.12)$$

Where $\overline{\ln(a)}^k$ denotes the average performance in the neighborhood's private school and q_k is the neighborhood's intrinsic quality. Note that the housing price differential between two parts of the same neighborhood assigned to different public schools disappears if a private school is created in the area, as the two parts of the neighborhoods now share the same desirability.

In our simulations, if a private school created in a neighborhood in period t becomes unsustainable in period $t + 1$ because of the lower average ability of the new residents that have moved to the neighborhood, then the private school should close. In fact, the simulations show that under reasonable values of the parameters, this situation never happens. Once a private school is created, the quality of the neighborhood increases relatively to other neighborhoods while the quality of the public school in the area decreases, lowering the rank of surrounding neighborhoods. This leads in equilibrium to an increase in the average income of neighborhoods that have shifted to private schooling and a decrease in surrounding areas.

Private schools and school choice. Under the model's assumptions, private schools tend to disappear when school choice is introduced, because it becomes almost impossible to find a neighborhood in which all families are willing to support a private school instead of sending their child to the public school of their choice. If a family has a child with high-ability, it would prefer sending her to the best public school instead of the private school, causing a failure in the support for the private school. As a result, private schools might only subsist in the top income neighborhoods, if the average ability of peers is high enough compared with the best public school. This could happen if the size of the neighborhood is small enough compared to the size of public schools and if the correlation between income and ability is very high. However, our simulation show that the occurrence of such a scenario is rare enough to be neglected for credible values of the parameters.

2.5 Simulations

In this section, we use our model to perform simulations of alternative school enrollment schemes (both without and with private schools) for different values of the parameters. Our goal is to predict the effect of various policies that have been debated in France. We start with the analysis of the benchmark equilibrium of strict school zoning, which corresponds to the existing situation in Paris. We then analyze the impact of redesigning school catchment areas on school and residential social segregation and characterize the educational gains and losses incurred by pupils depending on their ability and family income. Finally, we assess the effect of switching from school zoning to school choice.

2.5.1 Simulations set-up

City and school catchment areas. In order to perform our simulations, we represent the city as a checkerboard constructed as follows (cf. figure 2.2):

- The city has $K = 100$ neighborhoods and $J = 16$ school zones.
- Each school zone is composed of blocks belonging to 4 neighborhoods that are entirely assigned to that particular school, 4 neighborhoods that overlap 2 school zones and 1 neighborhood that overlaps 4 school zones; the total number of different combinations of school zones and neighborhoods is equal to 144.
- The checkerboard is divided into 400 elementary cells of the same size, allowing us to define the relevant income bins to simulate the allocation of households to houses. The 144 combinations of neighborhoods and school zones have three different sizes: they may contain 1,2 or 4 elementary cells depending of the particular overlapping of neighborhoods and school zones. Figure 2.3 shows the precise layout of cells, blocks and neighborhoods in four typical school zones.

The quality of each neighborhood is randomly picked up from a standard log-normal distribution. The rank of a neighborhood's quality is positively correlated with the rank of the neighborhood's proximity to the city center, with the correlation coefficient being equal to λ .

Figure 2.2: *Layout of the city used in the simulations: a checkerboard of 100 neighborhoods and 16 school zones.*

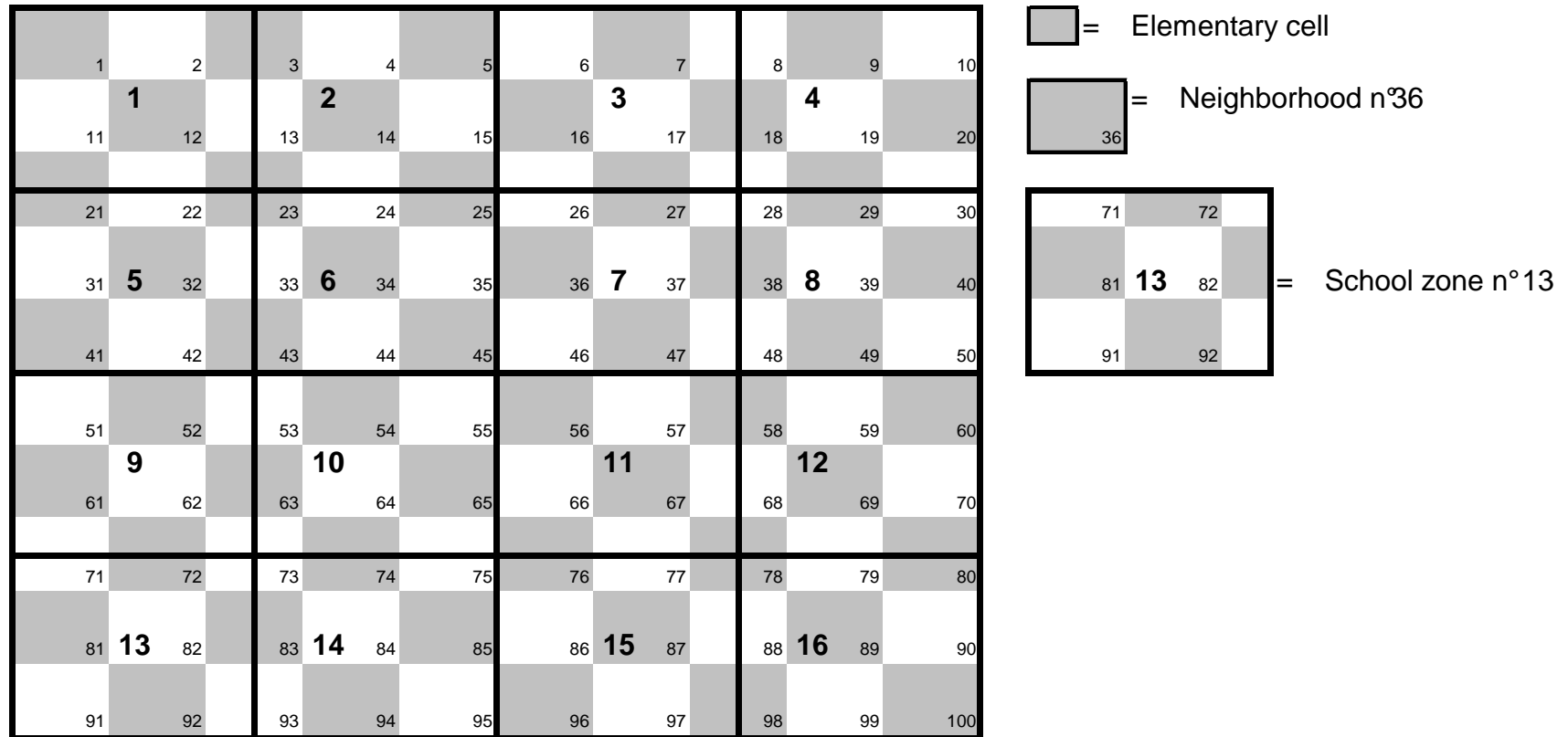
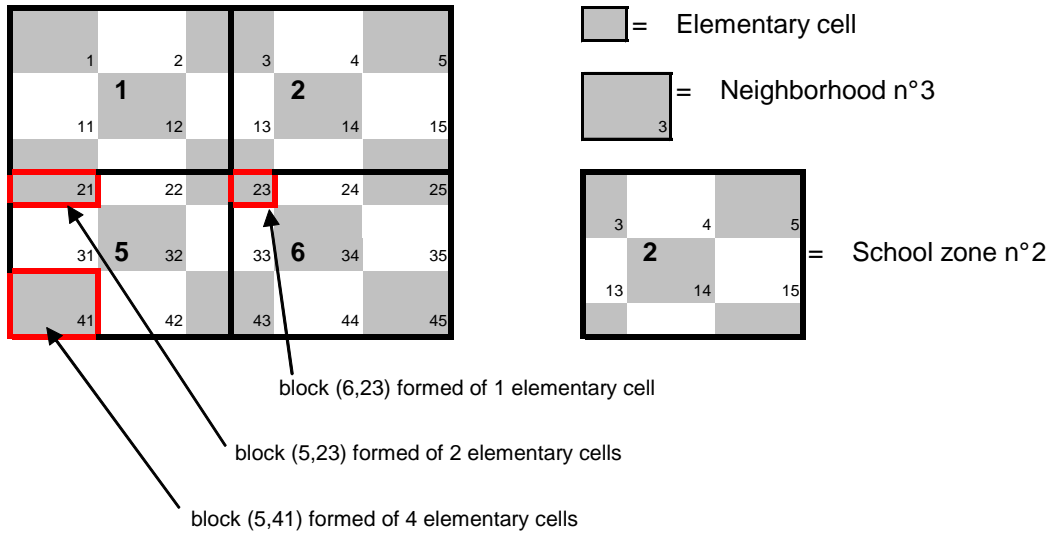


Figure 2.3: *Layout of school zones, neighborhoods, blocks and cells in the city.*



Families. The total number of families is set equal to 10 000. Income and abilities are picked up from a bivariate standard log-normal distribution with a correlation coefficient equal to ρ :

$$\begin{pmatrix} y_i \\ a_i \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$$

The utility function has the following specification:

$$U(c_i, h_{jk}) = \alpha \ln(c_i) + h_{jk}$$

where c_i is the private consumption of family i and h_{jk} is the desirability of the block B_{jk} in which it resides.

Parameters of the model. The model has four parameters:

- ρ : the correlation coefficient between family income and child ability.
- λ : the correlation coefficient between neighborhood quality and the proximity to the city center.

- β : the taste for school quality, i.e. the weight on school quality $\overline{\ln a^j}$ in the total desirability h_{jk} of block B_{jk} , relatively to the weight on the exogenous quality q_k of neighborhood k .
- s_0 : the secession threshold for the creation of a private school. In order for a private school to appear, the desirability to secede s_k of families living in neighborhood k must be superior or equal to s_0 .

2.5.2 Alternative school admission rules

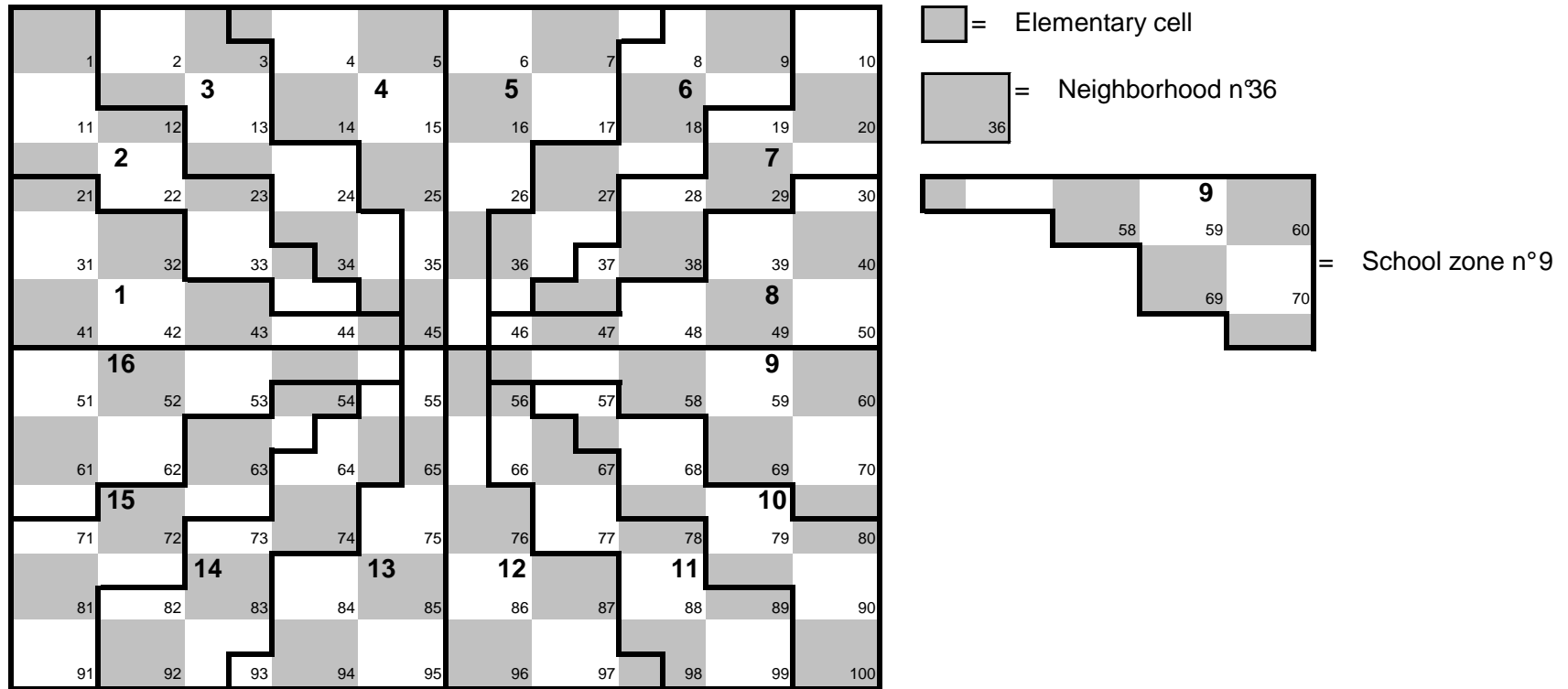
Before explaining the simulation procedure, we need to define precisely the school enrollment policies that we want to analyze and specify the main outcomes of interest. We focus on the segregation outcomes and on the distribution of educational gains and losses incurred by pupils from these policies. In addition, we investigate changes in school stratification that arise when private schools are introduced in the model.

Three alternative school enrollment schemes. We start with a benchmark equilibrium of **strict school zoning** where each child is assigned to a school on the basis of her address of residence.

Departing from this equilibrium, we then implement **school zone reassignments**. This policy consists in redesigning school catchment areas in order to change the composition of schools. We simulate a reassignment scheme that initially reduces stratification by abilities in all schools. Figure 2.4 shows how boundaries will be redrawn in our simulations. The original square-shaped zones are changed to slice-shaped zones radiating from the city center. Given the correlation between neighborhood quality and proximity to the center, this particular form of reassignment tends to homogenize school zones in terms of the intrinsic quality of neighborhoods.

Finally, departing from the strict school zoning equilibrium, we simulate the impact of implementing **open enrollment**. This policy consists in allowing parents to apply to the public school of their choice. Given our assumption that school principals have the right to choose their students and seek to maximize the average ability in their school, the resulting schooling equilibrium is completely stratified by ability. This assumption reflects the empirical evidence on school choice when no specific

Figure 2.4: *Simulating school zone reassignments: the new pattern of school attendance boundaries.*



admission rules are imposed. For instance, Hsieh and Urquiola (2006) find that the development of a voucher program in Chile yielded an increase in ability stratification, with higher ability children leaving the public schools for the private sector. However, school choice policies could be designed to restrict this type of stratification effects, by introducing admission rules in the form of lotteries for oversubscribed schools or specific admission criteria (preference given to children with low parental income or living closest to the school, etc.). We do not investigate these possibilities in our simulations and stick to a “cream skimming” version of school choice. Similarly, we do not model potential efficiency gains that could arise from increased competition among schools, as evidence on the subject is mixed²¹.

Segregation outcomes. Changes in school policies will affect both residential and educational segregation. In our simulations, we consider two measures of segregation:

- **Residential segregation:** we measure residential segregation at the school zone level by computing the ratio of the variance of the school-level average log income over the total variance of log income in the city population:

$$Seg_{res} = \frac{\text{Var}(\overline{\ln y^j})}{\text{Var}(\ln y_i)}$$

The higher the ratio, the higher the residential segregation between school zones (the coefficient is equal to zero when there is no segregation).

- **School segregation:** we measure educational segregation across schools by computing the ratio of the variance of the school-level average log ability over the total variance of log ability in the population:

$$Seg_{ab} = \frac{\text{Var}(\overline{\ln a^j})}{\text{Var}(\ln a_i)}$$

The interpretation of this ratio is similar to that of residential segregation.

²¹In a much debated paper, Hoxby (2000) uses natural boundaries such as rivers to identify exogenous variations in choice and shows that cities which have a larger number of school districts also have more productive schools. However, Rothstein (2006) calls into question the existence of such effects after replicating Hoxby’s estimations.

Distributional analysis of educational gains and losses. Under the assumption that peer effects are linear in means, changes in school policies will have no effect on aggregate educational achievement. However, alternative school enrollment policies are likely to affect pupils differently depending on their parental income and individual ability. We therefore focus on the distribution of educational gains and losses along these two dimensions.

The welfare implications of school enrollment policies are not limited to these educational effects, since housing prices are also affected by the reforms. In a context of local funding of schools through income or property taxes, it would be important to take into account the wealth effects of school policies that arise through changes in voting over school spending (see Epple and Romano, 2003). However, as our model does not include educational spending as an input in the school production function (the French school system being centrally financed), we abstract from these other welfare effects and concentrate exclusively on peer effects.

More precisely, we use two measures of educational gains:

- **The average peer ability change by decile of ability and income:** The individual gain (or loss) induced by a school enrollment policy change for a given child is denoted Δ_i and is equal to the difference between the average log ability in the school j' she attends after the implementation of the new policy and that of the school j she was previously assigned to:

$$\Delta_i = \overline{\ln(a)}^{j'} - \overline{\ln(a)}^j$$

The average peer group change by decile of ability and income is computed as the average Δ_i for all children.

- **The proportion of “winners” by decile of ability and income:** we compute the proportion of net “winners” in each decile of ability and income. A child will win (respectively lose) from a school enrollment policy change if the average ability of her schoolmates increases (respectively decreases). Note that the proportion of winners is not necessarily equal to 1/2, since the size of the effects can be asymmetric. For example, the average proportion can be inferior

to 1/2 if a small fraction of high-ability children becomes concentrated in a school as a result of the policy, leaving all children in the other schools worse off.

Educational gains are computed taking the initial school zoning equilibrium (with and without private schools) as the reference point. We do not specify a social choice function, but the two above measures of educational gains give an idea of the relative weight that should be put on low-income children if one wanted to favor this group when choosing among several policies, knowing that the net gains over the whole population of pupils are zero because of the linear-in-means specification of peer effects.

Specific outcomes for private schools. In order to study how the introduction of private schools affects our results, we compute three additional indexes:

- **the percentage of children attending private schools;**
- **the difference in the log average ability of children in public and private schools;**
- **the segregation by ability within both public and private schools.**

These outcomes are useful to assess the extent to which private schools cream-skin the best students from public schools.

2.5.3 The simulation algorithm

In order to estimate the impact of various school zoning policies, we perform a series of simulations using different sets of values for the parameters of interest. Residential choice and the allocation of pupils to the various schools in equilibrium is computed using an algorithm²² that consists of several steps and loops. An important feature of our algorithm is that it does not move each of the 10 000 households but works on elementary cells. We use the equilibrium property that households must be perfectly stratified by income to group them into 400 income bins corresponding

²²We coded the algorithm using SAS.

to the 400 elementary cells. Each cell is represented by an average income and average ability, with perfect ranking along these two dimensions. All the households belonging to a given cell will therefore always be moved together. Another important feature of the algorithm is that we do not need to calculate housing prices. Since the single crossing condition ensures that higher income households will always pay higher housing prices for blocks of higher desirability, we directly allocate households/cells according to the ranking of the desirability of blocks.

No private schools

Benchmark equilibrium: strict school zoning with no private schools (SZN).

The algorithm proceeds as follows:

- Step SZN1: We simulate 10 000 households characterized by parental income and child ability. The values of both characteristics are taken from a bivariate standard log-normal distribution, with a correlation coefficient equal to ρ .
- Step SZN2: We simulate 100 neighborhoods, randomly distributed throughout the city, under the constraint that the correlation between the rank of their intrinsic quality and the rank of their distance to the city center is equal to λ .
- Step SZN3: The initial allocation of households is the one that would arise in the absence of school zoning. Households are grouped into 400 income bins and are allocated to the different neighborhoods so that the ranking of income bins matches the ranking of the intrinsic quality of these neighborhoods. Note that this initial allocation will determine the equilibrium allocation: a different initial spatial allocation of households would yield a different allocation in equilibrium. However, our choice for the initial allocation can be justified as a starting point, since it corresponds to the equilibrium that arises when there are no school zones.
- Step SZN4: Square-shaped school zones are defined, each one grouping a total of 25 elementary cells. The initial allocation of households to the different neighborhoods yields a school quality for every school zone (calculated as the average log ability of pupils residing in the school's catchment area). The

exogenous intrinsic quality of neighborhoods and the endogenous quality of schools give rise to a unique ranking for the desirability of the 400 elementary cells, for a given value of parameter β (the taste for school quality).

- Loop SZN: The 400 income bins are reallocated to the different cells so that the ranking of income bins matches the ranking of elementary cell desirability. This again modifies the quality of schools, initiating a new reallocation of income bins to the various cells. The process is repeated until it converges to an equilibrium. An equilibrium is reached when the spatial allocation of income bins in period t is the same as the previous period's allocation. Depending on the values of the different parameters, an equilibrium is obtained within 10 to 20 successive iterations.

School zone reassignment with no private schools (SRN). The equilibrium obtained under strict school zoning with no private schools (SZN) serves as the starting point for the simulation of school zone reassignment. The algorithm proceeds as follows:

- Step SRN1: The initial allocation of income bins on the checkerboard is the one obtained in equilibrium under strict school zoning, for given values of the parameters λ , ρ , β and an initial spatial allocation of neighborhood quality throughout the city. The initial square-shaped school zones are replaced by new slice-shaped ones. The initial allocation of income bins defines the initial quality of these newly created school zones, which together with the neighborhood intrinsic qualities defines the new desirability of the various cells.
- Loop SRN: The 400 income bins are then reallocated to the different cells until an equilibrium is found, following the same process as in loop SZN.

Open enrollment with no private schools (OEN). The equilibrium obtained under strict school zoning with no private schools (SZN) serves as the starting point for the simulation of open enrollment. The algorithm proceeds as follows:

- Step OEN1: The initial allocation of income bins on the checkerboard is the one obtained in equilibrium under strict school zoning, for given values of the

parameters λ , ρ and β and an initial allocation of neighborhood quality throughout the city. School zones are then suppressed and replaced by open enrollment. Under that scheme, pupils are univocally allocated to the various public schools on the basis of their ability. Hence, in equilibrium, each school enrolls pupils belonging to a same bin of ability.

- Step OEN2: under the assumption that demand for desirability is invariant to its level (see subsection 2.3.1.1), the equilibrium allocation of households to the various cells under open enrollment is identical to the equilibrium obtained under strict school zoning when the parameter β (the taste for school quality) is set to 0. In equilibrium, households are thus allocated to the different neighborhoods so that the ranking of income bins matches the ranking of the intrinsic quality of neighborhoods.

Introducing private schools

The three equilibria defined above (strict school zoning, school zone reassignment and open enrollment) are extended to account for the endogenous emergence of private schools. Residential choice and the allocation of pupils to public and private schools in equilibrium is then computed using an algorithm that involves several nested loops.

In our model (strict school zoning and school zone reassignment), private schools consist of local clubs formed by parents at the neighborhood level. In order for a private school to be created, it must be the case that the average difference between the log ability of the pupils residing in a given neighborhood and the average log ability in the surrounding public schools is higher than an exogenously predetermined threshold s_0 . The way private schools appear and disappear is described below.

Strict school zoning with private schools (SZP). The algorithm that accounts for the introduction of private schools in the strict school zoning scheme proceeds as follows:

- Steps SZP1 to SZP3 are identical to steps SZN1 to SZN3.

- Step SZP4: Square-shaped school zones are defined and no private school initially exists. The initial allocation of households to the different neighborhoods yields a public school quality for every school zone (calculated as the average log ability of pupils living in the public school’s catchment area).
- Step SZP5: Given the initial quality of public schools, the incentive to create a private school is evaluated separately for each neighborhood. Households living in neighborhood k will decide to “secede” and create their own private school if the difference between the average log ability in neighborhood k and the weighted average of the initial quality of the surrounding public schools (the weight of each surrounding public school being equal to the number of households in neighborhood k that live within its corresponding catchment area) exceeds the threshold s_0 .
- Step SZP6: After having calculated the quality of the newly created private schools, the algorithm updates the quality of the various public schools to account for the fact that some pupils are now enrolled in private schools. Together with the intrinsic quality of neighborhoods, these public and private school qualities determine a unique ranking of the desirability of the 400 elementary cells, given the value of parameter β (the taste for school quality). The uniqueness of cell desirability ranking derives directly from the particular way in which we model the emergence of private schools: in any cell belonging to a seceding neighborhood, private school quality will always surpass the quality of the local public school so that all households, if offered the possibility of living in such a neighborhood, would always prefer sending their children to the local private school rather than then to the local public school (assuming that existing public and private school qualities are held constant and considered as given by households when making their residential choices).
- Loop SZP consists of three successive subloops:
 - Subloop SZP1: The 400 income bins are spatially reallocated to the different cells so that the ranking of income bins matches the ranking of elementary cell desirability. Depending on whether the cell has seceded

and created its own private school in the previous period or not, children belonging to a given parental income bin will end up in a public or a private school.

- Subloop SZP2: The new allocation of households modifies the ability composition of both public and private schools created in the previous period. Given these new ability compositions, the sustainability of existing private schools is evaluated in the following way: every seceding neighborhood evaluates whether, should it dissolve its private school and send its pupils back to the surrounding public schools (other private schools being assumed to remain sustainable), it would still have the incentive to secede (as defined in step SZP6). If it does not, then the school is dissolved and its pupils are sent back to the surrounding public schools. If it does, then the private school is considered sustainable and enrolls the neighborhood's pupils. In practice, the simulations show that once created, private schools hardly ever disappear²³.
- Subloop SZP3: Once the sustainable private schools are identified, the ability composition of public and private schools is updated. Together with the intrinsic quality of neighborhoods, this determines a new ranking of cell desirability and initiates a new spatial reallocation of income bins to the various cells. Subloops SZP1 to SZP3 are iterated until they converge to an equilibrium. An equilibrium is reached when the spatial allocation of bins in period t is the same as the previous period's allocation and the seceding neighborhood creating their own private schools are identical from one period to the next.

School zone reassignment with private schools (SRP). The equilibrium obtained under strict school zoning with private schools (SZP) serves as the starting point for the simulation of school zone reassignment. The algorithm proceeds as follows:

²³This is because newly created private schools will tend both to attract wealthier households and to reduce the quality of the surrounding public schools, thus increasing the average ability differential between the seceding neighborhood and the surrounding public schools.

- Step SRP1: The initial allocation of income bins on the checkerboard is the one obtained in equilibrium under strict school zoning with private schools (for given values of the parameters λ , ρ , β , the seceding threshold s_0 and the initial spatial allocation of neighborhood quality throughout the city). The initial squared-shaped school zones are replaced by the new slice-shaped ones. Pupils enrolled in private schools are supposed to temporarily remain enrolled in their private schools. Therefore, the initial quality of the newly created public schools zones is purely determined by the initial allocation of income bins belonging to non-seceding neighborhoods.
- Step SRP2: Given the new ability composition of public schools, the sustainability of existing private schools is evaluated in the same way as in subloop SZP2. If a private school does not meet the sustainability criterion, then the school is dissolved and its pupils are sent back to the surrounding public schools. If it does, then the private school is considered sustainable and enrolls the neighborhood's pupils. In practice, the simulations show that some of the private schools that emerged under strict school zoning disappear after the reassignment of school zones.
- Loop SRP consists of three successive subloops that are identical to subloops SZP1 to SZP3 in the algorithm used to compute the equilibrium under strict school zoning with private schools (SZP). These subloops are iterated until they converge to an equilibrium. An equilibrium is reached when the allocation of bins in period t is the same as the previous period's allocation and when the seceding neighborhoods creating their own private schools are identical from one period to the other.

Open enrollment with private schools (OEP). The equilibrium obtained under strict school zoning with private schools (SZP) serves as the starting point for the simulation of open enrollment with private schools. As explained in the discussion of how private schools are introduced in the model (section 2.4), no private school is sustainable under an open enrollment scheme for credible values of the parameters. The computation of the equilibrium under open enrollment with private schools is

thus identical to the computation of the equilibrium under open enrollment with no private schools (OEN).

Random assignment of neighborhood quality

Conditional on the values taken by the parameters λ , ρ , μ , s_0 and the initial spatial allocation of neighborhood quality throughout the city, the different algorithms defined above will converge to a unique equilibrium for every distinct school enrollment policy.

In order for our prediction to be independent of a particular allocation of neighborhood quality on the checkerboard, we perform 100 simulations for each set of parameters λ , ρ , μ , S_0 by randomly changing the initial geographic allocation of neighborhood quality. We then average the equilibrium values of the parameters of interest across the 100 replications.

2.5.4 An example

We perform several simulations using the following values of the parameters: $\rho = 0.5$; $\lambda = 0.5$; $\beta = 0.5$; $s_0 = 0.4$. Figure 2.5 shows a simulated distribution of the logarithm of intrinsic neighborhood quality, household income and pupil ability.

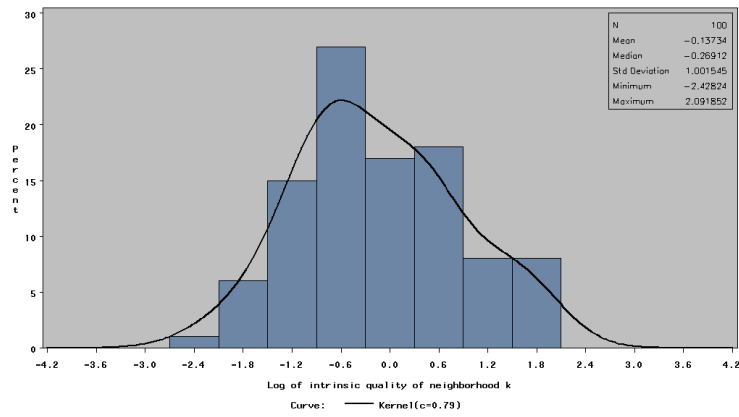
The initial spatial allocation of neighborhoods by quintile of intrinsic quality is displayed in figure 2.6. Neighborhoods with the highest intrinsic quality tend to be concentrated in the center of the checkerboard while low-quality neighborhoods tend to be located in the periphery. But as the correlation between neighborhood quality and proximity to the city center is not equal to one, the checkerboard also exhibits some good neighborhoods in the periphery and lower quality ones near the center.

The benchmark equilibrium: strict school zoning without private schools

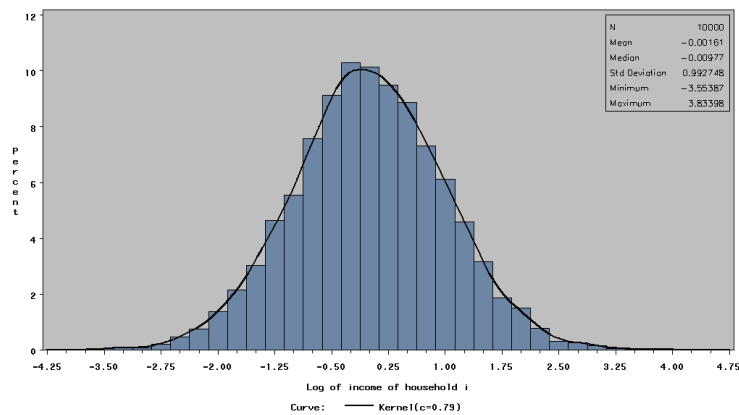
The equilibrium spatial distribution of the quintiles of school quality and equilibrium desirability of cells is shown in figure 2.7. Better schools tend to be located in the better neighborhoods close to the city center. As expected, the equilibrium desirability of blocks partly reflects the spatial distribution of neighborhood qualities. As the equilibrium is perfectly stratified by income, the wealthiest households live in

Figure 2.5: *Distribution of the log of intrinsic neighborhood qualities, the log of family income and the log of pupil ability in the sample of 10,000 households used in the simulations.*

(a) Log of intrinsic neighborhood quality



(b) Log of family income



(c) Log of pupil ability

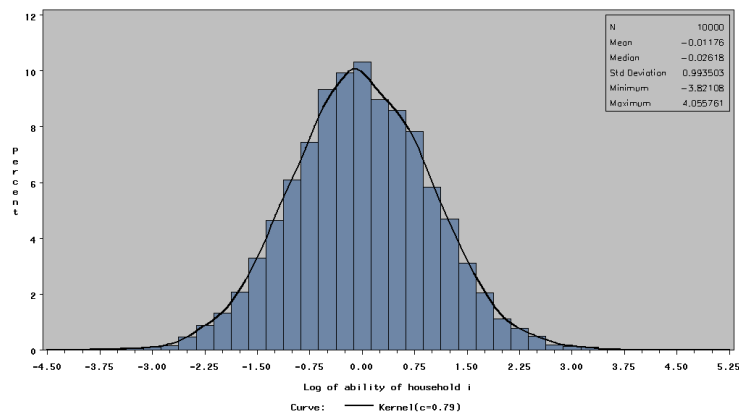
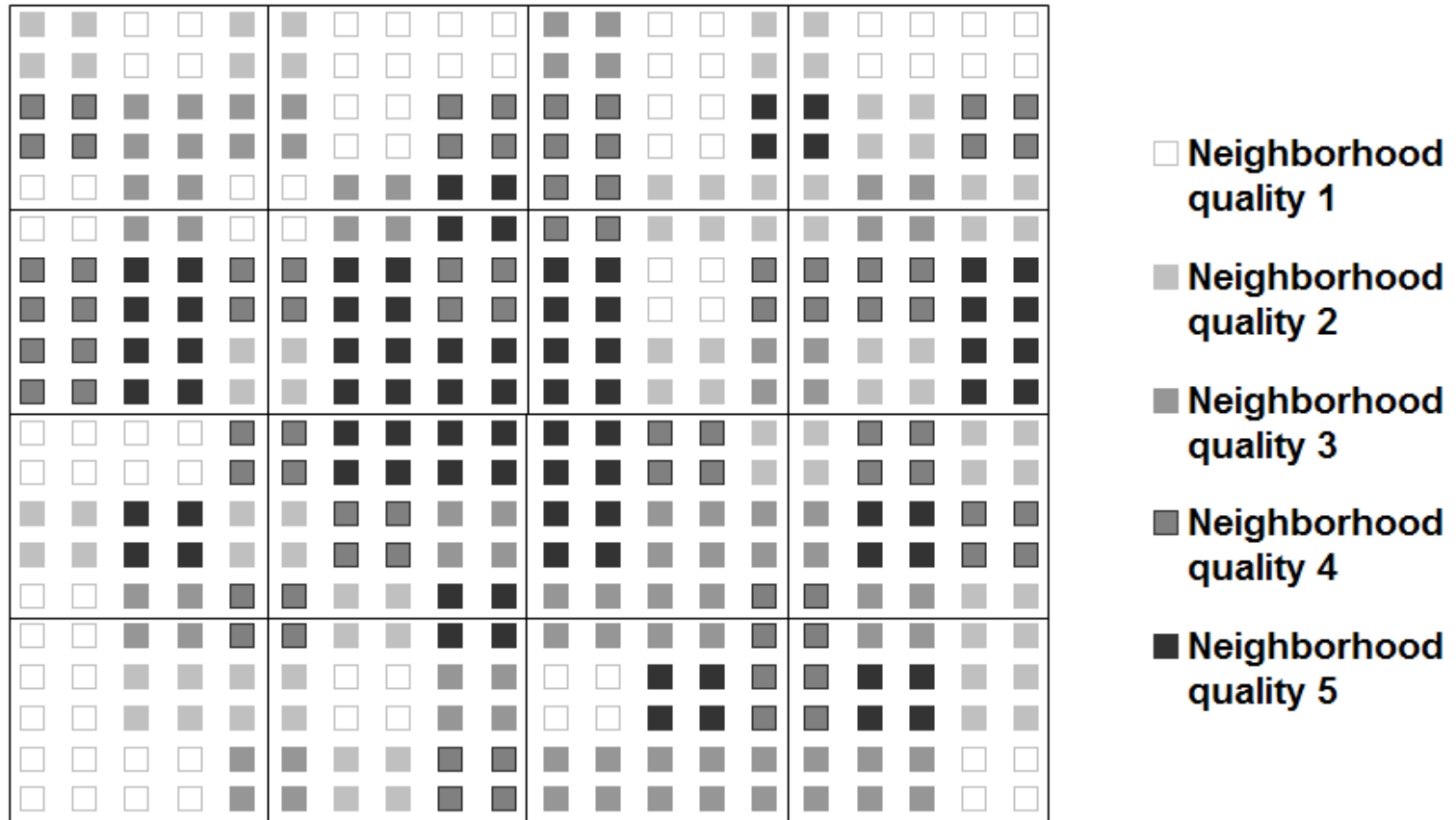


Figure 2.6: Simulation example: spatial distribution of neighborhood intrinsic quality.

Quintiles of the intrinsic quality of each of the 100 neighborhoods
 Strict school zoning with no private schools ($\rho=0.5$; $\lambda=0.5$; $\beta=0.5$)



the most desirable blocks. The blocks located near the center, with a higher average quality, tend to concentrate a wealthier population than neighborhoods located in the periphery. However, it might be the case that some blocks with a poor intrinsic quality do not house the poorest households, because the quality of the local middle school is good. In our example, this is the case for some of the neighborhoods located in school zones 5 and 9 (in the western part of the checkerboard).

Figure 2.8 shows the equilibrium correlation between mean income in the cells (which form the blocks) and the various components of block desirability. Households are perfectly sorted by income and block desirability, but this is not necessarily the case for each component of block desirability. On average, blocks with higher income groups tend to be located in neighborhoods with higher intrinsic quality and better schools, but some households prefer to reside in lower quality neighborhoods where the local public school is very good. Conversely, school zones mix several level of income, because of the heterogeneity in neighborhood quality in each zone. With the values of the parameters chosen for the simulation, residential segregation is equal to 0.5 and ability segregation between schools is 0.12.

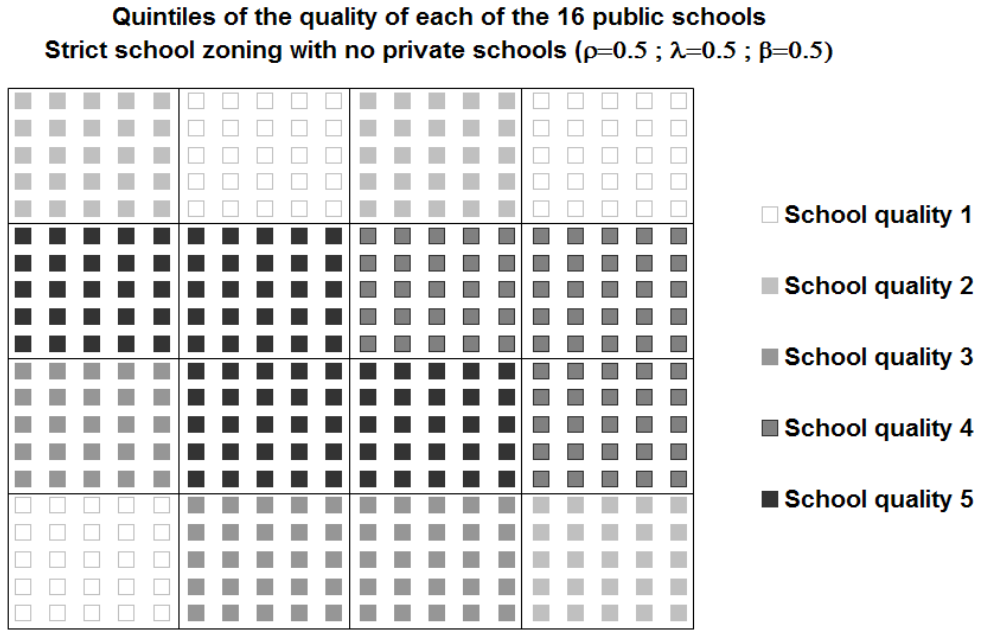
School reforms without private schools

Two types of school policy reforms are simulated: open enrollment and school zone reassignments. Changes in segregation outcomes and distributional gains are reported in tables 2.1 and 2.2.

Open Enrollment. As this policy disentangles residential and schooling choices, the spatial distribution of families in the city is uniquely determined by the intrinsic quality of neighborhoods. The wealthiest households live in the best neighborhoods (see figure 2.6). As a result, residential segregation, measured within the areas defined by the previous school zones, decreases significantly. In terms of educational gains, high-ability children unambiguously benefit from joining peers of higher ability, whereas low-ability children experience a decrease in peer quality. Given the positive correlation between ability and income, children of families with income below the median are on average worse off under this scheme (see table 2.2).

Figure 2.7: Simulation example: equilibrium spatial distribution of school quality and final cell desirability under strict school zoning.

(a) School quality



(b) Final cell desirability

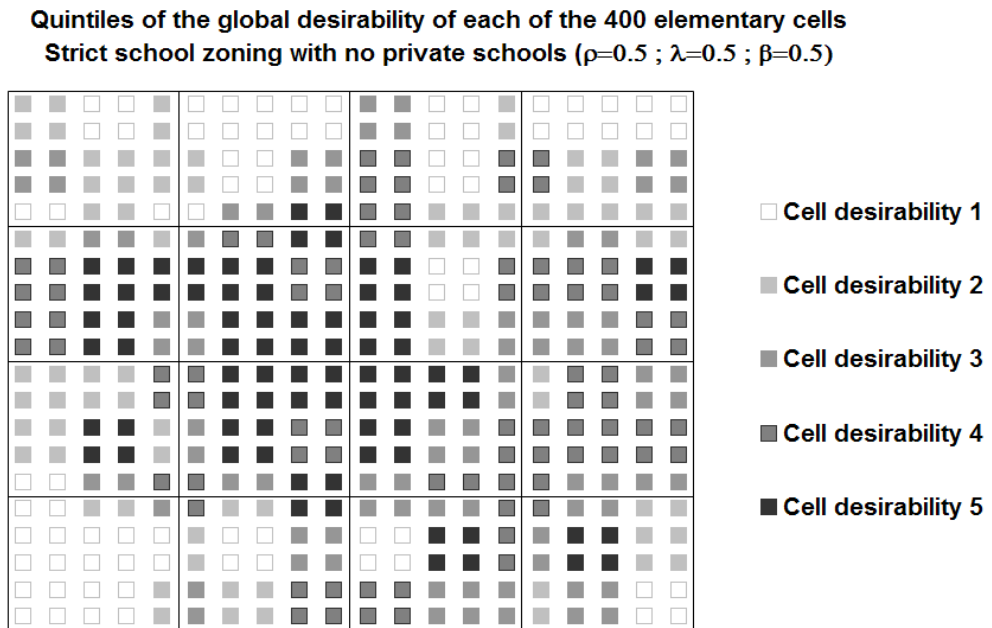


Figure 2.8: *Simulation example: relationship between income and cell desirability, intrinsic neighborhood quality and mean ability in the local public school under strict school zoning.*

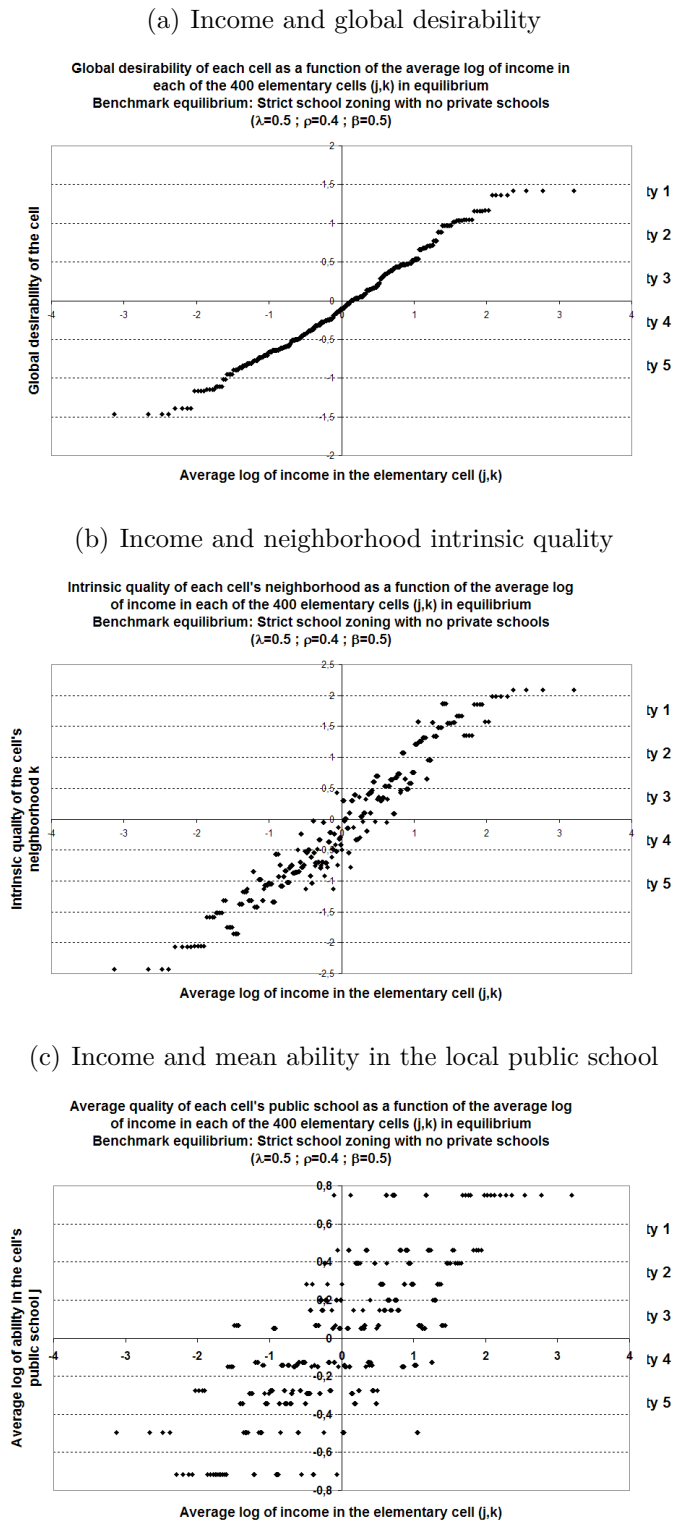


Table 2.1: *Simulation example: residential and educational segregation outcomes from alternative school enrollment policies.*

PARAMETER VALUES: $\lambda=0.5$. $\rho=0.5$. $\beta=0.5$. $s_0=0.4$								
	Residential segregation (log income)	Educational segregation (log ability)	Residential segregation (log income)	Educational segregation within public schools: log abilities	Educational segregation within private schools	Average log ability in public schools	Average log ability in private schools	Private Enrollment (%)
	<i>Without private schools</i>			<i>With private schools</i>				
Benchmark equilibrium: strict school zoning (s.e.)	0.49 (0.04)	0.12 (0.01)	0.35 (0.05)	0.12 (0.01)	0.11 (0.02)	-0.16 (0.02)	0.61 (0.05)	0.20 (0.03)
School zone reassignments: initial level (s.e.)	0.18 (0.05)	0.045 (0.012)	0.16 (0.04)					
School zone reassignments: final level (SE)	0.32 (0.06)	0.08 (0.02)	0.17 (0.05)	0.11 (0.01)	0.11 (0.02)	-0.21 (0.03)	0.55 (0.03)	0.27 (0.03)
Open enrollment (s.e.)	0.26 (0.03)	0.98 (0.00)	0.26 (0.03)					

Notes: λ : correlation between neighborhood quality and proximity to the city center; ρ : correlation between family income and child ability; β : parental taste for school quality; s_0 : seceding threshold to create a private school. Results from 100 simulations performed on random initial spatial distribution of neighborhood quality, given λ . Standard errors of the estimates are in parentheses.

Table 2.2: *Simulation example: distribution of educational gains induced by school zone reassignments and open enrollment in the presence or absence of a private sector.*

PARAMETER VALUES: $\lambda=0.5$. $\rho=0.5$. $\beta=0.5$. $s_0=0.4$

PERCENTAGE OF WINNERS:

<i>Without Private Schools</i>	Decile	1 th dec.	2 nd dec.	3 rd dec.	4 th dec.	5 th dec.	6 th dec.	7 th dec.	8 th dec.	9 th dec.	10 th dec.	Total
School zone reassignments (SRN)	by income	0.72	0.64	0.61	0.58	0.51	0.48	0.46	0.56	0.45	0.24	0.53
	by ability	0.62	0.58	0.57	0.56	0.54	0.52	0.50	0.49	0.47	0.42	0.53
Open enrollment (OEN)	by income	0.32	0.42	0.42	0.46	0.46	0.51	0.53	0.59	0.60	0.67	0.50
	by ability	0.00	0.00	0.04	0.19	0.37	0.62	0.80	0.95	1.00	1.00	0.50

<i>With Private Schools</i>	Decile	1 th dec.	2 nd dec.	3 rd dec.	4 th dec.	5 th dec.	6 th dec.	7 th dec.	8 th dec.	9 th dec.	10 th dec.	Total
School zone reassignments (SRP)	by income	0.37	0.51	0.52	0.51	0.47	0.45	0.52	0.55	0.38	0.11	0.44
	by ability	0.44	0.46	0.46	0.47	0.47	0.45	0.43	0.43	0.41	0.35	0.44
Open enrollment (OEP)	by income	0.38	0.49	0.47	0.49	0.50	0.52	0.53	0.56	0.52	0.49	0.50
	by ability	0.00	0.00	0.10	0.25	0.39	0.63	0.75	0.87	0.97	0.99	0.50

AVERAGE EDUCATIONAL GAINS:

<i>Without Private Schools</i>	Decile	1 th dec.	2 nd dec.	3 rd dec.	4 th dec.	5 th dec.	6 th dec.	7 th dec.	8 th dec.	9 th dec.	10 th dec.	Total
School zone reassignments (SRN)	by income	0.13	0.09	0.09	0.06	0.01	-0.02	-0.06	0.00	-0.07	-0.23	0.00
	by ability	0.07	0.04	0.03	0.03	0.01	0.00	-0.02	-0.03	-0.05	-0.08	0.00
Open enrollment (OEN)	by income	-0.41	-0.15	-0.19	-0.07	-0.09	0.03	0.07	0.21	0.24	0.36	0.00
	by ability	-1.48	-0.93	-0.58	-0.32	-0.13	0.12	0.32	0.59	0.91	1.50	0.00

<i>With Private Schools</i>	Decile	1 th dec.	2 nd dec.	3 rd dec.	4 th dec.	5 th dec.	6 th dec.	7 th dec.	8 th dec.	9 th dec.	10 th dec.	Total
School zone reassignments (SRP)	by income	-0.07	0.02	0.02	0.01	-0.03	-0.04	-0.01	0.02	0.06	0.03	0.00
	by ability	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.00
Open enrollment (OEN)	by income	-0.24	0.01	-0.08	0.01	0.00	0.07	0.09	0.15	0.03	-0.04	0.00
	by ability	-1.37	-0.85	-0.52	-0.28	-0.10	0.12	0.29	0.54	0.83	1.35	0.00

Notes: λ : correlation between neighborhood quality and proximity to the city center; ρ : correlation between family income and child ability; β : parental taste for school quality; s_0 : seceding threshold to create a private school. Results from 100 simulations performed on random initial spatial distribution of neighborhood quality, given λ . Standard errors of the estimates are in parentheses.

School zone reassignments. School catchment areas are redrawn in the form of slices radiating from the center, as shown in figure 2.4. The aim of school zone reassignments is to reduce segregation by ability in schools, but it also affects residential segregation since the new school catchment areas tend to be more mixed in term of intrinsic neighborhood quality. With the parameter values chosen for the simulation, the initial schooling segregation is reduced by two thirds after the re-assignment. However, some families react to the policy and endogenously recreate residential segregation in the new equilibrium.

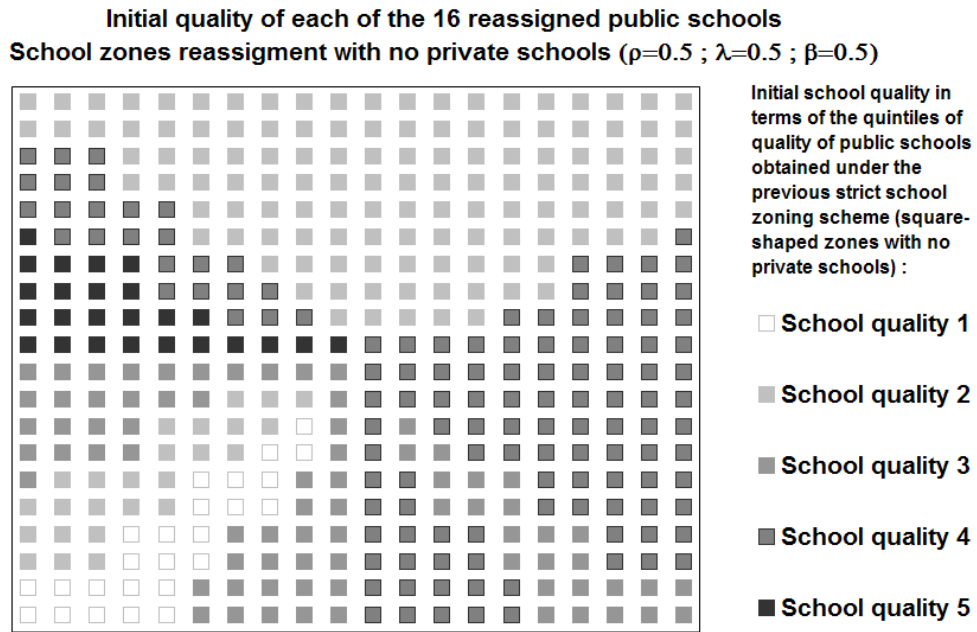
Figure 2.9 compares the initial spatial distribution of school quality when boundaries are redrawn with the final spatial distribution of school quality that arises in equilibrium. The difference between the two layouts is due to the parental reaction triggered by the policy. In order to compare these school qualities with the school qualities that emerged under the strict school zoning scheme (cf. figure 2.7(a)), we keep the quintile thresholds defined over the distribution of school qualities in the benchmark equilibrium. We observe that the number of very high quality schools has dropped from four in figure 2.7(a) to one in figure 2.9(b). Both final and residential levels of segregation are lower with the reassigned boundaries than in the benchmark equilibrium of strict school zoning. Children of lower income families experience positive gains from this policy, while children from higher income families experiences losses. If one assigns a high weight on children coming from low-income backgrounds in a social choice function, this policy would be clearly preferred to open enrollment. However, these results are dramatically modified when we introduce private schools, because the private sector offers an outside option for dissatisfied parents to opt out from the assigned local public school.

2.5.4.1 School reforms when private schools are introduced

Strict school zoning with private schools. Starting from the benchmark strict school zoning equilibrium, we now allow for the creation of private schools. Private schools are set up in the neighborhoods where the desirability to secede s_k is superior to the threshold 0.4. Figure 2.10 shows that, in equilibrium, private schools appear both in zones where the public school quality is high and in zones where the public

Figure 2.9: *Simulating school zone reassignments: comparing initial school quality to equilibrium school quality.*

(a) Initial school quality



(b) Equilibrium school quality

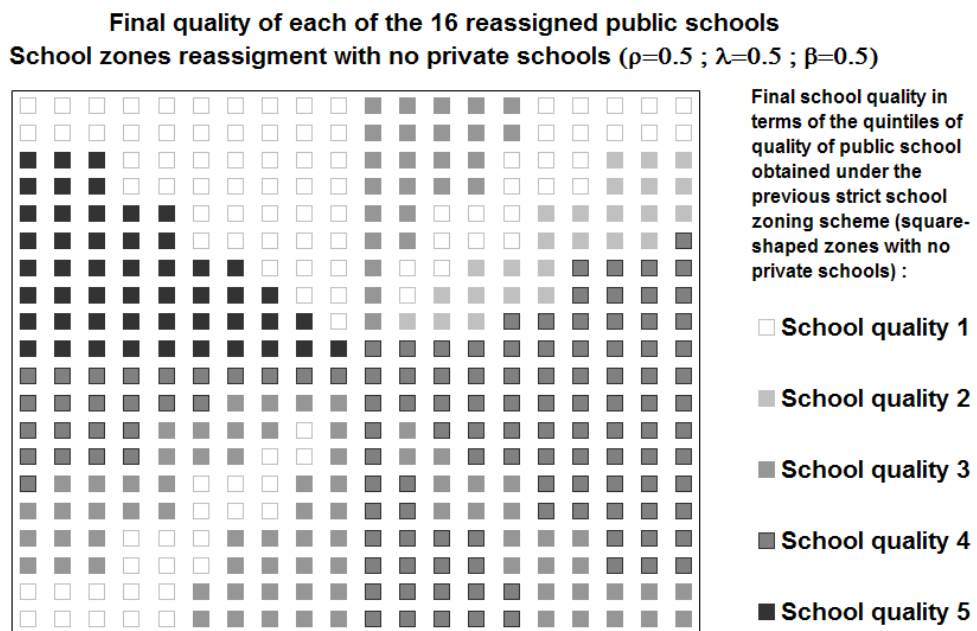
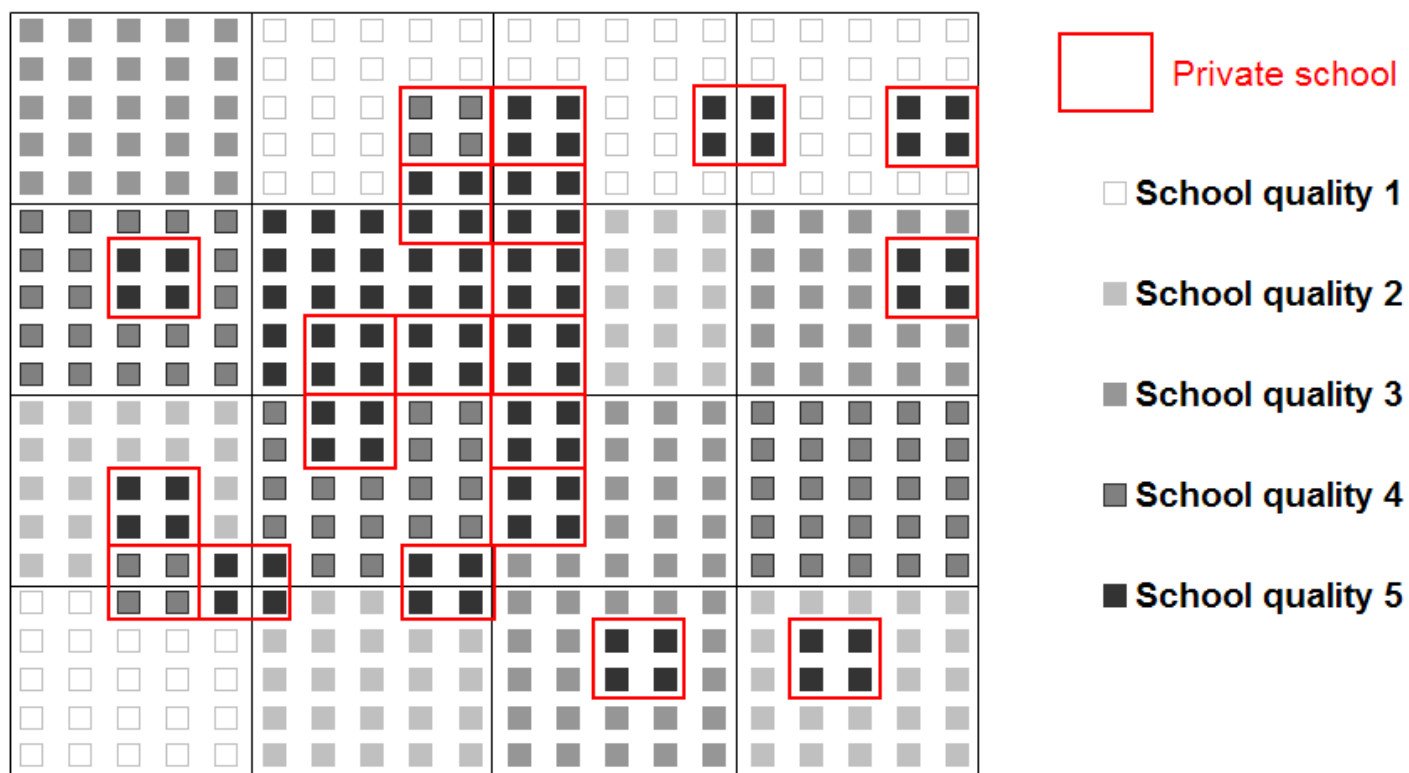


Figure 2.10: *Simulating strict school zoning with private schools: equilibrium public and private school quality.*

Quintiles of the quality of public and private schools
Strict school zoning with private schools ($\rho=0.5$; $\lambda=0.5$; $\beta=0.5$)



school quality is relatively low, with total private enrollment equal to 20%. In fact, private schools are located in the best neighborhoods of each school zone. Their cream-skimming effect, which is implied by the model, is very clear on the map, but it is not complete: private schools do not enroll all the higher ability students, but they cream-skin the best students groups in the school zones where they are created, implying a decrease in public school quality. The introduction of private schools also yields a decrease in residential segregation (cf. table 2.1).

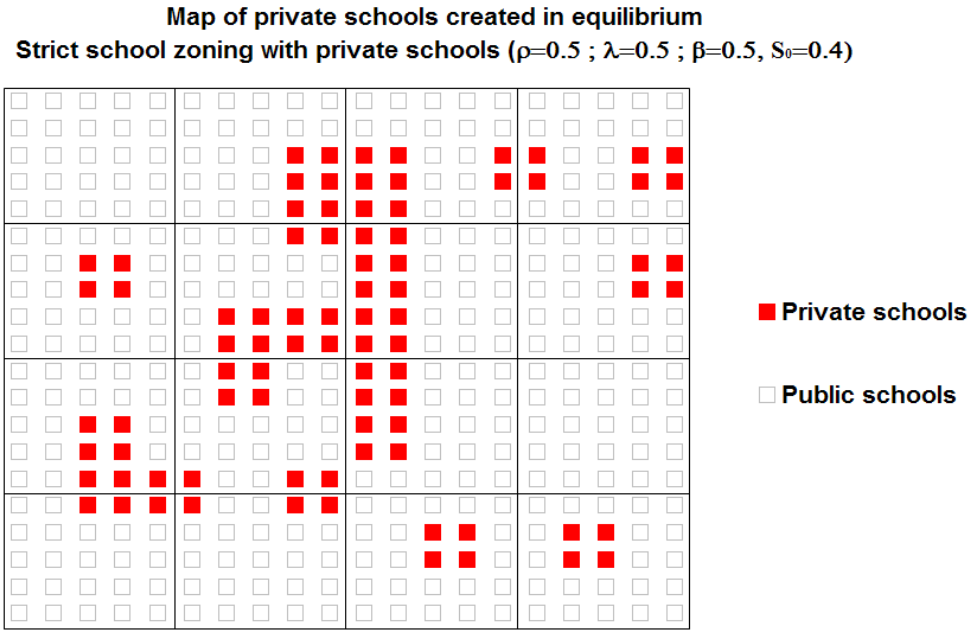
Open Enrollment. Given our assumptions and parameter values, private schools are no longer sustainable when open enrollment is introduced. The results are therefore similar to those described earlier for open enrollment without private schools. However, the distribution of educational gains must now be evaluated in reference to strict school zoning with private schools. Table 2.2 shows that the average gains and losses incurred at both ends of the income and ability distributions are somewhat smaller when private schools operate than when open enrollment is introduced without private schools. Indeed, low-income/low-ability children were already attending public schools of low quality under the strict zoning scheme and, conversely, high-income/high-ability children were already attending private schools with high-ability peers. Hence the effect of open enrollment on the characteristics of school peers is not as strong when this system replaces strict school zoning with private schools as when it replaces strict school zoning without a private sector.

School zone reassignments. Shifting school attendance boundaries initially leads some private schools to shut down, if the quality of the reassigned public school is sufficiently high. However, figure 2.11 shows that in equilibrium, the size of the private sector tends to *increase* after the reassignment, because the number of dissatisfied high-income/high-ability families rises when the quality of public schools is levelled. With the parameter values chosen for the simulation, the size of the private sector rises from 20% to 27%. As a result, the cream-skimming effect of private schools yields a fall in the average public school quality.

Figure 2.12 shows that some public schools may paradoxically end up with an even *lower* quality after the reassignment than under the previous design of school

Figure 2.11: Comparing the size of the private sector under strict school zoning, before and after the reassignments of school zones.

(a) Private schools before reassignments



(b) Private schools after reassignments

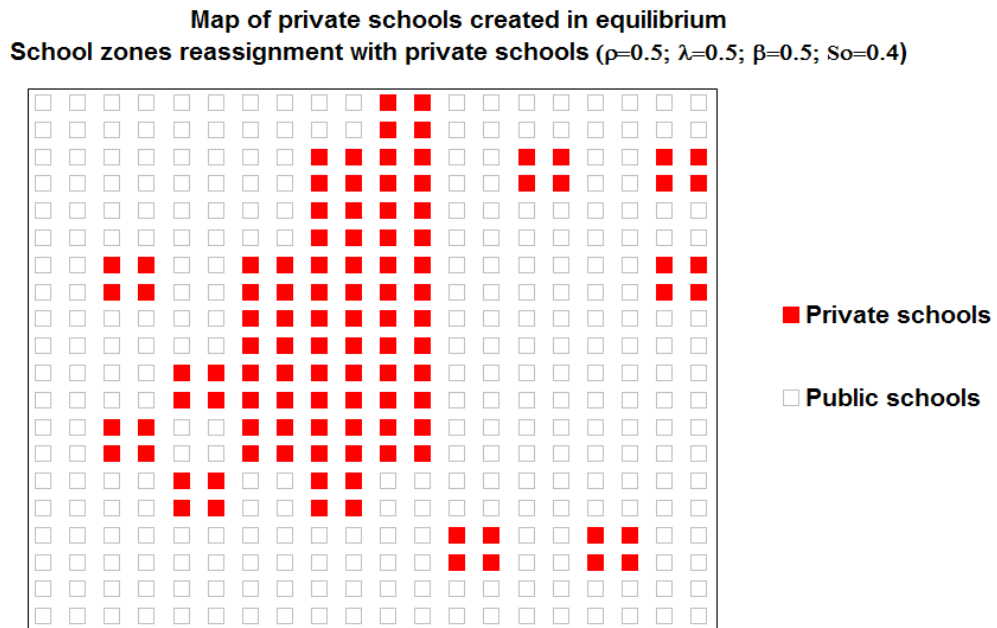
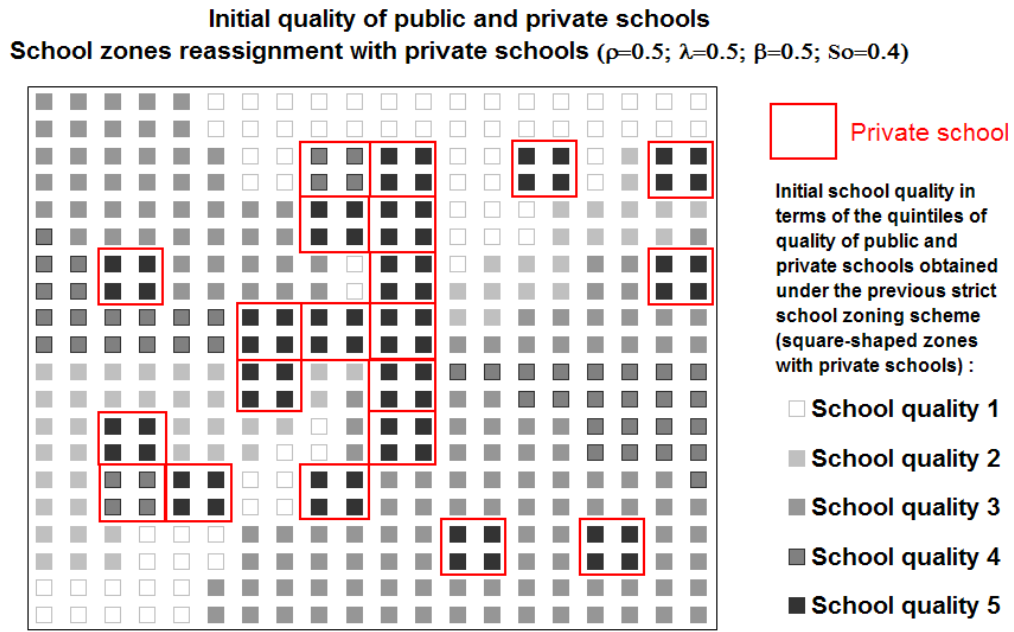
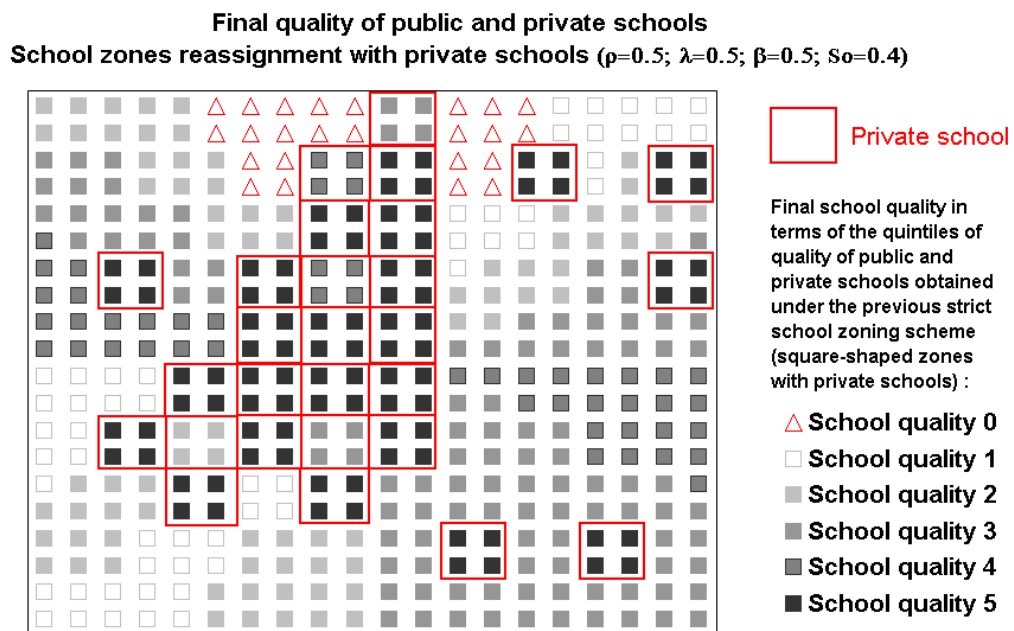


Figure 2.12: Comparing public and private school quality under strict school zoning immediately after the reassignments of school zones and in the final equilibrium.

(a) Initial quality after reassignments



(b) Equilibrium quality after reassignments



Note : "School quality 0" stands for schools which after the reassignment of school zones end up with a quality lower than the lowest performing school under the previous square-shaped strict school zoning scheme

catchment areas (see the school located in the northern part of the checkerboard, which has been classified of as “school quality 0” to indicate that the quality of the school falls *below* the lowest school quality observed before the reassignment). The gains for low-income/low-ability children are therefore wiped out, compared to the situation without private schools. At the upper end of the distribution, the higher income households who have escaped from the public sector as a result of the policy change enjoy better peers (cf. table 2.2).

This result suggests that although school zone reassignments would significantly reduce educational inequalities in a world without private schools, this policy may have the opposite effect if parents have the possibility to opt out of the public school system. Before turning to the calibration of the model using data from the city of Paris, let us investigate how the values of parameters affect the model’s predictions.

2.5.5 Changing the values of the parameters

Departing from our example where $\rho = 0.5$; $\lambda = 0.5$; $\beta = 0.5$; $s_0 = 0.4$, we study how the outcomes are affected when each of the four parameters is modified (for values ranging from zero to one) while the others are kept constant.

Benchmark equilibrium without private schools. Increases in ρ , λ or β leads to higher residential segregation under strict school zoning, the largest impact being that of β . However, stratification of schools by pupil ability is much more affected by an increase in ρ than by an increase in the two other parameters, since this strengthens the correlation between income and abilities.

School zone reassignments without private schools. An increase in λ , the correlation coefficient between neighborhood quality and distance to the city center, does not affect much final residential and ability segregation when school catchment areas are redrawn. As in the benchmark equilibrium, final residential segregation increases with β since a higher taste of families for schools raises the incentives to relocate after school zone reassignments. As a result, the gains in peer quality for low-income households decrease as β increases. High values of ρ yield high values of

ability segregation in the school zone reassignment equilibrium, but school segregation by ability is always less pronounced than in the strict school zoning benchmark equilibrium, so the benefits from school zone reassignments for low-income children tend to increase with ρ .

Open enrollment without private schools. The model predicts the same high level of ability segregation in open enrollment for any values of the parameters. Unsurprisingly, residential segregation is affected by λ , the correlation between neighborhood quality and the city center. Average gains are negative for children coming from families with income below the median, although the losses for this group tend to be smaller for higher values of λ . This can be explained by the fact that high-ability children from low-income backgrounds benefit more from the open enrollment policy, since they previously attended schools with a low average ability due to the high level of residential segregation. Similarly, higher values of β yield lower average losses for low-income groups, for a given ρ . On the contrary, higher values of ρ tends to increase losses for low-income groups, since it increases the heterogeneity in gains.

Introduction of private schools. As expected, the size of private enrollment varies dramatically with the size of the threshold s_0 . Increasing the private sector attenuates the effect of both school enrollment policies examined: gains from school zone reassignments for the low income groups decrease significantly whereas their losses decrease in open enrollment.

The effect of λ is interesting: a high initial correlation between neighborhood quality and distance to the city center limits the development of private schools, since high-income households tends to be more concentrated in the center and already enjoy good public schools. In this case, however, school zone reassignments trigger a higher increase in private school enrollment and the gains or losses of this policy are finally very small for all values of λ . Changes in β have the same qualitative impact, because a strong taste for school quality also induces a high level of residential segregation and an small initial size of the private sector.

The incentives to segregate induced by a high ρ increase the size of the private sector under strict school zoning. This significantly decreases the quality of public

schools. As a result, school zone reassignments no longer yield positive benefits for low-income children, since the quality of public schools is already low and the size of the private sector further increases.

Finally, for all values of the parameters, the presence of the private sector tends to level the gains and losses of an open enrollment policy across income levels. Note that when private schools operate, the first decile of the income distribution is nevertheless worse off under open enrollment than under strict school zoning (except when β is very low).

2.6 Calibrations for the city of Paris

In this section, we calibrate our model using data from the city of Paris to simulate the likely consequences of alternative school enrollment policies.

2.6.1 Calibration of the key parameters

We need to calibrate the four key parameters ρ , λ , β and s_0 . Among these parameters, only λ , the correlation between neighborhood quality and proximity to the center, can be estimated directly. The other coefficients are not directly observable and have to be inferred by matching the predictions of our model with what we observe in reality. In order to calibrate the model, we try to replicate three distinct features of the Parisian data:

- **The level of residential segregation at the school zone level:** We use data from the French 1999 census to retrieve information on the occupation of the household head at the most disaggregated residential level, since we do not have any direct information linking income and residence. We then impute income by occupational group, using the average log income of the household head for each occupational category calculated with the French Housing Survey. We find a residential segregation in the 0.3-0.4 range.
- **The level of segregation by scores in public schools:** We calculate the ratio of the variance in test scores using the database on *the Diplôme National*

du Brevet (the national exam taken by all pupils at the end of middle school) in year 2004 and find a ratio equal to 0.2.

- **Private school enrollment:** it is equal to 30% in Paris.

The correlation between neighborhood quality and proximity to the center (λ). In order to calculate λ , we need an approximation of neighborhood intrinsic quality q_k . We use housing prices in order to calculate it, since they reflect the quality of neighborhood attributes, but these prices have to be corrected for the capitalization of public school performance. We achieve this by using our empirical estimation of the impact of school performance on housing prices (in the next chapter). Taking the worst school zone as the reference, we calculate the implied capitalization of school quality for each school zone and subtract it from housing prices (after correcting for the characteristics of the flat). The intrinsic quality of neighborhood k is the average of these corrected price over all sales in the area:

$$\ln q_k = \overline{\ln p_{jk}} - \hat{\theta} \Delta \bar{e}^j$$

where $\ln p_{jk}$ is the average hedonic price of neighborhood k in school zone j , $\hat{\theta}$ is our estimated coefficient of the effect of school quality on housing prices ($\hat{\theta} = 0.02$)²⁴ and $\Delta \bar{e}^j$ is the difference in standardized average test scores between school j and the worst Parisian school.

We consider Parisian neighborhoods to correspond to the Census blocks. Under the name *Iris*, they are defined by the French Insee²⁵ as a small geographical area of approximately 2,000 inhabitants sharing similar socio-economic characteristics. After dropping the *Iris* with too few housing sales²⁶, we can divide Paris in 718 neighborhoods.

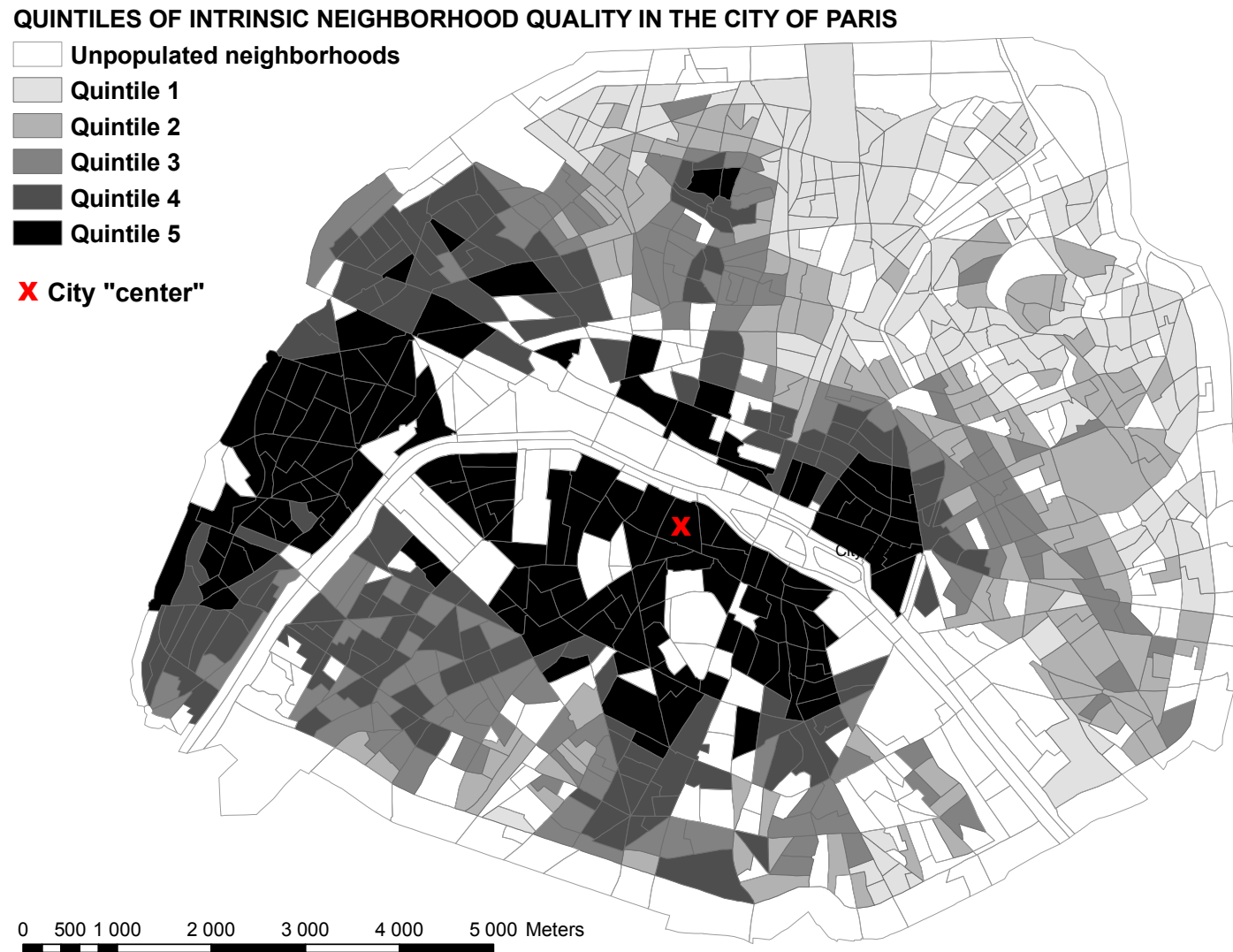
Figure 2.13 shows that the spatial distribution of neighborhood quality (one point represents one *Iris*) is clearly not random, but the geographical center of Paris does not precisely coincide with the highest quality neighborhood. As our simulations

²⁴This coefficient means that one standard deviation increase in test scores raises housing prices by 2%.

²⁵*Institut National de la Statistique et des Études Économiques*.

²⁶We dropped areas in which less than 100 sales occurred between 1997 and 2003.

Figure 2.13: Simulation example: spatial distribution of the estimated quintiles of intrinsic neighborhood quality in the city of Paris.



are based on a checkerboard with the best quality neighborhoods located near the center of the rectangle, we have to redefine the city center as the best neighborhood quality location in order to calculate the correlation between neighborhood quality and proximity to the center. More precisely, we select the area with the highest neighborhood quality and define it as the city center (this places the city “center” in Saint-Germain-des-Prés). We then calculate the ranking of neighborhood quality and the ranking of the proximity of each neighborhood to this center. The rank correlation is equal to 0.51. If we decide alternatively to position the city center in the middle of the area formed by neighborhoods belonging to the top 1% of neighborhood quality, we find a correlation equal to 0.61. We therefore take values of λ between 0.5 and 0.6

The correlation between family income and child ability (ρ). The correlation coefficient ρ cannot be directly estimated with our data on educational achievement because we cannot link parental income and child ability, as all we know is the occupation (socio-economic status) of parents. Using the average log income by occupation would probably lead to a severe underestimation of the correlation because it would ignore the correlation between parental income and ability at the *individual* level.

We therefore rely on existing empirical estimates of the intergenerational transmission of income. Estimates of the correlation between fathers’ and sons’ long run income range from 0.4 (Solon, 1992; Lefranc and Trannoy, 2004) to 0.6 (Mazumder, 2003; Dearden et al., 1997). We take $\rho=0.5$ as our benchmark parameter and test the sensibility of our estimate to lower values of ρ ²⁷.

The taste for school quality (λ). Finally, once we have determined ρ and λ , we can calibrate the relative taste for school quality β , *i.e.* the weight that families put on school quality relatively to neighborhood quality in the total desirability of a block. We take the value of β that matches the level of residential and school segregation observed in Paris, given the values of ρ and λ .

²⁷Lefranc and Trannoy (2004) estimate a correlation coefficient of 0.4 in France, but the relation could possibly be higher in Paris. As a lower ρ yields lower levels of segregation in public schools, we would have to choose a very high value of β to match the ability segregation in Parisian public schools.

In practice, the admissible range of parameters β and ρ varies with the size of peer effects. If peer effects are very strong, even a moderate level of β might generate high variations in educational outcomes at the school level. As we have already mentioned, empirical estimates of peer effects have not yet reached a consensus. We stick to the linear-in-means model, but we allow the size of peer effects to vary. Hoxby (2001) finds that a 1 point change in the average test score raises a student's own score between 0.15 and 0.4 point. We take these two values as the lower and upper bounds for the level of peer effects.

The seceding threshold s_0 to create a private school. We calibrate the value of this parameter so that the proportion of middle school children attending private schools matches the Parisian situation, where private schools account for 30% of total enrollment.

The data are best matched when setting the parameters values as follows: $\lambda = 0.6$; $\rho = 0.5$; $\beta = 0.5$; $s_0 = 0.3$. With these values, residential segregation is found equal to 0.38, segregation by ability in public schools to 0.12 and private school enrollment to 30%. Given the level of ability segregation, taking the lower bound for peer effects (0.15 point) yields an index of test score segregation in schools of 0.15 while taking the upper bound for peer effects yields an index of 0.21.

2.6.2 School enrollment policy simulations

The results of school enrollment policy simulations using the parameter values calibrated to match Parisian features are shown in tables 2.3 and 2.4.

Open enrollment. Given that pupils from high-income families tend to be enrolled in private schools, the fact that the private sector disappears after the introduction of open enrollment yields educational gains and losses within deciles (gains for high ability children and losses for low-ability children) such that the net average effect is small for all deciles (see figure 2.14). High-ability children still unambiguously gain from the policy, whereas low-ability children are hurt by the policy. However, high-ability children from low-income background experience high gains in peers quality

Table 2.3: Calibration for Paris: residential and educational segregation outcomes from alternative school enrollment policies.

PARAMETER VALUES: $\lambda=0.6$. $\rho=0.5$. $\beta=0.5$. $s_0=0.4$

	Residential segregation (log income)	Educational segregation within public schools: log abilities	Educational segregation within private schools	Average log ability in public schools	Average log ability in private schools	Private Enrollment (%)
Benchmark equilibrium: strict school zoning (s.e.)	0.38 (0.04)	0.12 (0.01)	0.13 (0.02)	-0.23 (0.03)	0.48 (0.04)	0.30 (0.04)
School zone reassignments: initial level (s.e.)	0.14 (0.04)	0.07 (0.02)				
School zone reassignments: final level (s.e.)	0.15 (0.03)	0.11 (0.00)	0.14	-0.29	0.42	0.39
Open enrollment (s.e.)	0.31 (0.03)	0.98 (0.00)				0.00

Notes: λ : correlation between neighborhood quality and proximity to the city center; ρ : correlation between family income and child ability; β : parental taste for school quality; s_0 : seceding threshold to create a private school. Results from 100 simulations performed on random initial spatial distribution of neighborhood quality, given λ . Standard errors of the estimates are in parentheses.

Table 2.4: Calibration for Paris: distribution of educational gains induced by school zone reassignments and open enrollment.

PARAMETER VALUES: $\lambda=0.6$. $\rho=0.5$. $\beta=0.5$. $s_0=0.4$

PERCENTAGE OF WINNERS:

	Decile	1 th dec.	2 nd dec.	3 rd dec.	4 th dec.	5 th dec.	6 th dec.	7 th dec.	8 th dec.	9 th dec.	10 th dec.	Total
School zone reassignments	by income	0.34	0.50	0.51	0.49	0.45	0.45	0.50	0.51	0.14	0.04	0.39
	by ability	0.42	0.44	0.43	0.44	0.43	0.40	0.38	0.37	0.34	0.28	0.39
Open enrollment	by income	0.42	0.51	0.49	0.50	0.50	0.51	0.51	0.55	0.49	0.48	0.50
	by ability	0.00	0.01	0.12	0.26	0.40	0.61	0.74	0.86	0.96	0.99	0.50

AVERAGE GAINS:

	Decile	1 th dec.	2 nd dec.	3 rd dec.	4 th dec.	5 th dec.	6 th dec.	7 th dec.	8 th dec.	9 th dec.	10 th dec.	Total
School zone reassignments	by income	-0.07	0.01	0.01	0.00	-0.02	-0.02	0.01	0.04	0.02	0.01	0.00
	by ability	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00
Open enrollment	by income	-0.14	0.05	-0.05	0.03	0.01	0.05	0.03	0.11	-0.03	-0.06	0.00
	by ability	-1.33	-0.83	-0.51	-0.28	-0.10	0.12	0.28	0.53	0.80	1.32	0.00

Notes: λ : correlation between neighborhood quality and proximity to the city center; ρ : correlation between family income and child ability; β : parental taste for school quality; s_0 : seceding threshold to create a private school. Results from 100 simulations performed on random initial spatial distribution of neighborhood quality, given λ . Standard errors of the estimates are in parentheses.

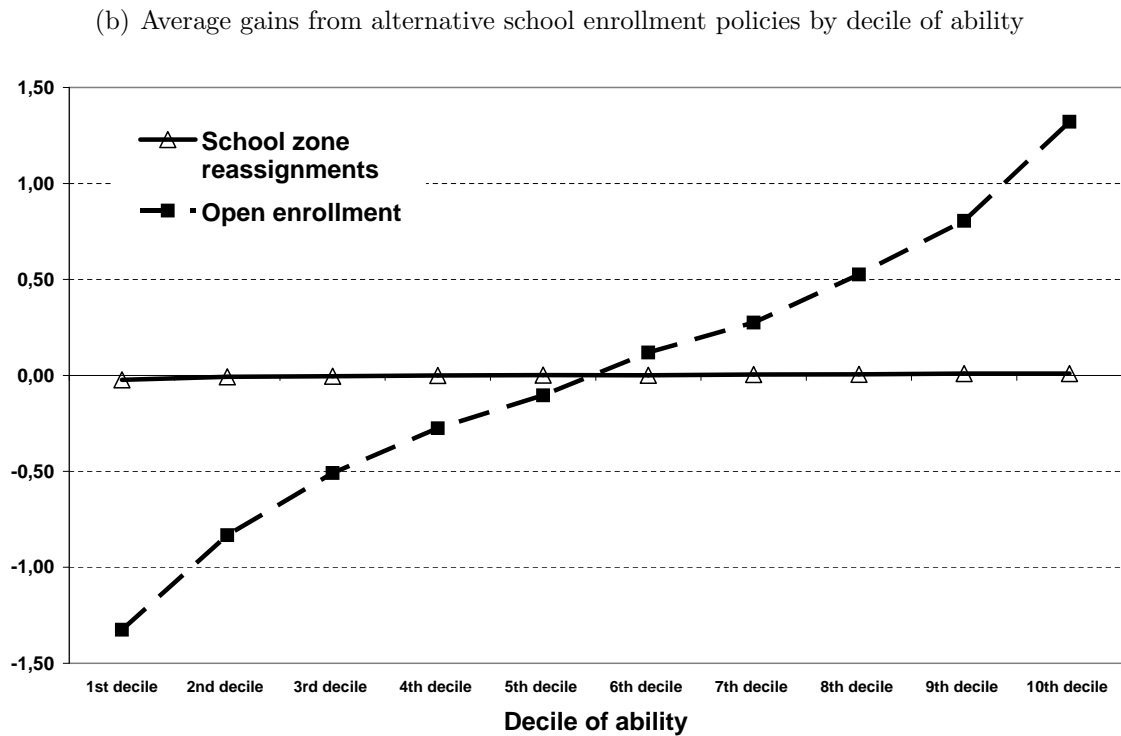
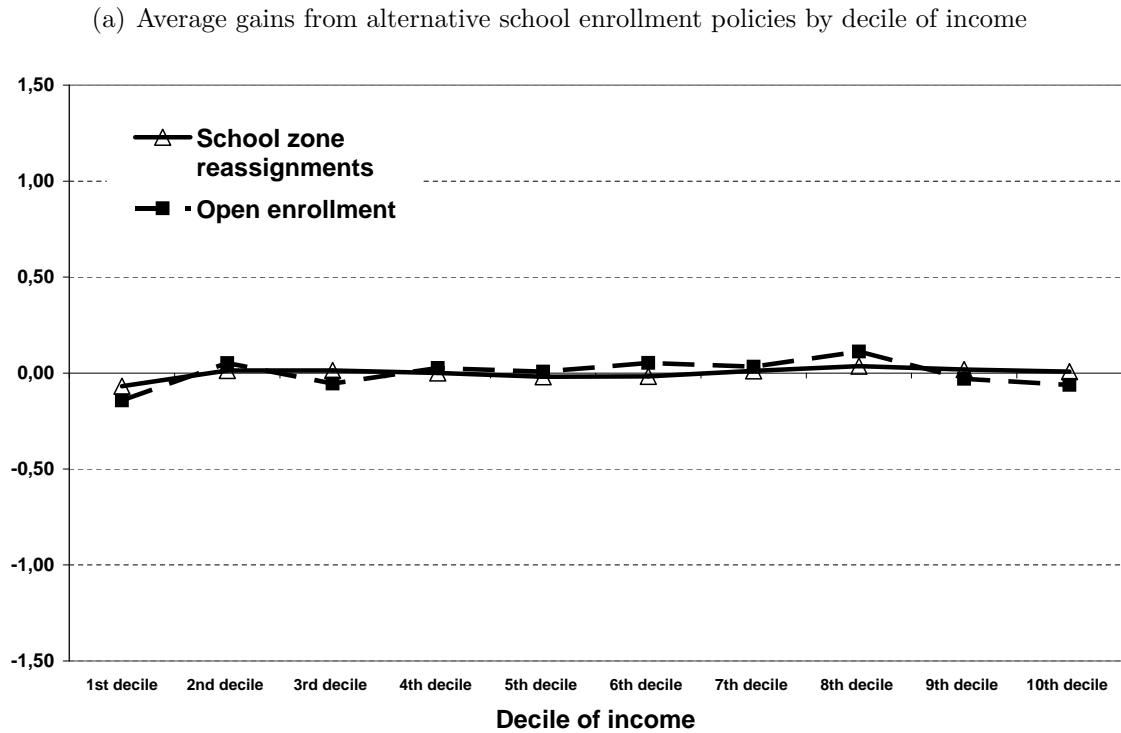
as children from high-income families previously attending private schools are now enrolled in public schools. Even if the proportion of high-ability children is lower in the bottom of the distribution, the high educational gains for these children tend to compensate for the losses incurred of their low-ability counterparts.

At the other end of the distribution, the gains for high-income families are small (and might even be negative), since they were already benefitting from high-ability peers in private schools. In particular, children from high-income families with low abilities who were previously enrolled in private schools must now attend the worst public schools. High-ability children from high-income families do not experience such high gains in peer quality, since they benefited from relatively high private school quality before the reform.

School zone reassignments. In our simulations, school zone reassignments initially reduce school segregation by almost 45%, but parental reaction to the policy wipes out this initial reduction in equilibrium: in the presence of private schools, redrawing school zones boundaries is useless because it does not yield an equalizations of public school quality in equilibrium. This comes from the increase in private school attendance from 30% to 39% that further cream skims public schools. Interestingly, residential segregation does not increase after the reassignment, since dissatisfied parents now have the possibility to create a private school and do not need to change residence to attend a better public school. Figure 2.14 shows that average gains are close to zero for all deciles of ability and income, with children in the first decile experiencing even lower peer quality than before the reassignments.

These results suggest that in the presence of private schools, redrawing school boundaries would not be a very effective policy to reduce stratification across public schools. On the contrary, open enrollment does not appear to harm low-income children more than their well-to-do counterparts. Indeed, open enrollment yields very high gains for high-ability children coming from deprived backgrounds, which balance the losses experienced by their low-ability counterparts.

Figure 2.14: *Calibrations for Paris. distribution of educational gains induced by school zone reassignments and open enrollment, dy deciles of income and ability. Parameter values: $\lambda=0.6$. $\rho=0.5$. $\beta=0.5$. $s_0=0.4$.*



2.7 Conclusion

In this theoretical study, we have addressed the issue of how school admission rules affect housing markets, residential segregation and student performance. Our main objective was to assess the analytical properties of alternative school enrollment schemes within the framework of a multi-community model that matches some of the salient institutional features of the French education system, which is centrally financed and comprises a large and subsidized private sector.

Our simulations calibrated on Parisian data show that the impact of shifting school attendance boundaries or switching from strict school zoning to open enrollment is highly dependent on the presence of private schools. Indeed, both the segregation outcomes and the welfare implications of these policies change dramatically when parents can choose to opt out from public school by enrolling their children in private establishments.

In the absence of private schools, redrawing school boundaries in a way that equalizes the socio-economic composition of schools reduces residential segregation by income, lowers educational segregation by ability and benefits to children from low-income families. However, the final impact of this policy on the level of educational inequalities is lower than initially targeted, as the behavioral response of parents who decide to relocate in a different part of the city tends to endogenously recreate part of these inequalities. When private schools are introduced in the model, shifting school attendance boundaries no longer has a sizeable effect on educational segregation and wipes out the benefits enjoyed by children from low-income families. This is because in addition to changing residence, high-income parents now have the option of avoiding the newly desegregated public schools by enrolling their children in the private sector. In our model, the expansion of the private sector (in the order of 30%) will actually tend to harm the children enrolled in the lowest performing public schools, because their schools' performance falls below the lowest levels observed before the reassignments.

The consequences of switching from strict school zoning to open enrollment are also very different if private schools are present or not. In both cases, allowing parents to freely choose their school reduces residential segregation by income and increases

educational stratification by ability. In our model, however, children from low-income families are hurt by this policy only if private schools did not operate in the initial strict school zoning scheme. In the absence of private schools, allowing parents to choose their school unambiguously lowers the average ability of low-income children's schoolmates, because of the strong positive correlation between ability and parental income. But if private schools were operating in the strict school zoning scheme, then switching to open enrollment does not have an adverse effect on pupils from low-income families. This is because private schools tend to disappear when parents are allowed to choose their public school, so children from high-income families tend to join public schools and improve their average performance. The education losses incurred by low-ability pupils from low socio-economic backgrounds are then offset by the large educational gains enjoyed by high-ability pupils from low-income families.

It must be noted that these results are partly driven by the way we model private schools. The assumption that private schools would disappear if an open enrollment policy was implemented is somewhat extreme. First, the choice of sending a child to a private school instead of the local public school is not uniquely determined by the public-private school quality differential and other considerations such as religious orientation may directly influence private school attendance. Second, the fact that the support for private schools collapses after the introduction of school choice is essentially driven by the increase in the quality of top public schools. Yet if a number of admission criteria were established to regulate open enrollment (such as parental income or distance to school) and limit stratification by ability, the attractiveness of good public schools would be reduced and private schools could still be operating in the new equilibrium. Hence there are reasons to believe that the welfare implications of a regulated system of open enrollment would be fairly different from those of the policy we have modelled. Nonetheless, it is reasonable to expect school choice policies to reduce private school enrollment and to increase the average performance of public schools as a result.

Similarly, the ineffectiveness of school zone reassignments is essentially driven by private school enrollment. Given the constraints imposed by the French system, it seems very unlikely that the size of the subsidized private sector would instanta-

neously increase by 30%. Moreover, as private school attendance is costless in our model, it triggers a high demand for private schools. Despite these limitations, we believe that one of the strengths of our model is that it captures the fact that reassignment policies would probably increase the dissatisfaction of high-income parents and induce them to avoid the local public schools. It also reproduces a specific feature of the French context, namely that private schools tend to be concentrated in wealthy neighborhoods. This contrasts with the calibrations usually carried out in the US context, where families who choose to send their children to private schools move to lower quality neighborhoods, in order to avoid paying the high local tax rates that characterize high-income neighborhoods.

The main lesson from this theoretical exercise is that in the French educational context, the interaction between private and public schools plays a crucial role in shaping the effects of alternative school enrollment policies. This highlights the need for a study of the determinants of parental choice between public and private education, since the reaction of families switching from one type of schooling to the other is likely to dramatically affect the reforms' outcomes.

Chapitre 3

Sectorisation et marché immobilier : analyse empirique de l'impact de la performance des collèges parisiens sur le prix des logements

3.1 Introduction

Motivation. In France, as in many other countries, there is an ongoing debate on how residence-based assignment to schools affects both educational and residential segregation. Theoretical models (including Bénabou, 1993; Fernández and Rogerson, 1996) have shown that the existence of peer effects in education yields income and residential sorting in equilibrium. One of the key parameters of these models is the willingness of parents to pay for school performance. The higher the value of this parameter, the higher the level of sorting when admission to schools is tied to residence. However, these effects might be mitigated in the presence of private schools, which operate under different rules of admission. In this study, we test the theoretical predictions of models of housing markets in which public and private schools coexist (Nechyba, 1999, 2000, 2003). We estimate the impact of school performance on

housing prices and investigate how this effect varies with the availability of private schools in the neighborhood.

Related literature. Several empirical papers have sought to test the empirical prediction that housing prices should be higher in areas where schools perform better. Earlier papers have estimated hedonic regressions in which housing prices are regressed on school performance, controlling for housing and neighborhood characteristics¹. Unfortunately, these early estimates are plagued by endogeneity problems, since better schools tend to be located in wealthier neighborhoods and pupils drawn from privileged socio-economic backgrounds have higher academic achievement. If the estimation strategy fails to correct for observable and unobservable neighborhood characteristics, potentially correlated both with housing prices and school performance, then the estimated marginal willingness to pay for a better school will suffer from severe biases. In her 1999 paper, Black proposed a method to solve this problem by comparing housing prices across primary school attendance district boundaries. The identifying assumption is that changes in school performance are discrete at boundaries, while changes in neighborhood characteristics are smooth. The difference in mean housing prices located on opposite sides of attendance district boundaries can therefore be related to differences in school test scores only. When restricting the sample to the set of sales located within 0.15 mile of a boundary, Black finds that a 5% increase in primary schools' test scores (approximately one standard deviation) is associated with a 2.1% increase in housing prices, which is half the value of the "naive" OLS estimate. On UK data and using an alternative estimation strategy to correct for spatial fixed effects², Gibbons and Machin (2003, 2006) find an effect of the same order of magnitude for primary schools. A study by Bayer et al. (2003) yields similar results using a discrete choice model instead of a hedonic model. Furthermore, these authors provide evidence of significant heterogeneity in

¹In the French context, Gravel and Trannoy (2006) have conducted the first study using data from a Parisian suburb. For a review of other earlier papers, see Black (1999).

²In the UK, the estimation strategy cannot rely on school boundary fixed effects, since school catchment areas are not strictly defined. Proximity to primary schools is only one criterion among others for admission in schools and Cheshire and Sheppard (2004) point out that living in a house near a boundary might render admission to a school more uncertain than living in a house located in the immediate vicinity of a school.

the marginal willingness of households to pay for school performance. A few papers have exploited changes in school boundaries over time. Bogart and Cromwell (2000) measure the effect of reshaping school catchment areas and introducing school bus-ing and find a negative impact of these policy changes on housing prices. Reback (2005) exploits the adoption of open enrolment for public schools in Minnesota and finds that housing prices tend to increase in school districts where pupils are able to transfer to their preferred schools, while they decline in districts that receive these transfer pupils.

Some papers have looked more precisely at which components of school quality are most valued by parents³. The results of these studies are mixed, but the most recent papers, including Downes and Zabel (2002) and Brasington and Haurin (2006), suggest that parents react to school performance measures based on average test scores but not to value-added measures of school quality. In addition, Clapp et al. (2008) find that school demographic composition, such as the percentage of Hispanic students, also affects housing prices.

A few papers have investigated whether the release of new information about school quality yields changes in housing prices. Figlio and Lucas (2004) study the introduction of a new school accountability system in Florida and find that housing markets first respond strongly to the new letter grading of schools but that the effect dissipates after a few months. Indeed, the new grading appears to provide relatively noisy information on school quality, with a lot of fluctuation of school letter grades over time. Using the introduction of a similar school accountability system based on test scores and value-added measures in the Charlotte-Mecklenburg county in North Carolina, Kane et al. (2003) do not find any significant impact of the release of new measures of school performance on housing prices, but a significant impact of the long-run test scores averages. Finally, Kane et al. (2006) argue that the effect of school quality on housing prices does not only reflect differences in school performance, but also differences in neighborhood characteristics arising from households sorting along school attendance boundaries. Using data on the Charlotte-Mecklenburg county, they show that neighborhood characteristics change discontinuously at boundaries when

³For a comprehensive review of these papers, see Brasington and Haurin (2006).

school quality differentials are sufficiently large. These differences across boundaries might be particularly important in this county, which was operating under a court-imposed desegregation order that drew school boundaries so that students living in the same neighborhood were sent to different schools. However, these results suggest that it is important to compare socio-demographic characteristics across boundaries when studying the impact of school performance on housing prices.

Almost all of these studies focus on public schools and although many acknowledge that the presence of private schools might affect the premium paid for public schools, none of them has analyzed this question in depth. To our knowledge, this contribution is the first to investigate the effect of private schools on the housing price premium attached to public school performance.

The study's focus. In respect to this existing literature, our contribution innovates in two main directions. First, we improve the estimation strategy originally proposed by Black (1999) to deal with the endogeneity of school performance in the housing price equation. We do so by developing a matching framework which consists in the careful comparison of sales across public middle school attendance boundaries. We also investigate the validity of this identification strategy by checking precisely if other neighborhood characteristics vary discontinuously across boundaries.

Second, we investigate whether school choice reduces the capitalization of public school performance in housing prices. We use comprehensive data on both schools and real estate transactions in the city of Paris over the period 1997-2004 to study how the availability of private schools locally affects the willingness to pay for better public schools. The organization of secondary schooling in the city of Paris, which combines residence-based assignment to public schools with a well-developed and almost entirely publicly funded private school system, offers a valuable empirical context for analyzing this issue.

Using different measures of school performance, we find that a standard deviation increase in school performance raises housing prices by 1.5 to 2.5%. The size of this effect is similar to existing estimates in the US and UK contexts and can explain roughly 5% of the observed difference in housing prices between adjacent school zones.

We also find that the price premium attached to good public schools exhibits spatial heterogeneity and varies with the availability of private schools in the neighborhood. In line with the theoretical predictions of general equilibrium models of school choice, the presence of good private schools in certain neighborhoods tends to attenuate the capitalization of public school performance in housing prices, by providing an advantageous outside option to parents.

The remainder of this chapter is as follows: section 3.2 describes the estimation strategy; section 3.3 briefly presents the French educational system and school admission rules; section 3.4 gives a description of the data; section 3.5 presents the basic results; section 3.6 performs some robustness checks and section 3.7 evaluates the impact of private schools on the capitalization of public school performance in housing prices.

3.2 Estimation strategy

In this section, we present our empirical framework to assess the impact of school performance on housing prices. We start by briefly discussing the bias induced by the endogeneity of school performance in the classic hedonic model of housing prices, before explaining how the endogeneity problem can be overcome by the inclusion of a full set of school attendance boundaries fixed effects in the housing price equation. After discussing the limitations of the fixed effect approach, we turn to our identification strategy, which consists in a careful matching of sales across common middle school attendance boundaries.

3.2.1 The model

Within a city, the price per square meter (taken in natural logarithm) of a transaction i located in neighborhood n , belonging to school zone s during school year t , is assumed to depend linearly on a school performance z_s index, a vector of observable features of the flat (floor, age of the building, number of rooms, number of bathrooms,

etc.), a time-varying neighborhood effect $\theta_{n,t}$ and an error term $\epsilon_{i,n,s,t}$:

$$\ln p_{i,n,s,t} = \alpha + \beta \cdot z_s + X_{i,n,s,t} \gamma + \theta_{n,t} + \epsilon_{i,n,s,t} \quad (3.1)$$

In this relatively flexible housing equation, the school performance index z_s is assumed to capture the medium-run characteristics of the local public middle school that are valued by households when making their residential choice. The time-varying neighborhood effect $\theta_{n,t}$ captures local factors, such as local amenities or the neighborhood's socio-demographic composition, that affect prices on the housing market. They are supposed to vary continuously through space. Finally, the error term $\epsilon_{i,n,s,t}$ includes all the unobservable characteristics that influence the price of a particular sale i . The coefficient β , which measures the impact of school performance on housing prices, is our parameter of interest.

In our model, neighborhood time-varying effects $\theta_{n,t}$ are not necessarily independent of school performance z_s . In particular, $\theta_{n,t}$ incorporates the socio-economic composition of a particular neighborhood, which can influence both school performance through peer quality and housing prices by making an area more or less enjoyable to live in. However, the crucial assumption that will be maintained throughout the analysis is that within a common neighborhood n observed in year t and given a flat's observable features $X_{i,n,s,t}$, the unobservable component of housing prices $\epsilon_{i,n,s,t}$ is mean-independent of school quality:

$$E(\epsilon_{i,n,s,t} | z_s, X_{i,n,s,t}, \theta_{n,t}) = 0$$

Importantly, the unobservable component of housing prices $\epsilon_{i,n,s,t}$ is not assumed to be i.i.d.. Unobservable shocks that determine housing prices in addition to flat, neighborhood and school characteristics are indeed most likely to exhibit serial and spatial correlation.

3.2.2 The endogeneity of school performance

The standard hedonic housing price function describes the price of a particular sale as a function of the flat's observable features, which include its intrinsic characteristics

(size, floor, etc.) as well as its neighborhood characteristics. The corresponding coefficients are supposed to measure the marginal purchaser's willingness to pay for each specific characteristic. The typical hedonic function for housing prices takes the following form:

$$\ln p_{i,c,s,t} = a + b \cdot z_s + X'_{i,c,s,t}c + N'_{i,c,s}d + L'_t e + u_{i,c,s,t} \quad (3.2)$$

where $p_{i,c,s,t}$ is the price of sale i , located in census district c , in school attendance zone s during school year t ; z_s is the performance index of school s , $X_{i,c,s,t}$ the vector of the flat's features, $N_{i,c,s}$ the vector of neighborhood socio-demographic characteristics (at the census district level), L_t a vector of time dummies and $u_{i,c,s,t}$ is the error term. The OLS estimate of parameter b is supposed to measure β , the marginal willingness to pay for a better performing school.

As pointed out by Black (1999), this methodology will produce upwardly biased estimates if the error term $u_{i,c,s,t}$ includes unobservable neighborhood characteristics that are correlated with school performance z_s and have an independent effect on housing prices. If the true underlying model of housing prices is described by equation (3.1), the hedonic model will suffer from omitted variable bias as long as the variables included in the right hand side of the regression equation fail to fully account for the time-varying neighborhood fixed effect $\theta_{n,t}$.

3.2.3 School attendance boundaries fixed effects and their limitations

Black (1999)'s estimation strategy consists in focusing exclusively on the set of sales that take place in the vicinity of a school attendance boundary. Under the assumption that unobservable determinants of housing prices vary continuously through space, the causal impact of school performance on housing prices can be estimated by comparing sales across common school attendance boundaries, which create spatial discontinuities in school performance.

The housing price equation (3.1) for sales located in the vicinity of a school at-

tendance boundary can then be written in the following way:

$$\ln p_{i,b,s,t} = \alpha + \beta.z_s + X_{i,b,s,t}\gamma + \theta_{b,t} + \epsilon_{i,b,s,t} \quad (3.3)$$

where the subscript b indicates that the sale is located in the vicinity of school boundary b .

The identifying assumption is that all sales located in the vicinity of a common attendance boundary are in the same neighborhood. Hence their prices are influenced by the same boundary fixed effect θ_b so that the remaining heterogeneity is uncorrelated with school quality:

$$A1': E(\epsilon_{i,b,s,t} | z_s, X_{i,b,s,t}, \theta_b) = 0$$

To estimate parameter β , Black replaces the vector of observed characteristics in the traditional hedonic equation (3.2) by a full set of boundary dummies:

$$\ln p_{i,s,b,t} = a + b.z_s + X'_{i,s,b,t}c + K'_b f + L'_t e + u_{i,s,b,t} \quad (3.4)$$

where K_b is the vector of boundary dummies, L_t a set of time dummies (year and quarter) and $u_{i,s,b,t}$ is an error term which is clustered at the school attendance level s .

Note that this school boundary fixed effects model embodies three relatively strong restrictions:

1. First, features of houses are supposed to have the same impact on prices in the whole the area under study. Yet there is some evidence⁴ that the influence of features such as the age of a building or the number of rooms in a flat is likely to differ noticeably across the various areas of a city like Paris⁵.
2. Second, since the full set of school boundary dummies is not interacted with time dummies, school attendance boundary fixed effects are assumed to be constant across years ($\theta_{b,t} = \theta_b \forall t$) and temporal shifts in housing prices to be common to all neighborhoods (through the time dummies L_t). This assumption

⁴See Laferrère (2005) in the case of Paris.

⁵For instance, the price premium attached to a parking space is likely to be higher in the center of the city (where parking lots are scarce) than in the periphery (where they are easier to find).

may appear too strong, as trends in housing prices are rarely identical in the different parts of a city.

3. Third, in Black's estimations, standard errors are clustered at the school attendance level to account for the fact that unobservable shocks affecting housing prices may exhibit spatial and serial correlation within a given school catchment area. Hence it is implicitly assumed that unobservable shocks are uncorrelated across common school attendance boundaries. This seems rather implausible since sales located on both sides of a common boundary typically belong to the same neighborhood. This implies that there must be some degree of spatial and serial correlation between the idiosyncratic shocks affecting the price of sales located in the vicinity of a common boundary.

The violation of either of these three assumptions would result in biased standard errors and would therefore lead to invalid inference. A more general estimation strategy could consist in including a full set of interaction between boundary time boundaries, as well as a full set of interactions between flat features and area dummies. But in addition to being particularly tedious, performing such a regression would induce a important loss in the model's degrees of freedom.

A final undesirable feature of the school boundary fixed effect strategy is that the comparison of sales located on both sides of a common attendance boundary does not take into account the distance between these sales. In particular, this methodology assumes that whatever the length of a particular border, flats located on both sides but at opposite ends of a particular boundary share the same unobservable characteristics. This may not be true in the case of very long boundaries passing through relatively different neighborhoods. In that case, a more efficient estimation strategy would consist in restricting a sale's comparison group to the set of transaction that are located within a certain radius from that sale.

3.2.4 An alternative estimator

In order to circumvent the limitations of the boundary fixed effect model, we adopt a matching framework that restricts a transaction's comparison group to a

carefully constructed counterfactual sale located on the other side of the common school attendance boundary and within a given radius of that specific transaction. The principles governing our estimation strategy are explained below and figure 3.1 shows how it applies for a particular set of sales.

First, to account for the fact that the price premium attached to a specific flat feature may vary throughout the area under study, we correct the prices of all sales for area-specific flat features effects, these areas being known to be relatively homogenous and to follow similar trends in prices⁶. The prices of all sales are homogenized in the sense that there are expressed in terms of a “typical flat equivalent price”⁷. We call “hedonic prices” these corrected housing prices⁸. Appendix A explains in detail how they are computed.

We suppose that the price per square meter (taken in log) of a transaction i , assigned to public middle school s , located in the vicinity of a school attendance boundary b and completed during school year t is determined by the housing price equation (3.3). We call “reference” sales all housing transactions that are located within a certain distance d of a school attendance boundary. For each reference sale i , we construct a fictive “counterfactual” sale i' which is located on the other side of the common school attendance boundary (s, s'). The price of this counterfactual sale is supposed to measure the amount for which the reference transaction would have been sold, had it been located in school attendance area s' rather than in school attendance area s , everything else being equal.

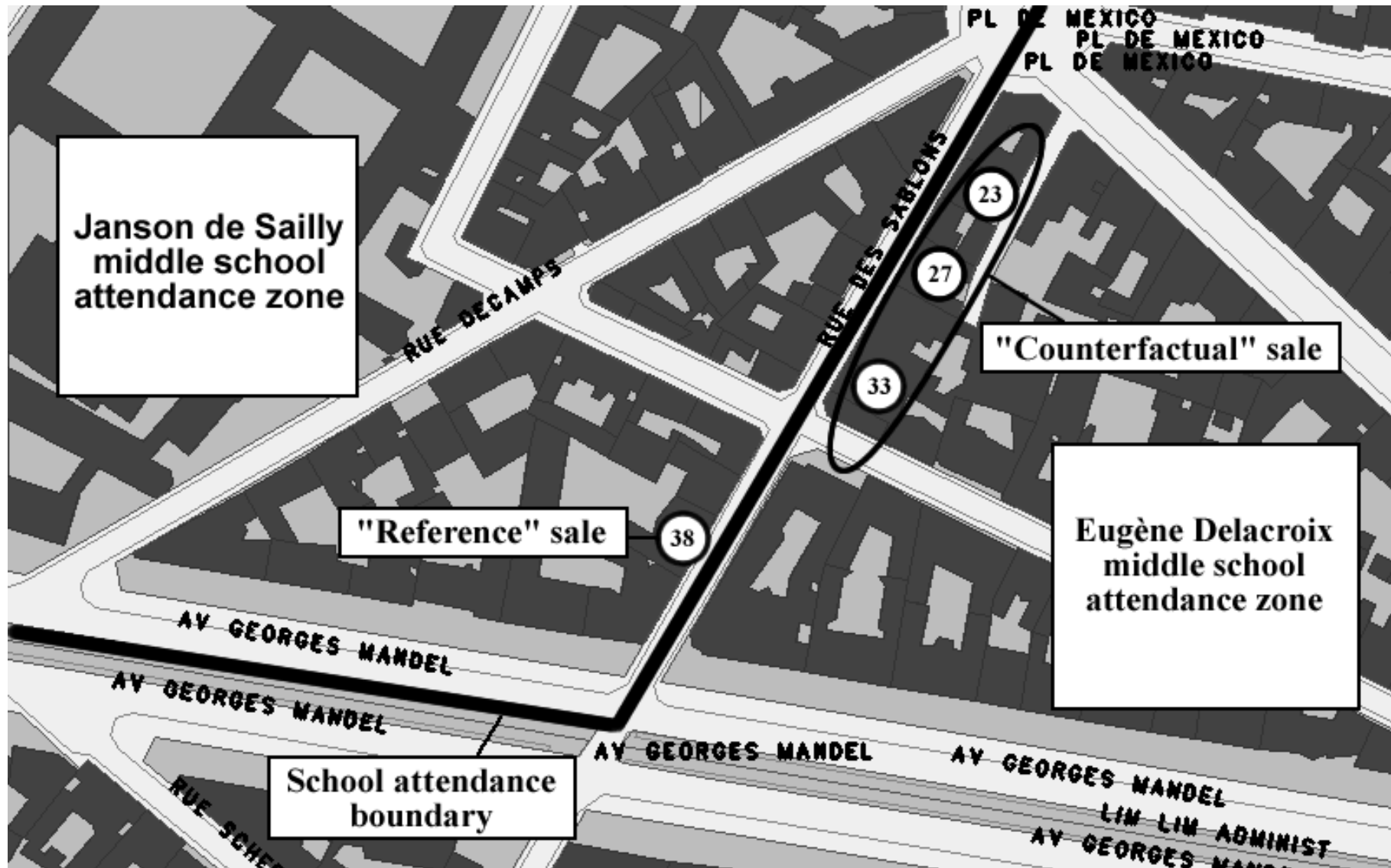
The price of the counterfactual transaction i' cannot be observed and has to be estimated. We do so by calculating the weighted geometric mean of the prices of all transactions j that took place in the vicinity and during the same school year t as the reference sale i , but that were assigned to school s' rather than to school s . More precisely, we define the set $H_{i',s',b,t}$ of sales used to construct the counterfactual sale i'

⁶In the city of Paris, we have defined 18 such areas.

⁷In the case of Paris, the typical flat has two rooms, belongs to a building constructed between 1850 and 1913, is located on the ground floor, has one bathroom and average size rooms, has no maid’s room nor parking space and was sold during the fourth quarter of a particular year.

⁸While this correction is not essential to our estimation strategy (since flats located on both sides of common attendance boundaries should share similar features on average), it will increase the precision of our estimates by reducing the noise associated with the heterogeneity of flats’ observable characteristics.

Figure 3.1: Matching sales across boundaries: an example.



as:

$$H_{i',s',b,t} = \left\{ j : d_{i,j} \leq d, s' \neq s \right\}$$

This set includes all the sales located within a radius d of the reference sale i , completed in year t and assigned to school s' .

The transactions used to construct the counterfactual sale are weighted by the inverse of their distance $d_{i,j}$ to the reference sale i , in order to give more importance to the sales that are located nearby relatively to the more distant ones. Hence, the counterfactual sale's estimated price per square meter $p_{i',b,s',t}$ is calculated as:

$$\ln p_{i',b,s',t} = \sum_{j=1}^J w_{i,j} \ln p_{j,b,s',t}$$

where $J = \text{Card}(H_{i',s',b,t})$ and the weights $w_{i,j}$ are defined as:

$$w_{i,j} = \frac{\frac{1}{d_{i,j}}}{\sum_{j=1}^J \frac{1}{d_{i,j}}}$$

Note that a given reference sale could potentially be matched with different counterfactual sales located along distinct school attendance boundaries. This would be the case in particular for sales in the “corners” of school attendance zones. To avoid using the same references sales in different matches (a feature that would induce serial correlation in housing prices across different boundaries), we decided to assign each sale to its closest boundary. Hence as long as they belong to distinct boundaries, our matches will be constructed using different sales.

By definition, the price of each counterfactual sale i' can be written as:

$$\ln p_{i',b,s',t} = \alpha + \beta \cdot z_{s'} + \theta_{b,t} + \epsilon_{i',b,s',t}$$

where the disturbance term $\epsilon_{i',b,s',t}$ is a weighted average of the disturbance terms of the sales used to construct the counterfactual:

$$\ln \epsilon_{i',b,s',t} = \sum_{j=1}^J w_{i,j} \ln \epsilon_{j,b,s',t}$$

The identification of the “public middle school effect” relies on the crucial assumption that the counterfactual sale i' and the reference sale i share the same unobservable time-varying neighborhood effect $\theta_{b,t}$. Under this assumption, the housing price differential between the reference sale and the constructed counterfactual can be written:

$$\ln p_{i,b,s,t} - \ln p_{i',b,s',t} = \beta(z_s - z_{s'}) + \epsilon_{i',b,s',t} - \epsilon_{i,b,s,t} \quad (3.5)$$

Parameter β can be estimated by running an OLS regression of the price differential between the reference and the counterfactual sale on the corresponding school performance differential. Since the identifying assumption is more likely to hold for matches that are geographically close⁹, we perform weighted OLS regressions, the weight given to a particular match being equal to the inverse of the distance between its components.

An important feature of the housing price differential equation (3.5) is that by construction, the composite error term $(\epsilon_{i',b,s',t} - \epsilon_{i,b,s,t})$ is not i.i.d.. In this equation, serial correlation arises for two reasons. First, as emphasized earlier, the unobservable housing price component of the reference sales $\epsilon_{i,b,s,t}$ may exhibit spatial and serial correlation. Second, the counterfactual error term $\epsilon_{i',b,s',t}$ is itself a weighted average of the unobservable housing price components of the different sales that are used to construct it. The error terms $\epsilon_{i',b,s',t}$ will therefore exhibit spatial correlation along a given boundary because the same sales will be used in the construction of several counterfactuals. Note however that spatial correlation in the error term will only affect the matches that belong to the *same* school attendance boundary. Hence, provided that we standard errors are clustered at the school attendance boundary level, the serial correlation in equation (3.5) will be entirely accounted for. We will do so in the rest of our analysis.

In the next section, we emphasize the features of the French educational system that are relevant to the present analysis as well as the rules that govern public and private school admission in the city of Paris.

⁹For each match, the distance between the reference and the counterfactual sale is computed using their respective geographic coordinates. The geographic coordinates of the reference sale are known. The counterfactual’s geographic coordinates are computed as the weighted average coordinates of the sales that were used to construct it, with weights equal to the inverse of their distance to the reference sale.

3.3 Public and private middle schooling in Paris

3.3.1 The French educational system

3.3.1.1 General organization

The French educational system is highly centralized and fairly homogenous until pupils reach the age of 14. Primary and secondary education is organized as a 12-year curriculum, divided into three stages: children spend 5 years in primary school (age 6 to 10), 4 years in middle school or *Collège* (age 11 to 14) and 3 years in high school or *Lycée* (age 15 to 17). The school year starts early in September and ends early in July¹⁰. Until the third year of *Collège*, all pupils follow exactly the same curriculum. A first round of orientation takes place at the beginning of the third year (*Quatrième*) with the lowest-achieving pupils being tracked into pre-vocational courses in the same middle school as the other pupils who follow the general curriculum. The main orientation round takes place at the end of middle school (*Troisième*): about two thirds of the pupils are admitted into the *Lycées d'Enseignement général* where they follow the general curriculum leading to the *Baccalauréat* final exam, whereas the remaining third is sent to the *Lycées Professionnels*, which provide vocational education. Note that *Collèges* and *Lycées* are normally separate institutions housed in different buildings.

3.3.1.2 Public schools

Education in France is predominantly public, public schools accounting for 86% of primary school enrollment and 79% of secondary school enrollment¹¹. Public education is centrally financed¹² and supervised at the local level by 35 Local Education Authorities (LEAs) called *Académies*. They are in charge of managing human and financial resources and of implementing the official educational programs produced by the Ministry of Education. Importantly for our study, the city of Paris is under

¹⁰For simplicity, we denote each school year using one single number: for instance, “school year 1997” stands for “school year 1997-1998”.

¹¹See de Monredon (2008) and Jaspar (2008).

¹²This aspect of the French educational system implies that contrary to the US system, public school finance does not depend on local taxation.

the supervision of one unique *Académie*, which decides how to allocate pupils and resources to schools.

Human and financial resources are for the most part allocated to schools on the basis of enrollment, so spending per pupil is very similar across different institutions¹³. An important aspect of the management of human resources within the different LEAs is that randomization play an important role in the allocation of teachers to schools. When applying for a position, qualified teachers rank the different *Académies* by order of preference but cannot choose between the schools of a given *Académie*. The allocation of teachers to schools then depends on the available positions and the number of points accumulated, which themselves are determined by a great variety of factors, including marital status, overall seniority or the number of years of teaching in the same school. The result of this allocation rule is that within a given *Académie*, the characteristics of teachers are fairly similar across schools.

3.3.1.3 Private schools

In France, private education accounts for 14% of primary education enrollment and 21% of secondary education enrollment. Respectively 13% of primary schools and 33% of middle and high schools belong to the private sector¹⁴. Private education is predominantly religious, 90% of private schools being Catholic. Several institutional features of private schools in France make them very different from their UK or US counterparts and should be kept in mind when interpreting our results.

A first specificity of French private schools is that most of them are publicly funded. Depending on their degree of independence with the State and the Ministry of Education, private schools are either *sous contrat* or *hors contrat*. *Sous contrat* schools are subject to State supervision. They follow the national curriculum and schedule and appoint qualified teachers who are paid by the State¹⁵. Part of *sous contrat* private schools' expenses are covered by local authorities. Schools are au-

¹³It has been shown that even the French *Zones d'Éducation prioritaire* program, which was supposed to channel additional resources to schools in disadvantaged areas, translated only into a 10% increase in spending per middle school pupil (Bénabou et al., 2005).

¹⁴The share of private secondary schools is larger than the share of private school enrollment because private schools are on average smaller than public schools.

¹⁵There is a specific competitive exam for recruiting private school teaching (called *CAFEP-CAPES*), which is very similar to the competitive exam for recruiting public school teachers.

thorized to set fees, but only to pay the costs that are not publicly funded such as optional courses, which often include religious education. As a result, private education is usually not free but the fees set by *sous contrat* schools are modest and usually in the range of 500 to 2,000 euros per annum, plus small additional costs. Many *sous contrat* schools also offer substantial rebates for high-performing pupils from low-income families. On the other hand, *hors contrat* private schools are not monitored nor financed by the State or local authorities and can freely design their own curriculum. Nevertheless, The vast majority of private schools, especially in primary and secondary education, are under contract with the State: out of the 68 private middle schools in Paris, only 3 are classified as *hors contrat*¹⁶.

A second specificity of private education in France is that the State exerts a direct control over the size of the private sector, through two main channels. First, the opening of new private schools is subject to very tight restrictions. New private institutions cannot be publicly financed without the approval of an *ad hoc* board (composed of members of the Ministry of Education, representatives of local communities and representatives of private schools) and must meet a “recognized need for education” (“*besoin scolaire reconnu*”)¹⁷. Second, the Ministry of Education can decide to keep the size of the private sector within certain limits by imposing a cap on the number of positions that are offered every year in the competitive examination for the recruitment of private school teachers. Indirect evidence seems to point out that the actual size of the private sector is kept smaller than needed to satisfy the demand for private education, especially in large cities¹⁸.

3.3.2 Public and private school admission rules

The rules governing school admission are probably the most important feature that distinguishes public middle schools from private middle schools in France.

¹⁶In our analysis, we will restrict our sample of private schools to *sous contrat* schools, information about pupils enrolled in *hors contrat* schools being unavailable.

¹⁷See art. L442-5 to L442-11 of the *Code de l'Éducation*.

¹⁸According to the President of the French National Association of Private Catholic Schools, over 25,000 applications to Catholic schools had to be rejected before the start of academic year 2006-2007 because of oversubscription (“*L'école privée affiche toujours complet*” [“Private schools are still over-subscribed”], *Le Figaro*, October 15, 2007).

3.3.2.1 Public schools

Since 1963, the allocation of pupils to public primary and secondary schools is organized on a geographical basis. According to this allocation rule, pupils entering either primary or middle school are sent to the different local public schools depending on their home address. During the period under study (1997-2004), school catchment areas were designed at the municipality level in primary education and by the LEA in secondary education¹⁹. Primary and middle school zoning was “strict” in the sense that each school catchment area contained one school only, which means that in principle parents could not choose their child’s public school. During the first half of the period, school catchment areas were publicly released in the form of booklets that could be looked up directly at the LEA or via parents’ associations. Starting in year 2000, they could be directly checked online through the LEA’s website²⁰. In Paris, school admission rules are different in the second stage of secondary schooling where the allocation of pupils to the different high schools (*Lycées*) combines zoning and choice. At the end of middle schooling, parents can send applications to any particular high school located within a broadly defined zone. High school principals then select their pupils among applicants using a variety of criteria, including previous academic achievement.

In the past 40 years, this system has not always been strictly enforced: starting in 1983, a certain number of *Académies* were authorized to relax the strict school zoning policy for middle schools only, thus allowing parents to choose between a number of public schools. After a few years, this new system was found unsatisfactory by many parents who had to enroll their children far away from home because of their poor academic performance. Moreover, the introduction of public school choice was seen as biased in favor of children with higher socio-economic background because their parents had better access to information about school performance. It is in this context that all *Académies* gradually switched back to strict school zoning in the

¹⁹As from school year 2004-2005, the management of school catchment areas was transferred to the county (*Département*), which is the main French subnational administrative division (France is subdivided into 100 *Départements*). This was part of a series of reforms aimed at giving more decision power to local authorities.

²⁰Since 2004, secondary school catchment areas are published on the website of the *Département* (see footnote above). For the city of Paris, they can be checked at the following address: <http://perimco.paris.fr/perimco/jsp/site/RunStandaloneApp.jsp?page=schoolsearch>.

mid-1990s. These various public middle school admission rules can be followed in the Paris LEA over the period 1963-2004. Strict school zoning was enforced between 1963 and the beginning of school year 1987-1988. It was then temporarily and partially relaxed until the beginning of school year 1997-1998. In certain *arrondissements* of the city²¹, parents of children enrolled in the last year of primary schooling were allowed to rank in order of preference three different middle schools out of all those located in their *arrondissement*. Pupils were then selected by school principals on the basis of their school results. This policy experiment was gradually reversed and definitely abandoned in September 1997, when the system moved back to strict school zoning²².

However rigid the French strict school zoning system may appear, it allows some exceptions. There are two ways parents can get round school catchment areas without actually changing residence. First, they can ask the LEA for a dispensation that entitles them to send their children to a school located outside their attendance zone. These dispensations can be granted on several grounds: if specific options or “rare” languages (*e.g.* Russian or Japanese) are not taught in the local school, if a child’s sibling is enrolled in a different school or exceptionally if the local school is located much further away from home than a school belonging to an adjacent zone²³. Every year, dispensations are granted to about 8% of Parisian pupils entering middle schooling, the rate of rejection being around 40%. While a substantial fraction of these dispensations have true practical justifications, some parents use them to avoid what they perceive as low-performing local middle schools. The second way of getting round the zoning system is to use the outside option provided by the extensive network of highly subsidized private middle schools.

²¹The city of Paris comprises 20 *arrondissements*, which are the main administrative subdivisions of the municipality. Each *arrondissement* has its own mayor (*maire d’arrondissement*) and a directly elected council. The main purpose of this administrative unit is to answer the queries of citizens regarding birth or marriage recordings, official business, etc. However, local taxation, economic development and the provision of most public goods are decided upon by the Paris city council (*Mairie de Paris*), which is presided by the Mayor of Paris.

²²Unfortunately, access to information on schools and pupils before year 1997 is currently restricted, impeding the study of the pre-1997 school admission system. We therefore had to limit our study to post-1997 school years.

²³This might happen when the local school is not located in the center of its catchment area.

3.3.2.2 Private schools

An important characteristic of private education in France is that it is not subject to any zoning scheme. Because they can select their pupils from anywhere in the city, private schools offer a relatively cheap (but subject to rationing) outside option to parents who are willing to avoid the constraints of strict school zoning. This particular feature explains that although private schools are predominantly Catholic, the choice of private education is now driven by educational rather than by purely religious motives²⁴ (Langouët and Léger, 1997). Given the capacity constraints imposed by State restrictions on the size of the private sector, private schools principals have the right to select their pupils among the pool of applicants. Admission rules are not always explicitly stated²⁵ and vary from school to school. The most commonly used criteria are academic excellence, presence of an older sibling in the school, good behavior and commitment to the school's values²⁶.

3.3.3 The importance of middle schooling

While most existing empirical studies of the impact of school performance on housing prices have focused on primary schools, we investigate what happens during the first part of secondary education. The main reason for taking this approach is that in the French school system, middle schooling is the educational stage that is most likely to have the largest influence on housing markets.

The conjunction of strict middle school zoning and relatively wide high school choice explains the important role played by middle schooling among other determinants of residential choice. As explained above, admission to high schools does not rely on a strong geographical basis and is instead jointly determined by individual performance and by the academic reputation of the *Collège* of origin. This is par-

²⁴As an indication of this, families do not hesitate to switch between public and private sectors, especially when their children reach the middle or high school level. For example, during school year 2004, about a quarter of pupils enrolled in Parisian private middle schools came from public Parisian primary schools (and 3% of pupils enrolled in public middle schools came from a private primary school).

²⁵It should be mentioned however that over the past few years, an increasing number of private schools have decided to publish their selection policies and procedures on their websites.

²⁶Note that applicants to Catholic middle schools are usually not required to be baptized or to have been previously enrolled in a religious primary school.

ticularly true in the city Paris where parents can send applications to any particular high school located within their attendance area, the city being divided into 4 large attendance areas (North, South, East and West). There are some very good, over-subscribed public high schools within each zone (usually located towards the city center), as there are less popular, under-subscribed ones (often in the periphery). Competition to enter the best institutions is quite vivid, and parents are conscious of the importance of sending their children to a good middle school as a means of increasing their chances of being admitted into a good high school. Given the stringency of public middle school zoning, there are reasons to believe that the school performance factor should act as an important determinant of residential choice in the Parisian context.

The city of Paris combines a number of features that make it an especially appropriate setting to study the impact of school performance on housing prices. First, school-driven residential competition is likely to be fierce because of the large number of unequally performing public middle schools that operate in the city. Second, the large share of private middle school enrollment (one third of total enrollment) makes it possible to investigate how the housing market incorporates public school performance when there is some degree of school choice. Finally, the large human and spatial scale of the city, combined with the fact that Paris is a single administrative entity and a single LEA, should provide enough detail and variation to yield credible and precise estimates.

3.4 Data and summary statistics

To estimate the impact of school performance on housing sales in Paris, we have collected data on school zones, school characteristics, individual property sales and local socio-demographic characteristics during school years 1997-2004.

3.4.1 School catchment areas in Paris

Data on school zones was provided by the Paris LEA (*Rectorat de l'Académie de Paris*). During the period under study, the LEA was in charge of drawing the

assignment zones of primary and middle schools. Because of their central role in our estimation strategy, school catchment areas are worth examining in some detail.

School catchment areas are released every school year in the form of booklets that indicate, for each street section, the assigned public middle school. Using the complete set of geolocalized addresses of the city of Paris, we were able to map each of the 108 public middle school catchment areas for every school year between 1997 and 2003. Figure 3.2 displays the particular layout of school catchment areas that prevailed in 2003-2004. This map shows that in most cases, middle school zones are contiguous. Still, a closer look reveals that if the typical school catchment area is made of one single piece, many catchment areas are actually splitted into two or three parts, as can be noticed from the particular layout of catchment areas displayed in figure 3.3 (see for example the Henri Bergson middle school zone). To our knowledge, this feature is not present in previously studied school attendance zoning plans. Another interesting characteristic is that many school zone boundaries are not straight, but rather have a zigzag shape. Given our estimation strategy, middle school boundaries thus seem to share two highly desirable properties: first, they tend to split otherwise similar neighborhoods, a feature that is needed for credible identification; second, they seem numerous enough to yield precise estimates.

A serious concern that has been raised about using school zone boundaries as an exogenous source of variation in public school performance is that they are likely to coincide with other administrative divisions, which might be associated with other discontinuities than school performance. In the US, school districts usually include several municipalities and sometimes cross different counties; in the UK, school zone boundaries are strictly defined only between different LEAs, which themselves coincide exactly with counties, metropolitan districts or London boroughs. In these countries, observed price differential across school zone boundaries could in some cases reflect differences in local tax rates or public goods, in addition to differences in school performance. In the Parisian context, school catchment areas can be clearly distinguished from other administrative divisions. The first thing to note is that the Paris LEA covers a single municipality: local tax rates are therefore equal throughout the covered area. Moreover, figure 3.4 shows that each of the 20 administrative

Figure 3.2: *School catchment areas in Paris during school year 2003-2004.*



Figure 3.3: School catchment areas in the 19th arrondissement of Paris. School year 2003-2004.



subdivisions of the city of Paris (*arrondissements*), which may differ at the margin in terms of public goods provision, comprises several school attendance zones. To ensure that these administrative subdivisions do not contaminate our estimations, we decided to exclude from the data all school attendance boundaries that coincide with *arrondissements* boundaries. By definition, this exclusion rule also eliminates the middle school boundaries that match high school attendance boundaries, since the limits of the four Parisian high school catchment areas happen to follow exactly the *arrondissements* borders, as shown in figure 3.5. Although we do not have data on the precise layout of primary school catchment areas, we are confident that their boundaries do not coincide with middle school boundaries. First, primary school catchment areas are much smaller than those of middle schools, since the number of public primary schools is three times larger than the number of public middle schools. Second, our data indicates that in the city of Paris, children enrolled in a given public primary school are not assigned to a unique middle school. On average, two thirds go to the middle school that accounts for the largest share of pupils coming from that particular primary school, 17% go to the middle school that accounts for the second largest share and the remaining 17% transfer to other middle schools. These statistics suggest that primary school boundaries seldom follow middle school boundaries.

A final important issue raised by school zone boundaries is their degree of stability over time. Very unstable boundaries would represent a serious threat to our estimation strategy, since public school performance would then have little chance of showing up in property prices on the borderline between adjacent school zones. The information that we gathered through informal talks with the LEA officials in charge of school catchment areas in Paris revealed that the reassignment process is essentially driven by demographic reasons. By the end of each school year, the LEA uses the information at its disposal on the residential location of primary school pupils who are about to enter middle school to predict the following year's middle school enrollment figures. In the event of important imbalances between the number of incoming pupils and the capacity of schools, the LEA chooses either to reallocate resources between schools (new classes being opened in certain schools while existing classes are closed in others) or to redraw school attendance boundaries in order to restore the

Figure 3.4: *School catchment areas and arrondissements boundaries in Paris in 2004.*

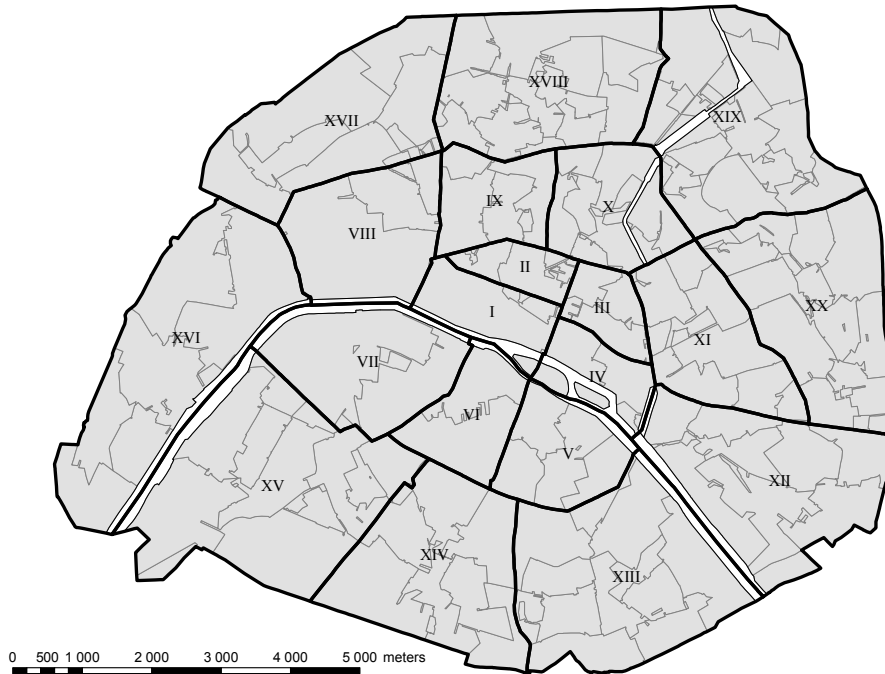
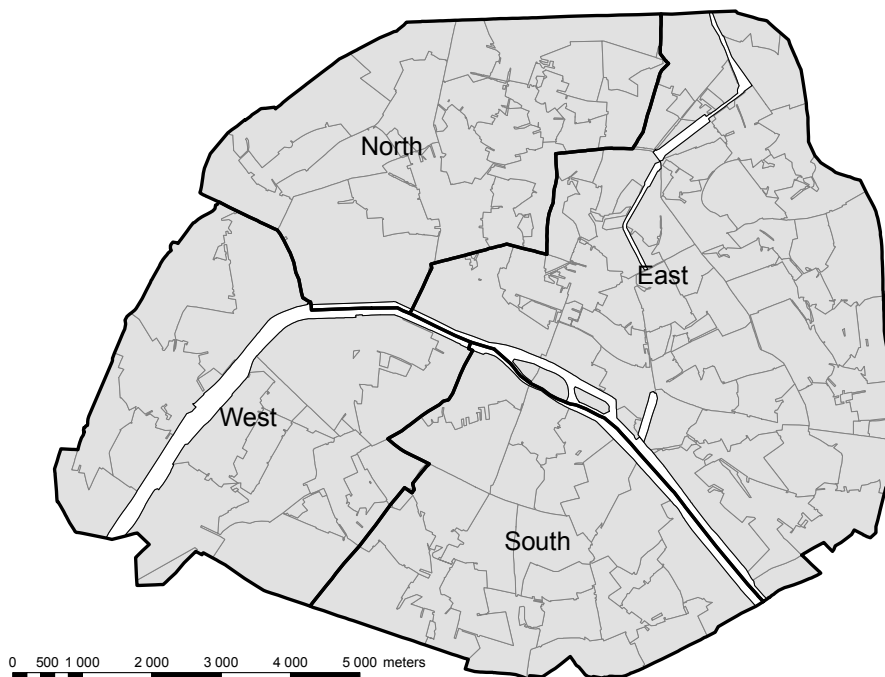


Figure 3.5: *School catchment areas and boundaries of high school districts in Paris in 2004.*



balance between expected enrollment and school capacity. Particular reassignments between school catchment areas are highly unpredictable by parents. However, they cannot be considered as exogenous events, since they usually coincide with underlying demographic trends that may affect the housing market independently from school performance. Hence we decided not to use these reassignments and to restrict our estimations to non-reassigned areas. Fortunately, our data show that over the period 1997-2004, school catchment areas in Paris remained largely unchanged, since less than 10% of all Parisian addresses were reassigned over this period of 7 years. It seems therefore unlikely that the instability of school attendance boundaries could invalidate our identification strategy.

3.4.2 Public and private schools

We retrieved information on school characteristics and computed measures of school performance from several datasets, all of which were provided by the Statistical Department of the French Ministry of Education²⁷.

3.4.2.1 School characteristics

General information about the characteristics of public and private middle schools operating in the Paris LEA can be found for school year 20003-2004 in the *IPES*²⁸ dataset. We use this dataset to compute three groups of indicators. The first group provides information on school size. The second group documents the characteristics of pupils: proportion of pupils coming from privileged socio-economic backgrounds²⁹, proportion of foreign pupils, proportion of females, fraction of pupils that repeat their first year of middle school and fraction of pupils in their final year of middle school who are admitted into general curriculum high schools (*Lycées d'enseignement général*) as opposed to those who enter vocational high schools (*Lycées professionnels*). The third group of indicators includes information on teachers' characteristics, available for

²⁷ *Direction de l'Évaluation, de la Prospective et de la Performance du Ministère de l'Éducation nationale (MEN-DEPP).*

²⁸ *Indicateurs pour le Pilotage des Établissements Secondaires.*

²⁹ A pupil is classified as coming from a privileged socio-economic background if the household head is either a professional (lawyer, practitioner, etc.), a general manager, a middle manager, a professor or a teacher.

public schools only: proportion of females, fraction aged under 40 and the distribution of the different levels of teacher qualifications³⁰.

3.4.2.2 School performance indexes

Our empirical analysis requires information on the performance of public and private schools.

The exact characteristics that parents value in schools are subject to an extensive debate in the literature. A number of studies have produced evidence that parents' school choices react to test-based school rankings (Black, 1999; Figlio and Lucas, 2004; Hastings et al., 2007). However, as pointed out by Mizala and Urquiola (2008), the problem with this evidence is that test scores are highly correlated both with peer group composition and, to a lesser extent, with school effectiveness (or "value added"). While disentangling these two components is a difficult task, recent research suggests that peer quality could well be the dominant factor (Rothstein, 2006). Mizala and Urquiola (2008) reach the same conclusion using data on Chile's SNED program³¹.

The institutional features of the French educational system outlined in section 3.3, as well as anecdotal evidence, suggest that parental perception of middle school performance relies heavily on peer quality. There are two main reasons for that. First, as previously noted, the highly centralized organization of middle schooling and the partially randomized allocation of teachers to the different establishments are unlikely to produce big differences in school effectiveness. Second, value-added measures of school performance have never been publicly released for middle schools, whereas they have long been available for high schools. The information that parents can use to rank middle schools thus relies exclusively of exam scores and peer composition. Official league tables showing the average exam scores at the middle school level were

³⁰In the French education system, a teacher's level of qualification depends on the degree she holds: in our classification, "high" stands for the *Agrégation* degree, "intermediary" for the *CAPES* degree and "low" for any lower degree.

³¹This program was aimed at helping parents identify the most effective schools, the information provided being quite different from that conveyed by a simple test-based ranking of schools. The authors' key finding is that schools' market outcomes were unaffected by the SNED program.

only recently introduced in France³² and are now widely accessible³³. It must be noted however that to satisfy parents' demand for information about middle school performance, rankings of Paris middle schools based on exam results were regularly published in the local press as from the beginning of the 1990s. For the period under study (1997-2004), parents living in the city of Paris can thus be considered to have a relatively good knowledge of how middle schools compare to each other in terms of pupil performance.

In our empirical analysis, we will use three distinct indexes to measure school performance: the average exam scores at the school level (DNB SCORE), the fraction of middle school pupils who are admitted into the high school general curriculum as opposed to vocational studies (GENERAL CURRICULUM) and the proportion of middle school pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND).

Individual exam score data could be retrieved for the entire Paris *Académie* from the *OCEAN*³⁴ national exam results database. Unfortunately, we could only use information for school year 2003-2004 since the exam results of previous years were not available. The *OCEAN* dataset records the individual score obtained by every pupil enrolled in the Paris LEA at the *Diplôme National du Brevet (DNB)*. This exam is taken by all French pupils in their final year of middle school (*Troisième*). Individual scores are a equally weighted combination of a continuous assessment and a final national examination, which is graded anonymously at the LEA level and consists in three parts: Math, French and History & Geography. Each section is scored out of 20. To ensure that our measure of performance is comparable across schools, we use only the national exam component of the *DNB* score. The performance of any particular school is therefore computed as the average Math, French and History & Geography score obtained by pupils at the *DNB* exam.

Because our exam score-based index is computed using a single year of data, one

³²The official justification for not publishing average exam scores was that doing so would have contradicted the very nature of strict school zoning, which theoretically prohibits parents to choose a public middle school outside of their local catchment area.

³³Examination results at the school level are consultable via the Internet (See for instance <http://f23.www.france-examen.com>).

³⁴*Organisation des Concours et Examens Académiques et Nationaux.*

cannot exclude that it acts a noisy measure of the medium-run school performance, which arguably determines parents' school choices. While the pitfalls associated with the use of short-run measures of school performance have been extensively discussed in the literature (Kane and Staiger, 2002; Chay et al., 2007; Mizala et al., 2007), we believe that they are likely to be less severe for middle schools, because their enrollment is typically larger than that of primary schools, a feature that mechanically reduces year-to-year variation in a school's average score. The median number of pupils taking the *DNB* exam in each middle school belonging to the Paris LEA in 2004 was 125, which is three to four times larger than the median enrollment used in studies that have called into question the reliability of school accountability measures. In any case, we decided to use an alternative index of school performance that would better reflect the medium-run performance of schools over the period 1997-2004. Using the *SCOLARITE* dataset³⁵, we were able to calculate for each school and each year the percentage of pupils in their final year of middle school who are admitted into general curriculum high schools the following year, as opposed to those who start vocational studies. These annual figures are then averaged at the school level over the period 1997-2004. In the French context, this variable can be considered as a good indirect measure of school performance, since it is closely linked to educational attainment and varies greatly across schools. It should nonetheless be noted that admission into general curriculum high schools is not tied to any particular threshold in the *DNB* exam score, but rather to a global assessment of individual performance by the teachers.

Our final and third index of school performance is a direct measure of peer quality. As explained earlier, there are many reasons to believe that in the French context, parents are highly concerned about peer group composition in their valuation of

³⁵The *SCOLARITE* dataset is available every year over the period 1997-2004. It contains individual information on all French pupils enrolled in public or private middle and high schools. The datasets contains information on each pupil's age, gender, citizenship, occupation of the household head, *arrondissement* of residence, school attended in the current (t) and previous ($t-1$) year as well as current and previous school educational level. However, this very rich dataset suffers from two limitations. First, the panel dimension of the data cannot be exploited because access to the pupil identifier is restricted. Second, this dataset does not contain the pupils' results to examinations and cannot be individually matched with the previously mentioned *OCEAN* national exam results database.

middle schools³⁶. Using the *SCOLARITE* dataset, we computed the fraction of pupils in each middle school that come from privileged socio-economic backgrounds, averaged over school years 1997 to 2004.

These three indexes are standardized by dividing each school's value by the corresponding standard deviation of school averages (including both public and private schools).

3.4.2.3 Summary statistics

Table 3.1 compares the characteristics of public and private schools during school year 2003-2004. These characteristics are displayed for different samples of Parisian middle schools: all public and private schools (column 1), public schools only (column 2), private schools only (column 3) and public schools splitted into four quartiles of average *DNB* exam score (columns 4 to 7).

Public middle schools account for 68% of total enrollment. On average, private schools are smaller than public schools: during school year 2003-2004, the average private school enrollment was 420, against 531 for public schools. Unsurprisingly, private school pupils enjoy much more favorable characteristics than their public school counterparts: 57% come from privileged socio-economic backgrounds *versus* 32% for public school pupils; only 4% are of foreign nationality *versus* 18% in public schools. The higher peer quality of private schools naturally translates into better pupil performance: with an average score of 10.97 points out of 20 on the *DNB* exam, private school pupils outperform their public school peers (9.42 points) by almost a standard deviation of distribution of the *DNB* score averaged at the school level. Moreover, they are more likely to be admitted into general curriculum high schools (78% *versus* 67%).

The last four columns of table 3.1 show that public schools are very heterogenous in their pupils' characteristics. Their performance is very highly correlated with peer quality: pupils enrolled in schools belonging to the top quartile of 2004 average *DNB* exam score come in majority from privileged socio-economic background (56%

³⁶A rather striking symptom of this parental concern for peer quality, especially in the Parisian context, is that maps showing how middle schools compare in their socio-economic composition have been regularly published in the local press.

Table 3.1: Summary statistics: Public and private middle schools in Paris in 2004. Sources: IPES dataset (2004) and OCEAN national examinations database (2004).

Variables	All schools (1)	Public schools (2)	Private schools (3)	Public schools			
				Quartiles of mean <i>DNB</i> exam score			
				Q1 (4)	Q2 (5)	Q3 (6)	Q4 (7)
<u>SCHOOL SIZE</u>							
Total enrolment	85,237	57,918	27,319	13,042	14,040	13,794	17,042
Average enrolment	490 (196)	531 (145)	420 (246)	483 (131)	520 (113)	511 (121)	609 (179)
Pupil-to-teacher ratio		14.5 (7.0)		12.1 (3.5)	13.6 (6.6)	14.6 (5.3)	17.5 (10.1)
<u>PUPILS' CHARACTERISTICS</u>							
% female	0.48 (0.11)	0.49 (0.03)	0.48 (0.18)	0.48 (0.03)	0.48 (0.03)	0.49 (0.02)	0.50 (0.04)
% with privileged socio-economic background	0.41 (0.24)	0.32 (0.20)	0.57 (0.22)	0.11 (0.07)	0.24 (0.12)	0.35 (0.14)	0.56 (0.12)
% foreign	0.13 (0.11)	0.18 (0.10)	0.04 (0.05)	0.27 (0.11)	0.20 (0.09)	0.15 (0.07)	0.09 (0.05)
% grade repetition in first year	0.07 (0.09)	0.06 (0.04)	0.08 (0.13)	0.09 (0.04)	0.06 (0.04)	0.06 (0.04)	0.05 (0.03)
% going into general curriculum	0.71 (0.17)	0.67 (0.15)	0.78 (0.18)	0.49 (0.09)	0.64 (0.08)	0.72 (0.06)	0.84 (0.09)
Mean <i>DNB</i> exam score (out of 20)	9.99 (1.89)	9.42 (1.62)	10.97 (1.94)	7.28 (0.83)	8.94 (0.31)	10.01 (0.30)	11.38 (0.70)
<u>TEACHERS' CHARACTERISTICS</u>							
% female		0.70 (0.08)		0.62 (0.08)	0.70 (0.08)	0.73 (0.06)	0.72 (0.07)
% aged under 40		0.36 (0.13)		0.50 (0.10)	0.36 (0.10)	0.29 (0.09)	0.28 (0.09)
% with high qualification		0.12 (0.06)		0.10 (0.05)	0.10 (0.05)	0.12 (0.05)	0.16 (0.08)
% with intermediary qualification		0.70 (0.08)		0.69 (0.08)	0.70 (0.08)	0.70 (0.09)	0.70 (0.08)
% with low qualification		0.18 (0.08)		0.21 (0.08)	0.20 (0.08)	0.18 (0.08)	0.14 (0.06)
Number of schools	174	109	65	27	27	27	28

Notes: The average 2004 *Diplôme national du Brevet (DNB)* score at the school level is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. School size, pupils' and teachers' characteristics are computed from the *Indicateurs pour le Pilotage des Établissements Secondaires (IPES)* dataset (2004). Standard deviations are in parenthesis.

versus 11% in the bottom quartile, which represents a difference of over twice the standard deviation), are much less likely to be of foreign nationality (9% *versus* 27%) and almost systematically enter general curriculum high schools (84% *versus* 49%). The comparison of school size and teachers' characteristics in the bottom and top quartiles of school performing shows much less variation, which is consistent with the idea that school peer group composition is the primary source of variation in French public middle school performance. With an average enrolment of 609 pupils, schools belonging to the top quartile of 2004 average *DNB* exam score are larger than those belonging to the bottom quartile (average enrollment of 483) and their pupil-to-teacher ratio is larger by about half a standard deviation (17.5 *versus* 12.1). If teachers working in the bottom quartile schools are younger (50% below the age of 40 *versus* 28%) than their colleagues working in the top quartile schools, they are only slightly less qualified (10% with high qualification *versus* 16%), as one would expect given the partially randomized process that governs the allocation of teachers to schools.

Table 3.2: *Correlation between the three standardized indexes of public middle school performance. Sources: OCEAN (2004) national examinations database and SCOLARITE pupil database (1997-2004).*

Public middle school performance index	DNB	GENERAL	PRIVILEGED
DNB SCORE	1.00		
(p-value)	<.0001		
GENERAL CURRICULUM:	0.84	1.00	
(p-value)	<.0001	<.0001	
PRIVILEGED BACKGROUND:	0.86	0.90	1.00
(p-value)	<.0001	<.0001	<.0001
Number of public middle schools	108		

Notes: The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004).

Table 3.2 shows the correlation between our three standardized indexes of school performance: the school average score at the 2004 *DNB* exam, the fraction of middle school pupils entering general curriculum high schools and the proportion of pupils

that come from privileged socio-economic backgrounds, these two last measures being averaged over years 1997-2004. As expected, these three indexes are very highly correlated, the correlation coefficients ranging from 0.84 to 0.90. These figures indicate that during the period under study, the performance of public middle schools was very stable. This is an important feature in the Parisian context, since school performance can be assumed to be capitalized in prices only if it not too volatile across school years.

Figures 3.6 to 3.8 display the spatial distribution of public middle school performance across the different school catchment areas in Paris. Given the high degree of correlation between our three indexes of school performance, these three maps share a very similar pattern: the best public middle schools tend to be located in the central and western parts of the city while the lowest-performing schools are usually found in the north-eastern quarter. However, these maps also indicate that the transition between different levels of school performance is not smooth and that many adjacent schools perform very unequally (especially in the eastern part of the city). Given our identification strategy, this is a desirable feature because large school performance differentials should yield more precise estimates.

3.4.3 Housing prices

Data on property sales come from the *BIEN*³⁷ dataset which is produced by the Notary Chamber of Paris and the Île-de-France³⁸. In France, all property sales have to be registered by a notary who collects the realty transfer fee to be paid to the Inland Revenue. The Notary Chamber has gathered the data for Paris and the Île-de-France since the early 1990s. The dataset is almost comprehensive and contains between 80 and 90% of all the transactions that took place since 1997³⁹. For each transaction, we have information on the price for which the property was sold, along with its detailed characteristics (size, number of bedrooms and bathrooms, date of construction, etc.) and its precise spatial location (Lambert II grid coordinates) with

³⁷ *Base d'Informations Économiques Notariales.*

³⁸ *Chambre des Notaires de Paris et d'Île-de-France.*

³⁹ In the early 1990s, when the Notary Chamber started to collect the data, not all Notary offices succeeded in transmitting their own data. However, the coverage rate has continuously improved since.

Figure 3.6: *Spatial distribution of the quartiles of public middle school performance as measured by the average Diplôme National du Brevet exam score in 2004. Source: OCEAN national examinations database (2004).*

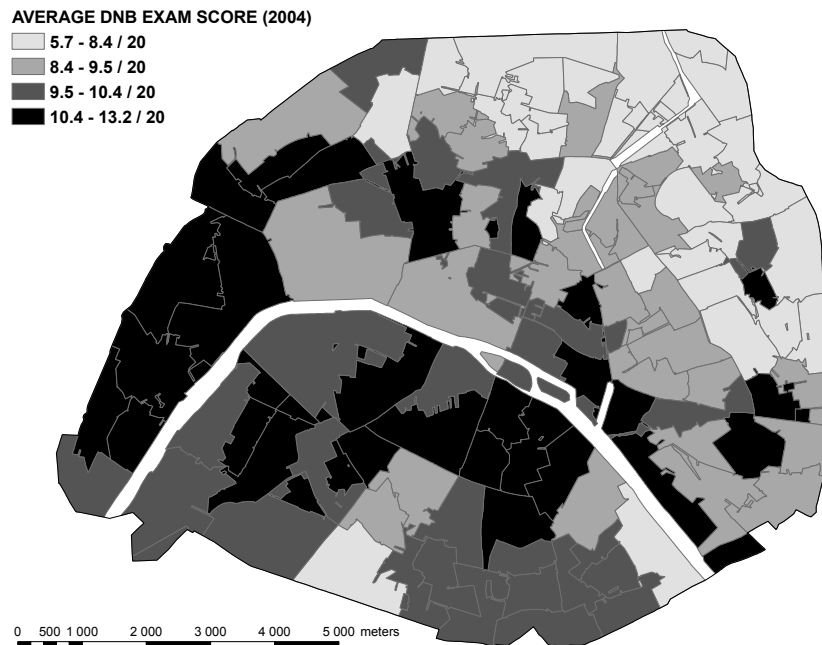


Figure 3.7: *Spatial distribution of the quartiles of public middle school performance as measured by the fraction of pupils admitted into general curriculum high schools (averaged over the period 1997-2004). Source: SCOLARITE database (1997-2004).*

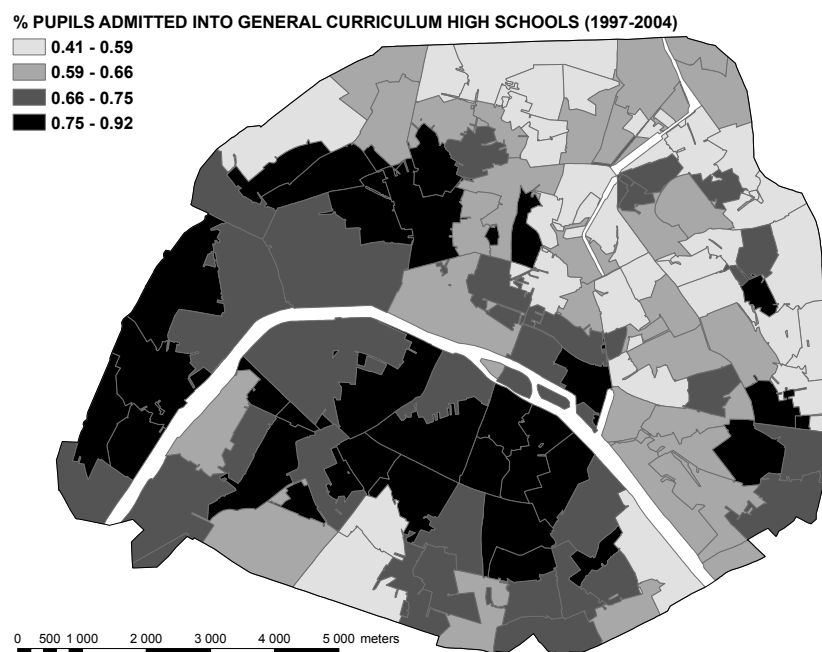


Figure 3.8: *Spatial distribution of the quartiles of public middle school performance as measured by the fraction of pupils from privileged socio-economic backgrounds (averaged over the period 1997-2004). Source: SCOLARITE database (1997-2004).*

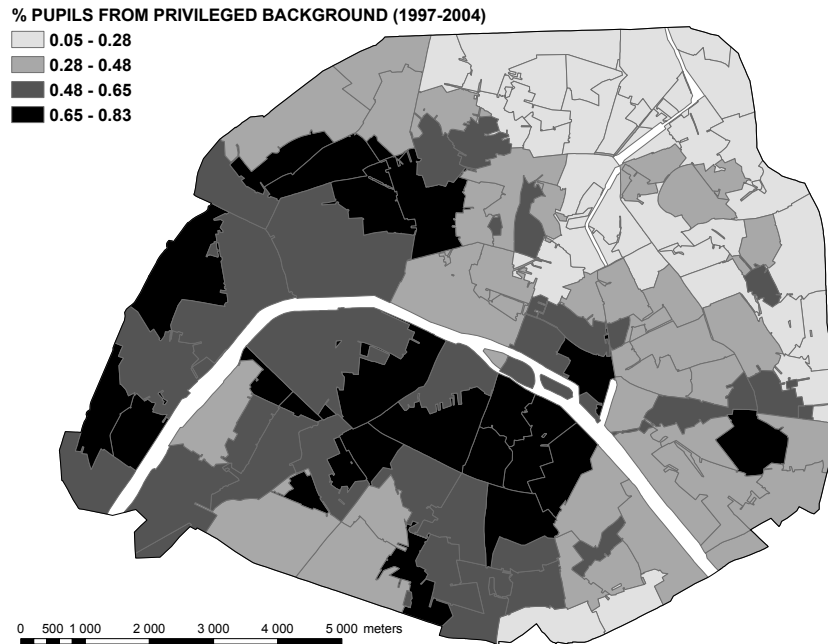
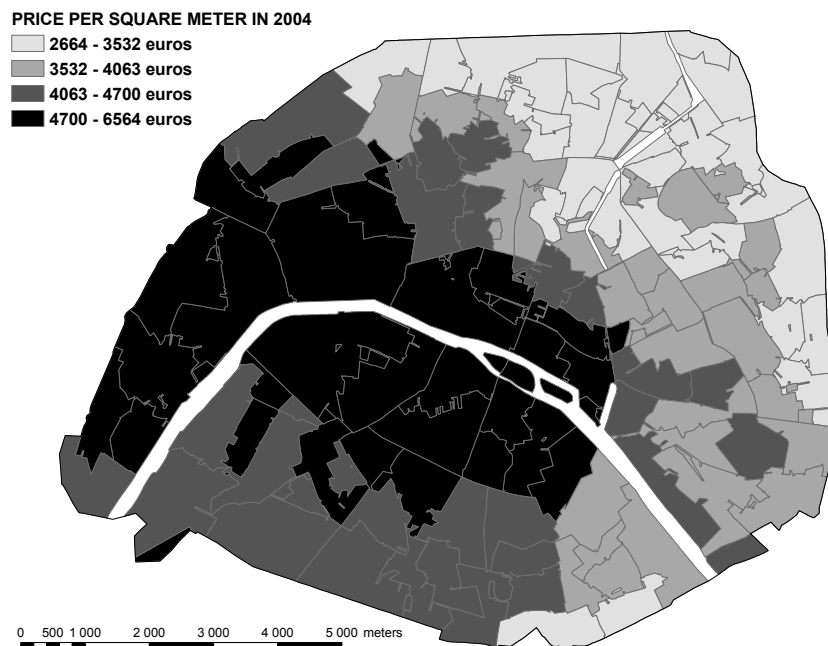


Figure 3.9: *Spatial distribution of the quartiles of average price per square meter (in 2004 euros) within each public middle school catchment area in school year 2003-2004.*



a precision of the order of 5 meters. Our sample is restricted to all arm's-length sales of second hand Parisian flats⁴⁰ that took place between September 1997 and August 2004. We further excluded sales belonging to the top and bottom percentiles of housing prices per square meter each year. We are left with a sample of about 200,000 transactions.

The first two columns of table 3.3 displays the characteristics of Parisian flats in the full sample of sales. The mean flat price in our sample is 183,054 euros with a standard deviation of 180,832 euros. Parisian flats are relatively small, the mean flat size being 52 m² (with a standard deviation of 35 m²) and are rarely made up of more than two rooms. Most buildings were built between the second half of the Nineteenth century and the the First World War and usually have 5 to 6 floors. Very few flats come with a parking space or a maid's room.

By combining the precise geographic coordinates of each sale in Paris and the mapping of school catchment areas, we were able to identify each transaction's assigned public middle school between 1997 and 2004. Figure 3.9 shows how the price per square meter (in 2004 euros) of sales that took place during school year 2003-2004 varies across the 108 corresponding public middle school catchment areas. The spatial distribution of housing prices across the city is fairly close to the spatial distribution of school performance displayed in figures 3.6 to 3.8. It is however much more continuous, the housing prices decreasing gradually as we move away from the central and western parts of the city. This feature suggests that housing prices and school performance are highly correlated, although no causal interpretation can yet be given to this phenomenon.

3.4.4 Neighborhood socio-demographic characteristics

We use the 1999 French National Census⁴¹ at the district level (*Iris*) to control for the socio-demographic characteristics of Parisian neighborhoods. The city of Paris is

⁴⁰We dropped newly-built property sales because their price differs greatly from the price of second-hand sales and because new properties represent only a very small share of overall property sold in Paris. For the same reasons, we excluded the few houses that were present in the sample. We also dropped transactions when the price or the number of rooms was missing. Finally, we decided to exclude the Île Saint-Louis and Île de la Cité islands, because of their very specific location and pattern of housing prices.

⁴¹*Recensement de la Population française 1999.*

Table 3.3: *Summary statistics: characteristics of flats located within 250, 300 and 350 meters from a school attendance boundary and sold during school years 1997 to 2004. Source: BIEN dataset.*

Distance to boundary	All sales		< 250 m		< 300 m		< 350 m	
	Mean (1)	s.d. (2)	Mean (3)	s.d. (4)	Mean (5)	s.d. (6)	Mean (7)	s.d. (8)
Price (in 2004 Euros)	183,054	(180,832)	181,915	(168,997)	183,516	(171,867)	184,804	(175,152)
Flat size (in m ²)	52	(35)	51	(34)	52	(34)	52	(35)
Price per m ² (in 2004)	3,284	(1,287)	3,320	(1,242)	3,325	(1,255)	3,330	(1,263)
FLAT CHARACTERISTICS								
Age of building (percent)								
Unknown	0.11		0.11		0.11		0.11	
Before 1850	0.05		0.05		0.05		0.05	
1850-1913	0.42		0.41		0.41		0.41	
1914-1947	0.15		0.16		0.16		0.16	
1948-1969	0.13		0.13		0.13		0.13	
1970-1980	0.11		0.11		0.11		0.11	
After 1981								
Floor (percent)								
Ground floor	0.10		0.11		0.11		0.11	
First	0.16		0.16		0.16		0.16	
Second	0.16		0.16		0.16		0.16	
Third	0.16		0.15		0.16		0.15	
Fourth	0.14		0.14		0.14		0.14	
Fifth	0.12		0.12		0.12		0.12	
Sixth or more	0.16		0.15		0.15		0.15	
Number of rooms (percent)								
One	0.24		0.24		0.24		0.24	
Two	0.37		0.36		0.36		0.36	
Three	0.22		0.22		0.22		0.22	
Four	0.10		0.10		0.10		0.10	
Five or more	0.07		0.07		0.07		0.07	
% without ind. bathroom	0.21		0.21		0.21		0.21	
% with maid's room	0.04		0.04		0.04		0.04	
% with lift	0.90		0.90		0.90		0.90	
% with parking space	0.13		0.12		0.12		0.12	
Number of sales	196,815		99,917		113,531		124,608	

Notes: The features of flats sold during school years 1997 to 2004 are computed from the *BIEN (Base d'Informations Économiques Notariales)* dataset, which is managed by the Notary Chamber of Paris and the Île-de-France. The sample is restricted to all arm's-length sales of second hand Parisian flats that took place between September 1997 and August 2004. Sales located on the Île Saint-Louis and Île de la Cité islands were excluded from the sample because of their very specific location and pattern of housing prices.

composed of 970 districts, a few of which are inhabited (parks, hospitals, railroads, etc.). The average number of residents living in a census district is 2,191.

Our controls include the average number of persons per flat, the proportion of families (*i.e.* households with at least one child below the age of 25), the proportion of public housing apartments, the proportion of owners, of graduates, of single-parent families, of foreigners, as well as information on the employment status and occupation of the household head (manager, self-employed, intermediary occupation, employee, manual worker, retired and economically inactive).

Summary statistics on socio-demographic characteristics are reported in the first two columns of table 3.4. These figures indicate that the population living in the city of Paris is rather socially privileged: 39% of Parisians hold a graduate degree, only 15% live in public housing apartments and among the population aged 15 or more, 40% are either managers, self-employed or hold an intermediary occupation whereas employees and manual workers represent only 21% of the total. Paris also appears as a relatively cosmopolitan city, the proportion of foreigners among household heads amounting to about a quarter. Finally, about one third of households own their apartment. The size of the standard deviations reveals however that the socio-economic composition of the city is far from being homogenous across the different neighborhoods.

3.5 Results

In this section, we apply our matching framework to estimate the impact of school performance on housing quality in the city of Paris. To investigate the magnitude of the bias induced by the endogeneity of school performance in the housing price equation, we start by computing the naive estimates using the traditional hedonic framework, before showing the results obtained when sales are matched across school attendance boundaries.

Table 3.4: *Summary statistics: Neighborhood characteristics and average public middle school performance at the district level for all sales and sales located within 250, 300 and 350 meters of a school attendance boundary, sold during school years 1997 to 2004. Sources: 1999 National Census and school catchment areas in academic year 1998-1999.*

Distance to school boundary	All sales		< 250 m		< 300 m		< 350 m	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>CENSUS VARIABLES</u>								
Nb of census districts	900		726		761		785	
Nb of individuals per district	2,292	(790)	2,345	(718)	2,343	(735)	2,333	(740)
Nb of households per district	1,204	(424)	1,244	(389)	1,240	(398)	1,234	(401)
Nb of persons per flat	1.89	(0.25)	1.87	(0.23)	1.87	(0.23)	1.88	(0.23)
<u>All households</u>								
% families with children under 25	0.22	(0.08)	0.22	(0.07)	0.22	(0.07)	0.22	(0.07)
% female-headed households	0.06	(0.03)	0.06	(0.02)	0.06	(0.02)	0.06	(0.03)
% owners	0.30	(0.11)	0.31	(0.10)	0.31	(0.10)	0.31	(0.10)
% public housing	0.15	(0.22)	0.13	(0.19)	0.13	(0.19)	0.13	(0.20)
% with graduate degree	0.39	(0.11)	0.40	(0.09)	0.40	(0.09)	0.40	(0.09)
% foreigners	0.23	(0.06)	0.22	(0.05)	0.22	(0.06)	0.22	(0.06)
% unemployed	0.12	(0.04)	0.11	(0.04)	0.11	(0.04)	0.12	(0.04)
Occupation:								
% managers	0.22	(0.07)	0.23	(0.06)	0.23	(0.06)	0.23	(0.06)
% self-employed workers	0.04	(0.02)	0.04	(0.02)	0.04	(0.02)	0.04	(0.02)
% intermediary occupation	0.15	(0.05)	0.14	(0.04)	0.14	(0.04)	0.15	(0.04)
% employees	0.14	(0.04)	0.14	(0.04)	0.14	(0.04)	0.14	(0.04)
% manual workers	0.06	(0.04)	0.06	(0.04)	0.06	(0.04)	0.06	(0.04)
% retired	0.18	(0.05)	0.18	(0.04)	0.18	(0.04)	0.18	(0.04)
% economically inactive	0.21	(0.05)	0.21	(0.05)	0.21	(0.05)	0.21	(0.05)
<u>SCHOOL PERFORMANCE</u>								
Average score at the 2004 DNB exam	9.40	(1.62)	9.47	(1.56)	9.47	(1.56)	9.47	(1.56)
% entering general curriculum high schools	0.66	(0.12)	0.67	(0.12)	0.67	(0.12)	0.67	(0.12)
% with privileged backgrounds	0.44	(0.20)	0.44	(0.20)	0.44	(0.20)	0.44	(0.20)
Number of public schools	108		105		105		105	

Notes: Neighborhood characteristics are computed at the district level from the French 1999 National Census. The city of Paris is composed of 970 distinct districts (*Irises*), some of which are uninhabited. The average population living in a census district is 2,191 residents. Flat prices are expressed in 2004 euros. Distance to school attendance boundaries is computed by combining the Lambert II geocoding of sales with the mapping of school catchment areas in academic year 1998-1999. The average school performance indexes within each school are computed from the *OCEAN* national examination database (2004) and the *SCOLARITE* pupil database (1997-2004).

3.5.1 Naive estimates

The first two columns of table 3.5 show the results of estimating the hedonic housing price equation (3.2), with and without controlling for the socio-demographic characteristics of the census district, using our three indexes of public middle school performance.

Column 1 indicates that when no controls are added except for year and quarter, the naive estimation of the impact of school performance on housing prices yields a strong apparent positive impact: for every standard deviation increase in the local public middle school's *DNB* score, the housing price per square meter goes up by 21.8%. For the other two indexes (fraction of pupils entering general curriculum high schools and proportion of pupils from privileged socio-economic backgrounds), a standard deviation increase in school performance is associated with respectively a 16.1% and a 19.0% increase in housing prices.

Column 2 shows that the coefficients are not particularly affected when one adds controls for the flats' characteristics, which seems to indicate that in the city of Paris, sales located in the attendance area of high-performing schools do not share on average more desirable features than those assigned to low-performing schools: the coefficients on the *DNB SCORE*, *GENERAL CURRICULUM* and *PRIVILEGED BACKGROUND* indexes only slightly decrease to 19.4%, 14.2% and 17.3% respectively.

However, because the best schools tend to be located in the most privileged neighborhoods, the coefficients associated with school performance indexes in the housing price equation drop dramatically when we control for detailed neighborhood characteristics at the district level. In column 3, a standard deviation increase in the school performance index is now associated with a 2.6% (*DNB SCORE*), 2.2% (*GENERAL CURRICULUM*) and a 3.6% (*PRIVILEGED BACKGROUND*) increase in housing prices.

Although the set of controls included in the right hand side of the hedonic regression equation is large, we cannot ensure that it is adequate to solve the endogeneity problem. To isolate precisely the causal impact of school performance on housing prices, we need to restrict our sample to sales located in the vicinity of a school

Table 3.5: *Regressions results - Naive estimates of the impact of middle school performance on housing prices. School years 1997-2004.*

Dependent variable: log of price per square meter (in 2004 euros)						
Distance to school attendance boundary:	All sales (1)	All sales (2)	All sales (3)	< 250 meters (4)	< 300 meters (5)	< 350 meters (6)
<u>MIDDLE SCHOOL PERFORMANCE INDEX:</u>						
DNB SCORE (2004)	0.218***	0.194***	0.026***	0.031***	0.034***	0.033***
Cluster robust standard error	(0.017)	(0.016)	(0.008)	(0.009)	(0.009)	(0.009)
GENERAL CURRICULUM (1997-2004)	0.161***	0.142***	0.022***	0.024***	0.026***	0.026***
Cluster robust standard error	(0.015)	(0.014)	(0.005)	(0.005)	(0.005)	(0.005)
PRIVILEGED BACKGROUND (1997-2004)	0.190***	0.173***	0.036***	0.037***	0.039***	0.038***
Cluster robust standard error	(0.013)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)
<u>CONTROL VARIABLES:</u>						
School year & quarter fixed effects	YES	YES	YES	YES	YES	YES
Flat features	NO	YES	YES	YES	YES	YES
Census socio-demographic variables	NO	NO	YES	YES	YES	YES
Number of sales	195,845	195,845	195,845	99,917	113,531	124,608
Number of clusters (middle schools)	108	108	108	105	105	105

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). Flat features include a set of dummy variables for the age of the building (before 1850, 1850-1913, 1914-1947, 1948-1969, 1970-1980, after 1981), the number of bathrooms (1, 2 or more), the presence of a parking space, of a maid's room, the floor (first to fourth or more with and without a lift), the number of rooms (from 1 to 5 or more) and the average room size (small, medium, large). Socio-demographic neighborhood characteristics are taken from the 1999 French Census and are available at the district level. They include the following variables: average number of persons per flat, % public housing apartments, % owners, % families (*i.e.* household with at least one child), % single-parent families, % foreign, % unemployed and occupation of the household head (manager, self-employed, intermediary occupation, employee, manual worker, retired and economically inactive). Standard errors are clustered at the public middle school attendance level.

attendance boundary.

3.5.2 Sales located near school attendance boundaries

The implementation of our matching strategy requires that the selected sales are located close enough to a public middle school attendance boundary.

To ensure that our results are not too sensitive to a particular value of the chosen bandwidth (*i.e.* the maximum distance to the school attendance boundary), we decided to use three different samples of sales that differ in the maximum admitted distance between a transaction and its closest school attendance boundary. The selected values for the maximum distance to boundary are 250, 300 and 350 meters, to ensure enough statistical power to our estimation strategy and to provide sufficiently large intervals (50 meters) for testing the stability of our results.

Columns 3 to 8 of table 3.3 show how the features of the flats included in our three samples compare with those of the full sample of sales. The comparison between their neighborhoods' socio-demographic characteristics is performed in columns 3 to 8 of table 3.4. Reassuringly, these tables indicate that along both these dimensions, the sales included in our sample are hardly different from those of the full sample.

The last three columns of table 3.5 show that the naive estimates obtained when running the simple hedonic regression (controlling for time trends, flat features and neighborhood characteristics) on the subsample of sales located within 250 meters (column 4), 300 meters (column 5) and 350 meters (column 6) are in the same order of magnitude as the full sample estimates (column 4), even if they appear slightly larger. We can therefore be reasonably confident that the conclusions that can be drawn from our estimations are not specific to the sample of sales located in the vicinity of a school attendance boundary, but apply more generally to the entire sample of Parisian sales.

3.5.3 Matching sales across school attendance boundaries

A way of solving the endogeneity problem of the school performance variable in the housing price equation is to compare sales that can be assumed to share the

Table 3.6: *Regression of cross-boundary housing price differentials on corresponding school performance differentials. Sales are matched across public middle school attendance boundaries. School years 1997-2004.*

Dependent variable: Cross-boundary differential in the log of housing price (in 2004 euros)			
Distance to school attendance boundary:	< 250 meters	< 300 meters	< 350 meters
	(1)	(2)	(3)
<u>MIDDLE SCHOOL PERFORMANCE INDEX:</u>			
Δ DNB SCORE (2004)	0.015**	0.015**	0.015**
Cluster robust standard error	(0.007)	(0.007)	(0.007)
Δ GENERAL CURRICULUM (1997-2004)	0.019***	0.021***	0.021***
Cluster robust standard error	(0.005)	(0.005)	(0.005)
Δ PRIVILEGED BACKGROUND (1997-2004)	0.024***	0.026***	0.027***
Cluster robust standard error	(0.008)	(0.008)	(0.008)
Average Δ DNB SCORE	0.60	0.60	0.61
(s.d.)	(0.48)	(0.48)	(0.48)
Average Δ GENERAL CURRICULUM	0.86	0.86	0.86
(s.d.)	(0.64)	(0.63)	(0.63)
Average Δ PRIVILEGED BACKGROUND	0.72	0.72	0.72
(s.d.)	(0.56)	(0.56)	(0.56)
Average distance between matched sales (in meters)	155	182	207
(s.d.)	(54)	(65)	(75)
Number of sales	99,917	113,531	124,608
Number of clusters (school attendance boundaries)	169	171	172

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. To account for the fact that the price premium attached to specific flat features may vary throughout the city of Paris, we correct the prices of all sales for zone-specific flat features effects. The prices of all sales within a given zone (out of 18) are homogenized in the sense that there are now expressed in terms of the typical flat's characteristics (The typical flat has two rooms, belongs to a building constructed between 1850 and 1913, is located on the ground floor, with one bathroom and average size rooms, without a maid's room or a parking space and was sold during the fourth quarter). Appendix A explains the exact steps that we have followed to compute these so-called "hedonic" prices. The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). Details on the matching of sales across public middle school attendance boundaries are given in section 3.2.4. The regressions are weighted by the inverse of the distance between each match's components (*i.e.* the reference sale and the constructed counterfactual sale). The average distance between matched sales is the average distance between the reference sale and its constructed counterfactual. The geographic coordinates of the counterfactual are given by the average coordinates of the sales that are used to construct it, weighted by the inverse of their distance to the reference sale. Standard errors are clustered at the public middle school attendance boundary level.

same unobserved characteristics corresponding to “neighborhood effects”. We do so by applying the matching strategy that was presented in section 3.2.4.

The three columns of table 3.6 display the results of regressing cross-boundary housing price differentials on the corresponding school performance differentials, calculated using the matching approach described above for each of our three indexes of school performance. The average school performance differentials are reported at the bottom of the table. In line with the pattern observed in figures 3.6 to 3.8, we see that public middle school attendance boundaries induce substantial discontinuous variations in school performance. The average cross-boundary performance differential is equal to 0.60 of a standard deviation (of schools’ averages) for the *DNB* score, 0.86 for the fraction of pupils admitted into general curriculum high schools and 0.72 for the proportion of pupils coming from privileged socio-economic backgrounds.

The estimates obtained using the *DNB* exam score as a measure of school performance are significant at the 5% level. Moreover, they are remarkably stable across the different choices of maximum distance to boundary. The results displayed in the three columns show that a standard deviation increase in the average exam score raises the price per square meter by about 1.5%, which is roughly half the size of the naive estimate when we controlled for flat features and neighborhood socio-demographic characteristics (see column 3 of table 3.5).

The estimates obtained using the other two indexes of school performance are fairly similar across three different values chosen for the maximum distance to school attendance boundary and are larger in magnitude than the coefficient on the *DNB* exam score. A standard deviation increase in the fraction of middle school pupils who are admitted into general curriculum high schools is estimated to raise housing prices by about 2%, whereas the coefficient on the proportion of pupils coming from privileged socio-economic backgrounds is close to 2.5%. Hence, our results appear robust to different definitions of school performance and are in the range of 1.5 to 2.5%

3.5.4 How large is the effect?

Our estimates of the impact of middle school performance on housing prices in Paris are of the same order of magnitude as existing estimates for primary schools in other countries: similar to the 2.1% effect for every standard deviation increase in school performance for Boston suburbs (Black, 1999); slightly smaller than the 3.7% effect obtained by Gibbons and Machin (2003) for the UK.

In order to interpret these results, it is worth getting some sense of the size of the measured effect. First, we calculate that other things being equal, moving from the worst to the best public middle school (which corresponds to 4.8 times the standard error of the average *DNB* exam score at the school level) would imply a price premium of 7.2% (13,180 euros for the average flat price). Second, we estimate the fraction of the housing price differential between school zones that can be explained by differences in school performance. In order to do so, we calculate the observed difference in the average flat price (taken in logs) between each adjacent pair of school zones and relate it to the flat price differential predicted by the corresponding difference in school performance. We find that the difference in school performance explains roughly 5% of the observed difference in housing prices between adjacent school zones.

These calculations indicate that although school performance plays a non-negligible role in the formation of housing prices, it is not the main driving force in the real estate market. However, our results do not imply that the way school performance determines parents' residential location should be neglected when school zone policies are designed. What is estimated is indeed an average effect of school performance on housing prices over the entire population of households, which is probably much lower than the effect for the subpopulation of parents living with school-age children.

3.6 Robustness checks

In this section, we perform several robustness checks. We start by comparing our results with those obtained using Black (1999)'s the boundary fixed effects approach. In an attempt to assess the validity of the identifying assumption, we also perform two

series of tests: first, we ensure that flats located on either side of common school attendance boundaries share similar observable features; second, we address the issue of socio-demographic sorting across school attendance boundaries to determine whether observable patterns can be credibly explained by school performance differentials.

3.6.1 The boundary fixed effects alternative specification

To see how our results compare with those obtained using the boundary fixed effects approach that has been used since Black (1999), we perform a series of regressions in which we replace the vector of neighborhood characteristics in the basic hedonic equation (3.2) by a full set of school attendance boundary dummies. As pointed out earlier, such an estimation strategy relies on stronger assumptions than our matching approach. The violation of these assumptions would not bias the estimates, but could lead to wrong inference.

Table 3.7 shows the results obtained when the set of sales is restricted to a radius of respectively 250, 300 and 350 meters from the closest school boundary, for each of our three indexes of school performances. The standard errors in parentheses are clustered at the school attendance boundary level (as in the case of our matching estimator) whereas those in brackets are clustered at the school attendance level (as in most papers that apply the boundary fixed effects method⁴²). Reassuringly, the estimates are very similar in magnitude to our previous matching estimates: the estimated impact on housing prices of a standard deviation increase in the *DNB* score is slightly larger when we use boundary fixed effects instead of matching (2.0% *versus* 1.5% for a distance of 300 meters) but the coefficients are almost identical for the fraction of pupils entering general curriculum high schools (2.3% *versus* 2.1% at 300 meters) and for the proportion of pupils coming from privileged socio-economic backgrounds (2.9% *versus* 2.6% at 300 meters).

The level of significance reached by the boundary fixed effects estimator is always higher than the corresponding levels reached by the matching estimator. This phenomenon could be interpreted as a symptom that the assumptions underlying the boundary fixed effects approach might be too restrictive in the Parisian context,

⁴²On the potential pitfalls of school-level clustering, see section 3.2.3.

Table 3.7: Estimation of the impact of school performance on housing prices including the full set of school attendance boundaries dummies. School years 1997-2004.

Dependent variable: log of price per square meter (in 2004 euros)			
Distance to school attendance boundary:	< 250 meters (1)	< 300 meters (2)	< 350 meters (3)
<u>MIDDLE SCHOOL PERFORMANCE INDEX:</u>			
DNB SCORE (2004)	0.019***	0.020***	0.021***
Clustered standard error (at the boundary level)	(0.004)	(0.005)	(0.004)
Clustered standard error (at the school level)	[0.006]	[0.006]	[0.006]
GENERAL CURRICULUM (1997-2004)	0.022***	0.023***	0.023***
Clustered standard error (at the boundary level)	(0.004)	(0.004)	(0.004)
Clustered standard error (at the school level)	[0.005]	[0.005]	[0.005]
PRIVILEGED BACKGROUND (1997-2004)	0.028**	0.029**	0.030**
Clustered standard error (at the boundary level)	(0.006)	(0.006)	(0.006)
Clustered standard error (at the school level)	[0.007]	[0.007]	[0.007]
<u>CONTROL VARIABLES:</u>			
School year & quarter fixed effects	YES	YES	YES
School attendance boundaries fixed effects	YES	YES	YES
Average distance to boundary (in meters) (s.d.)	131 (63)	148 (78)	163 (88)
Number of sales	99,917	113,531	124,608
Number of clusters (middle schools)	105	105	105
Number of clusters (school attendance boundaries)	169	171	172

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level.

The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). The regressions are weighted by the inverse of the distance between a sale and its closest school attendance boundary. Standard errors in parentheses are clustered at the school attendance boundary level. The standard errors obtained when clustering at the middle school level are displayed in brackets.

leading to an underestimation of the true standard errors. In particular, clustering the standard errors at the school level rather than at the school attendance boundary level seems inappropriate since using the former instead of the latter would artificially inflate the significance of the estimates by about 20%⁴³.

3.6.2 Testing the validity of the identifying assumption

Our estimation strategy relies on the assumption that on average, sales located on either side of a common attendance boundary share the same flat features and neighborhood characteristics so that price differentials are purely attributable to school performance differentials. Yet this hypothesis might be violated if apartments located on the “good” side of school attendance boundaries tend to display certain features (*e.g.* more housing units with a parking space) that are valued by buyers independently from school performance. If such characteristics tend to attract wealthier households, whose children’s educational attainment will mechanically drive up the performance of the local school, then one might worry that our approach could produce upwardly biased estimates.

First, we test whether the observable characteristics of flats such as the age of building, the number and size of rooms, the number of bathrooms, the presence of a parking space or a maid’s room, are similarly distributed on either sides of common attendance boundaries. Table 3.8 compares the features of apartments located on either side of common school attendance boundaries, within a distance of 250, 300 and 350 meters to these boundaries. The “good” side of a particular boundary corresponds to where the school’s average *DNB* exam score is the largest and the “bad” side where it is the smallest. Within each sample considered separately, the characteristics on the “good” side and the “bad” side of boundaries appear almost exactly similar. Moreover, the numbers show no obvious pattern that could explain that flats located on the “good” side of the boundary are more desirable on average than flats located on the “bad” side. In addition to these comparisons, we performed our matching

⁴³Note that the underestimation of the true standard errors induced by clustering at the school rather than at the boundary level is probably less severe in our analysis than in other empirical contexts. This is because we have excluded all school attendance boundaries that coincide with *arrondissements* boundaries, which explains that the number of school boundary clusters in our regressions is not much larger than the number of school attendance area clusters.

Table 3.8: *Comparison of flat features on the “good” and “bad” side of school attendance boundaries (in terms of the school’s average DNB exam score in 2004). School years 1997-2004.*

Distance to boundary	< 250 meters		< 300 meters		< 350 meters	
	“Bad” side	“Good” side	“Bad” side	“Good” side	“Bad” side	“Good” side
Flat size (in m ²)	51	52	51	52	51	52
(s.d.)	(34)	(34)	(34)	(34)	(34)	(35)
Price per m ² (in 2004)	3,305	3,338	3,306	3,346	3,313	3,350
(s.d.)	(1,247)	(1,236)	(1,262)	(1,246)	(1,271)	(1,253)
FLAT CHARACTERISTICS						
Age of building (percent)						
Unknown	0.11	0.11	0.11	0.11	0.11	0.11
Before 1850	0.05	0.05	0.05	0.05	0.05	0.05
1850-1913	0.42	0.41	0.42	0.40	0.42	0.40
1914-1947	0.16	0.17	0.16	0.17	0.16	0.16
1948-1969	0.13	0.13	0.13	0.13	0.13	0.13
1970-1980	0.11	0.11	0.11	0.11	0.11	0.11
After 1981	0.02	0.03	0.02	0.03	0.02	0.03
Floor (percent)						
Ground floor	0.10	0.11	0.10	0.11	0.10	0.11
First	0.16	0.16	0.16	0.16	0.16	0.16
Second	0.16	0.16	0.16	0.16	0.16	0.16
Third	0.15	0.16	0.15	0.16	0.15	0.16
Fourth	0.15	0.14	0.15	0.14	0.15	0.14
Fifth	0.12	0.12	0.12	0.12	0.12	0.12
Sixth or more	0.15	0.15	0.15	0.15	0.15	0.15
Number of rooms (percent)						
One	0.24	0.24	0.24	0.24	0.24	0.24
Two	0.37	0.36	0.37	0.36	0.37	0.36
Three	0.22	0.23	0.22	0.23	0.22	0.23
Four	0.10	0.10	0.10	0.10	0.10	0.10
Five or more	0.07	0.07	0.07	0.07	0.07	0.07
% without ind. bathroom	0.21	0.21	0.21	0.21	0.21	0.21
% with maid’s room	0.04	0.04	0.04	0.04	0.04	0.04
% with lift	0.90	0.90	0.90	0.90	0.90	0.90
% with garage	0.12	0.12	0.12	0.13	0.12	0.13
SCHOOL PERFORMANCE						
Average DNB exam score	9.11	9.95	9.11	9.94	9.11	9.94
(s.d.)	(1.39)	(1.44)	(1.39)	(1.43)	(1.39)	(1.43)
% entering general high schools	0.64	0.70	0.64	0.70	0.64	0.70
(s.d.)	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)	(0.12)
% with privileged background	0.41	0.49	0.41	0.49	0.41	0.49
(s.d.)	(0.19)	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)
Number of sales	53,181	46,736	60,672	52,859	67,014	57,594

Notes: The features of flats sold during school years 1997 to 2004 are computed from the *BIEN* (*Base d’Informations Économiques Notariales*) dataset, which is managed by the Notary Chamber of Paris and the Île-de-France. The sample is restricted to all arm’s-length sales of second hand Parisian flats that took place between September 1997 and August 2004. Sales located on the Île Saint-Louis and Île de la Cité islands were excluded from the sample because of their very specific location and pattern of housing prices. The average 2004 *Diplôme national du Brevet* (*DNB*) score at the school level (*DNB SCORE*) is computed from the *Organisation des Concours et Examens Académiques et Nationaux* (*OCEAN*) national examinations database. The proportion of middle school pupils entering general curriculum high schools (*GENERAL CURRICULUM*) and the fraction of pupils coming from privileged socio-economic backgrounds (*PRIVILEGED BACKGROUND*), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). For each boundary separating two public middle schools, the “bad” side corresponds to the school with the lowest average *DNB* exam score whereas the “good” side corresponds to the school with the highest average *DNB* exam score.

regressions without controlling for flat features, and found the same results⁴⁴. Hence our findings do not seem to be driven by differences in observable flat features. One might still argue that some unobservable flat features (*e.g.* one side gets more sun than the other) might bias the results, but there is no particular reason to believe that such characteristics should be distributed differently across school zones.

A more serious issue is that household sorting might occur at boundaries, even if apartments share the same features on average. Several papers using US data (Bayer et al., 2005, 2003; Kane et al., 2006) have shown that not only school performance, but also several socio-demographic characteristics (such as household income) could vary discontinuously between adjacent school catchment areas. In this case, comparing sales across boundaries would lead to overestimate the causal effect of school performance on housing prices.

To investigate the issue of sorting across attendance boundaries, we compare the observable socio-demographic characteristics of neighborhoods (at the district level) located on the “good” *versus* the “bad” side of school boundaries. Table 3.9 shows that within each of our three samples (250, 300 and 350 meters), neighborhoods socio-demographic characteristics are remarkably similar on both sides of boundaries and that households living on the “good” side of a boundary do not appear more privileged on average than households living on the “bad” side. This simple exercise does not support the view that in the case of Paris, school attendance boundaries mirror other preexisting lines of separation between distinct neighborhoods. Given the very irregular trace of school catchment areas (see figure 3.2), it seems rather plausible to consider that in our particular empirical context, school boundaries run through otherwise similar neighborhoods.

If large and general sorting effects across school attendance boundaries would presumably represent a threat to our identification strategy, one would still expect school boundaries to generate a limited amount of sorting among households that are likely to take school performance into account when making their residential choice. Most existing models of school-related sorting (reviewed by Fernández, 2003) argue that under the single-crossing condition, wealthier parents have a stronger incentive

⁴⁴The results are not displayed since they are basically the same as the initial regressions. The table is available upon request.

Table 3.9: *Comparison of socio-demographic characteristics of neighborhoods located on the “good” and “bad” side of school attendance boundaries (in terms of the school’s average DNB exam score in 2004). School years 1997-2004.*

Distance to boundary	< 250 meters		< 300 meters		< 350 meters	
	“Bad” side	“Good” side	“Bad” side	“Good” side	“Bad” side	“Good” side
<u>CENSUS VARIABLES</u>						
Nb of census districts	554	548	584	570	602	584
Nb of individuals per district	2,384	2,343	2,389	2,339	2,376	2,341
Nb of households per district	1,271	1,241	1,270	1,235	1,263	1,236
Nb of persons per flat	1.86	1.88	1.86	1.88	1.86	1.88
<u>All households</u>						
% families with children under 25	0.22	0.22	0.22	0.22	0.22	0.22
% female-headed households	0.06	0.06	0.06	0.06	0.06	0.06
% owners	0.32	0.32	0.31	0.31	0.31	0.31
% public housing	0.12	0.12	0.13	0.12	0.13	0.12
% with graduate degree	0.40	0.41	0.40	0.41	0.40	0.41
% foreigners	0.22	0.22	0.22	0.22	0.22	0.22
% unemployed	0.12	0.11	0.12	0.11	0.12	0.11
<u>Occupation:</u>						
% managers	0.23	0.23	0.23	0.23	0.23	0.23
% self-employed workers	0.04	0.04	0.04	0.04	0.04	0.04
% intermediary occupation	0.14	0.14	0.14	0.14	0.15	0.14
% employees	0.14	0.14	0.14	0.14	0.14	0.14
% manual workers	0.06	0.06	0.06	0.06	0.06	0.06
% retired	0.18	0.17	0.18	0.17	0.18	0.17
% economically inactive	0.21	0.21	0.21	0.21	0.21	0.21
<u>SCHOOL PERFORMANCE</u>						
Average DNB exam (2004)	9.11	9.95	9.11	9.94	9.11	9.94
(s.d.)	(1.39)	(1.44)	(1.39)	(1.43)	(1.39)	(1.43)
% entering general high schools	0.64	0.70	0.64	0.70	0.64	0.70
(s.d.)	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)	(0.12)
% with privileged backgrounds	0.41	0.49	0.41	0.49	0.41	0.49
(s.d.)	(0.19)	(0.20)	(0.19)	(0.20)	(0.19)	(0.20)
Number of sales	53,181	46,736	60,672	52,859	67,014	57,594

Notes: Socio-demographic neighborhood characteristics are taken from the 1999 French Census and are available at the district level. They include the following variables: average number of persons per flat, % public housing apartments, % owners, % families (*i.e.* household with at least one child), % single-parent families, % foreign, % unemployed and occupation of the household head (manager, self-employed, intermediary occupation, employee, manual worker, retired and economically inactive). The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). For each boundary separating two public middle schools, the “bad” side corresponds to the school with the lowest average DNB exam score whereas the “good” side corresponds to the school with the highest average DNB exam score.

to sort into neighborhoods assigned to high-performing schools than poorer parents. School attendance boundaries could therefore induce some sorting within the sub-population of families. We examined this question by looking separately at families (*i.e.* households with at least one child under the age of 25) and households with no children at home. The only dimension along which these two groups of households can be compared in the Census is the occupation of the household head. In order to detect eventual school-related sorting across attendance boundaries, we restricted our sample of sales to those located near boundaries with school performance differentials larger than half a standard deviation (using the average *DNB* exam score index). Within each of our three original samples, we were left with roughly half the number of observations.

Table 3.10 compares the distribution of household head occupations on both sides of school attendance boundaries, separately for families and for households with no children at home. While the patterns are identical when we consider households without children, they are somewhat different for families: the average fraction of managers on the “good” side of the boundary is 42% *versus* 39% on the “bad” side and the fraction of manual workers is 11% on the “bad side” *versus* 13% on the “good” side. Although these differences are not very large, they are significant and consistent with the claim that if sorting occurs at school attendance boundaries, it is entirely driven by families. We are therefore confident that our estimation strategy captures the effect of school performance on housing prices and that socio-demographic differences across boundaries are solely caused by school performance differentials.

So far, we have established that parents care about public school performance when they make their residential choices. In the next section, we investigate whether the availability of private schools influences the housing price premium attached to public school performance.

Table 3.10: *Comparison of the occupation of the household head (computed separately for families and households without children) in neighborhoods located on the “good” and “bad” side of school attendance boundaries where the school performance differential is larger than half a standard deviation (in terms of the school’s average DNB exam score in 2004). School years 1997-2004.*

Distance to boundary	< 250 meters		< 300 meters		< 350 meters	
	“Bad” side	“Good” side	“Bad” side	“Good” side	“Bad” side	“Good” side
<u>CENSUS VARIABLES</u>						
% families with children under 25	0.22	0.22	0.22	0.22	0.22	0.22
<u>Families only</u>						
Occupation:						
% managers	0.39	0.42	0.39	0.42	0.39	0.41
% self-employed workers	0.10	0.10	0.10	0.10	0.10	0.10
% intermediary occupation	0.16	0.15	0.16	0.16	0.16	0.16
% employees	0.15	0.14	0.14	0.14	0.15	0.14
% manual workers	0.13	0.11	0.13	0.11	0.13	0.11
% retired	0.03	0.03	0.03	0.03	0.03	0.03
% economically inactive	0.04	0.04	0.04	0.04	0.04	0.04
<u>Households with no children at home</u>						
Occupation:						
% managers	0.19	0.19	0.19	0.19	0.19	0.19
% self-employed workers	0.03	0.03	0.03	0.03	0.03	0.03
% intermediary occupation	0.11	0.11	0.11	0.11	0.11	0.11
% employees	0.10	0.10	0.10	0.10	0.10	0.10
% manual workers	0.04	0.04	0.05	0.04	0.05	0.04
% retired	0.21	0.21	0.21	0.21	0.21	0.21
% economically inactive	0.10	0.10	0.10	0.10	0.10	0.10
<u>SCHOOL PERFORMANCE</u>						
Average DNB exam (2004)	8.77	10.31	8.77	10.25	8.77	10.25
(s.d.)	(1.36)	(1.48)	(1.36)	(1.51)	(1.36)	(1.51)
% entering general high schools	0.63	0.72	0.63	0.72	0.63	0.72
(s.d.)	(0.10)	(0.13)	(0.10)	(0.13)	(0.10)	(0.13)
% with privileged backgrounds	0.38	0.54	0.38	0.53	0.38	0.53
(s.d.)	(0.17)	(0.21)	(0.17)	(0.21)	(0.17)	(0.21)
Number of Census districts	332	310	354	332	367	344
Number of sales	27,830	22,422	32,082	25,603	35,983	28,224

Notes: In the 1999 French Census, the occupation of the household head is available at the district level separately for families (i.e. households with a child under the age of 25) and for households with no children at home. The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). For each boundary separating two public middle schools, the “bad” side corresponds to the school with the lowest average DNB exam score whereas the “good” side corresponds to the school with the highest average DNB exam score.

3.7 The mitigating effect of private schools

Previous studies have mainly focused on public schools, ignoring the other options provided by private schools in their analysis⁴⁵. However, recent papers on school vouchers (Nechyba, 2003; Ferreyra, 2007) argue that private schools constitute an outside option that parents take into account when deciding in which area to buy. In this section, we exploit the combination of strict public school residence-based assignment and private school choice in the Parisian context to test whether private school availability influences residential choices. Indeed, the presence of a good network of private schools in some neighborhoods should lower the price premium that parents are ready to pay for a flat located in a good public school zone, or at least set an upper bound for that premium.

3.7.1 Theoretical predictions of school choice models

The way housing markets react to school performance when public and private schools coexist has been extensively analyzed by Nechyba in a series of theoretical papers (Nechyba, 1999, 2000, 2003). A number of testable predictions can be derived from his general equilibrium models of school finance, which include multiple school districts (either state or locally financed), multiple neighborhoods within school districts and different housing qualities. Local public schools are subject to zoning and coexist with private schools that can freely select their pupils. Using this particular set-up, Nechyba studies the consequences of private school choice on both school and residential stratification. The model is too complex to yield analytic solutions and it is calibrated to predict the effect of alternative school assignment policies. In the benchmark equilibrium in which no private school operates, peer effects yields substantial residential sorting⁴⁶. Private school choice decreases residential stratification and increases peer stratification in schools⁴⁷. The decrease in residential stratification is reflected in housing prices. In this section we test whether, in line with these

⁴⁵Gibbons and Machin (2003) include religious schools in their analysis, since they are publicly financed, but do not investigate whether these schools have a specific impact and do not include other private schools in their analysis.

⁴⁶There is only incomplete stratification because of exogenous heterogeneity in housing quality.

⁴⁷The overall effect on the performance of public schools varies with the type of school finance and on whether per public spending increases enough to compensate for the decrease in peer quality.

predictions, the presence of private schools in the neighborhood mitigates the effect of local public school performance on housing prices.

As explained in section 3.3, private schools offer an attractive outside option to parents who wish to avoid sending their children to a poor-performing local public school. In Paris, about a third of middle school pupils were enrolled in a private school during the period under study (see table 3.1). Contrary to public middle schools, the spatial distribution of private institutions is not even throughout the city. Therefore, under the assumption that parents don't usually want to enroll their children in a school that is located too far away from home, the impact of public school performance on housing prices should depend on the availability of local private schools in the neighborhood. The assumption that distance to school matters to parents who consider sending their children to a private institution can be indirectly tested by looking at how far away private middle school pupils reside from their school. Although we cannot calculate this distance precisely because we lack information on the home address of pupils, we know in which of the 20 Parisian *arrondissements* they reside. Table 3.11 shows that among Parisian pupils enrolled in a private school in 2003-2004, 53% attended a school located in their *arrondissement* of residence and 28% a school located in an adjacent *arrondissement*⁴⁸, so 81% of private school pupils can be considered as living reasonably close to their school. Other things being equal, the availability of private schools in a given area should therefore raise the probability of parents enrolling their children in the private sector and should lower the capitalization of public school performance in local housing prices.

In terms of our housing price model, the impact of private school availability on the capitalization of public school performance in housing prices can be viewed as changing the magnitude of parameter β which we now consider as a function of

⁴⁸Because of school zoning, the corresponding percentages for public schools are logically higher. Two reasons can potentially explain that not all public middle school pupils are enrolled in a school located in their *arrondissement* of residence. First, as explained in section 3.3, a number of pupils have been granted a dispensation by the LEA or cheat on their address. Second, some public middle school catchment areas overlap several *arrondissements*.

Table 3.11: *Location of the school attended by public and private school Parisian pupils with respect to their arrondissement of residence. School year 2003-2004.*

Type of middle school attended	Location of attended middle school with respect to the arrondissement of residence	Percentage
PUBLIC	Same <i>arrondissement</i>	85%
	Adjacent <i>arrondissements</i>	8%
	Other <i>arrondissements</i> in Paris	6%
	Outside Paris (Île-de-France)	1%
PRIVATE	Same <i>arrondissement</i>	53%
	Adjacent <i>arrondissements</i>	28%
	Other <i>arrondissements</i> in Paris	13%
	Outside Paris (Île-de-France)	6%

Notes: Source: The location of the school attended by Parisian public and private school pupils with respect to their arrondissement of residence are computed from the *SCOLARITE* pupils dataset covering school year 2003-2004. Statistics are computed for all middle school students residing in Paris and enrolled in a middle school in Paris or elsewhere in the Île-de-France. The city of Paris is divided into 20 *arrondissements*. Among middle school students attending a private school, 53% attend a school located in their *arrondissement* of residence and 28% in an adjacent *arrondissement*.

private school availability in the neighborhood (denoted a_b):

$$\ln p_{i,n,s,t} = \alpha + \beta(a_b) \cdot z_s + X_{i,n,s,t} \gamma + \theta_{n,t} + \epsilon_{i,n,s,t} \quad (3.6)$$

Under the assumption that both sales of a common match are sufficiently close to enjoy the same private school availability, the housing price differential equation can be rewritten:

$$\ln p_{i,b,s,t} - \ln p_{i',b,s',t} = \beta(a_b)(z_s - z_{s'}) + \mu_{i,i',b,s,s',t}$$

where $\mu_{i,i',b,s,s',t}$ is an error term which is clustered at the public school attendance boundary level⁴⁹.

With respect to the previous model, the impact of public school performance on housing prices is no longer spatially homogenous but is allowed to vary with private school availability. Within this framework, the estimated parameter $\hat{\beta}$ is the average impact of school performance on housing prices within the area covered by the sample. If private schools tend to mitigate the impact of public school performance on housing prices, we would expect the estimated coefficient to vary with the level of private school availability.

In addition, the analysis of the mitigating effect of private schools provides us with an indirect way of testing the validity of our identification strategy. If our estimation of the price premium paid for flats located on the “good” side of a boundary was solely due to unobservable differences in neighborhood characteristics and was not driven by school quality, this premium should remain significant in all areas, irrespective of the density of private schools in the neighborhood. On the contrary, if we find that this premium is lowered by the proximity to private schools, we would have good reasons to believe that the price premium is not driven by unobservable differences across schools boundaries.

⁴⁹In terms of our previous notation, $\mu_{i,i',b,s,s',t} = \epsilon_{i,b,s,t} - \epsilon_{i',b,s',t}$.

3.7.2 Estimation strategy

To test the hypothesis that private school availability influences the capitalization of public school quality in housing prices, we construct two different indexes of private school availability in a particular location. Our first index is aimed at capturing local private school proximity while the second is designed to measure local private school density.

For each transaction i , our index of local private school proximity (denoted PSP_i) is computed as the inverse of the distance between a sale and its closest private middle school:

$$PSP_i = \frac{1}{\min_j(d_{i,j})} \quad \forall i$$

where $d_{i,j}$ denotes the distance between transaction i and private school j . The higher the value of PSP_i , the closer the private middle school is to transaction i .

Our samples of sales located in the vicinity of a school attendance boundary are then splitted into four groups of equal size. Each sale is allocated to one of the four quartiles depending on the value of its private school proximity index⁵⁰. Figure 3.10 shows how corresponding areas are distributed in the city of Paris for the full set of existing Parisian addresses⁵¹. Reassuringly, our index of private school proximity is well distributed among neighborhoods and does not cut Paris into four geographically distinct zones, a feature that could bias our estimates.

Although simple and intuitive, this index of private school proximity may not be full adequate to reflect the availability of private schools in a particular area. Since the value of this index depends only on the distance between a sale and its closest private school, it does not take into account the local density of the private school network. The presence of a private institution in a given neighborhood does not in itself guarantee to prospective owners that they will be able to enroll their children in this particular private school, especially if it is oversubscribed. Hence it is reasonable to assume that the outside option provided by the private sector

⁵⁰Note that choosing the inverse of the distance between the transaction and the private school rather than a power of that distance (*e.g.* distance squared) has no incidence on the distribution of sales in each of the four quartiles, since it leaves the rank ordering of sales unchanged.

⁵¹Since sales were not necessarily completed in all Parisian addresses during the period under study, the full set of existing addresses in larger the set of sales included in the *BIEN* dataset.

Figure 3.10: *Map of local private middle school proximity in Paris.*

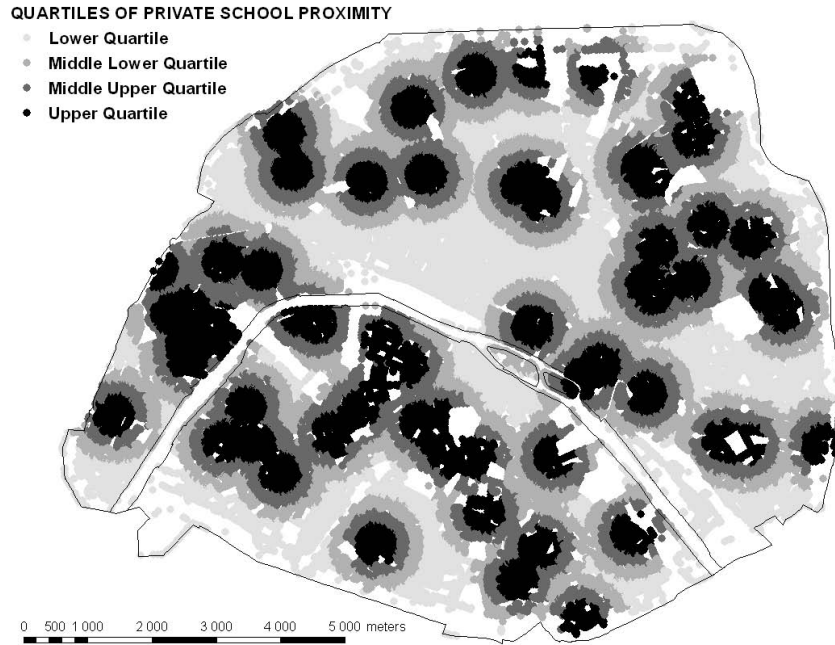
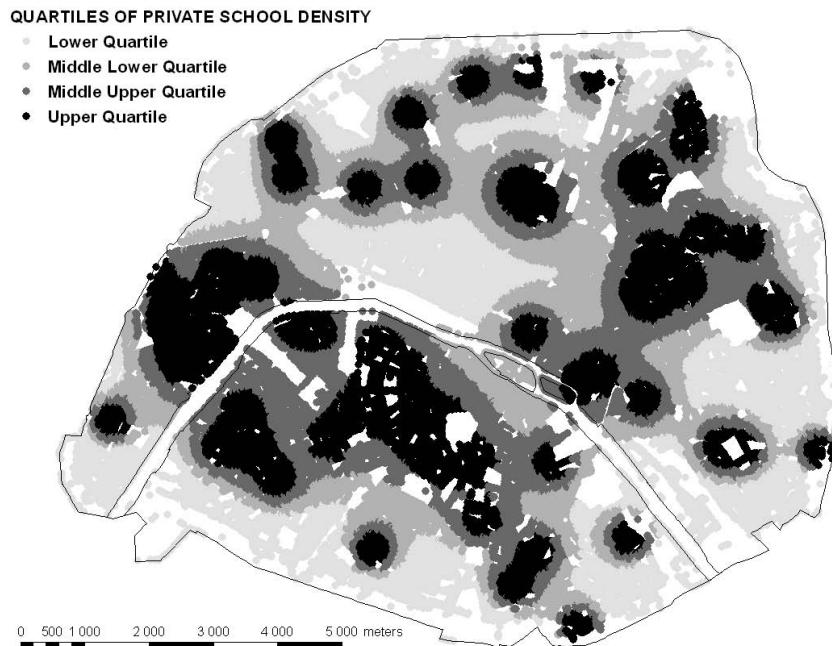


Figure 3.11: *Map of local private middle school density in Paris.*



will be more attractive in neighborhoods where a large number of schools operate than in neighborhoods where they are scarce. To take this dimension into account, we construct an index of local middle school density (denoted PSD_i). For each transaction i , we calculate the average value of the inverse of its distance (squared) to every private middle school in Paris:

$$PSD_i = \sum_{j=1}^{N_j} \frac{1}{d_{i,j}^2} \quad \forall i$$

where N_j denotes the total number of private middle schools that operate in the city of Paris (65) and $d_{i,j}$ is the distance between transaction i and private school j . The higher the value of this index, the higher the density of private schools in a transaction's neighborhood. Using the inverse of the distance squared allows us to give much more weight to the closest private schools relatively to the more distant ones. We then split our samples of sales into four groups depending on the value taken by the index for each particular sale. Figure 3.11 shows the corresponding spatial distribution of private school density. Logically, this map appears very similar to the map displaying the quartiles of local private school proximity (figure 3.10). However, some differences arise in certain neighborhoods: in particular, some areas which are classified as having a low level of private school proximity because the closest private school is far away are now classified as having a higher level of public school density because they are surrounded by many private schools.

To estimate the mitigating effect of private schools, we allow the index of public middle school performance to vary with the quartile of private middle school availability (proximity or density). The regression of the log price differential between sale i and its counterfactual i' in time t on the corresponding school performance differential thus takes the following form:

$$\log p_{i,b,s,t} - \log p_{i',b,s',t} = \beta_1 Q_1 \cdot \Delta_{z_s, z_{s'}} + \beta_2 Q_2 \cdot \Delta_{z_s, z_{s'}} + \beta_3 Q_3 \cdot \Delta_{z_s, z_{s'}} + \beta_4 Q_4 \cdot \Delta_{z_s, z_{s'}} + \mu_{i,i',b,s,s',t}$$

where $\Delta_{z_s, z_{s'}}$ is the school performance differential between sale i and sale i' 's assigned public schools s and s' , Q_j ($j \in \{1, 2, 3, 4\}$) are dummy variables that indicate to which quartile of private school availability the reference sale belongs and $\mu_{i,i',b,s,s',t}$ is the error term clustered at the school attendance boundary level.

3.7.3 Results

Table 3.12 shows the results obtained when we measure private school availability using the index of private school proximity. Column 1 displays the baseline coefficient on school performance in the housing price equation already presented in table 3.6. The coefficients on the interaction terms between school performance and each of the four quartiles of private school proximity are displayed in columns 2 to 5. Each panel corresponds to a particular choice of the maximum distance to public school attendance boundary: 250 meters (panel A), 300 meters (panel B) and 350 meters (panel C). Within each panel, the results are displayed separately for our three indexes of public middle school performance.

The coefficients show a very distinctive pattern and are remarkably consistent across the different choices of school performance indexes and maximum distance to boundary. Columns 2 to 5 indicate that as private school proximity increases, the impact of public school performance on housing prices becomes smaller. While for the bottom quartile of private school proximity, the coefficient on the *DNB* exam score is about 3.0%, it falls to about 2.0% for the middle lower quartile and becomes small and insignificant (0.3-0.8%) for the upper quartiles (columns 4 and 5).

The results are very similar when we use the other two indexes of school performance. A standard deviation increase in the fraction of public middle school pupils who are admitted into general curriculum high schools raises housing prices by about 3.0% in areas belonging to the bottom quartile of private school proximity and only 1.4% in areas belonging to the top quartile. The difference is even larger when the fraction of pupils coming from privileged socio-economic backgrounds is used to measure school performance: the coefficient falls from 4.2-4.7% to 0.8-1.4% when we move from the bottom to the top quartile of private school proximity.

As emphasized in the previous section, the degree to which private schools serve as an advantageous outside option to avoid the local public middle school could be better captured by our second measure of private school availability which uses the density rather than the proximity of private schools. In table 3.13, the PSD index replaces the PSP index. The results are very similar to those obtained previously but the contrast between the areas belonging to the lower and upper quartiles of

Table 3.12: *Regression of cross-boundary housing price differentials on corresponding school performance differentials by quartile of proximity to the closest private middle school. Sales are matched across public middle school attendance boundaries. School years 1997-2004.*

Dependent variable: Cross-boundary differential in the log of housing price (in 2004 euros)					
	All sales (1)	Quartiles of proximity to the closest private middle school			
		Lower Quartile (2)	Middle Lower Quartile (3)	Middle Upper Quartile (4)	Upper Quartile (5)
<u>PANEL A: SALES <250 METERS OF BOUNDARY</u>					
Δ DNB SCORE (2004)	0.015**	0.029*	0.019***	0.007	0.006
Cluster robust standard error	(0.007)	(0.017)	(0.007)	(0.008)	(0.011)
Δ GENERAL CURRICULUM (1997-2004)	0.019***	0.027***	0.022***	0.013**	0.010
Cluster robust standard error	(0.005)	(0.011)	(0.007)	(0.006)	(0.008)
Δ PRIVILEGED (1997-2004)	0.024***	0.042**	0.029***	0.014*	0.008
Cluster robust standard error	(0.008)	(0.018)	(0.009)	(0.008)	(0.009)
Number of observations	99,917	24,978	24,973	24,985	24,981
<u>PANEL B: SALES <300 METERS OF BOUNDARY</u>					
Δ DNB SCORE (2004)	0.015**	0.028*	0.020**	0.006	0.008
Cluster robust standard error	(0.007)	(0.017)	(0.008)	(0.008)	(0.010)
Δ GENERAL CURRICULUM (1997-2004)	0.021***	0.029***	0.026***	0.013***	0.014**
Cluster robust standard error	(0.005)	(0.011)	(0.007)	(0.005)	(0.007)
Δ PRIVILEGED (1997-2004)	0.026***	0.043**	0.032***	0.015**	0.012
Cluster robust standard error	(0.008)	(0.019)	(0.009)	(0.007)	(0.008)
Number of observations	113,531	28,377	28,388	28,380	28,386
<u>PANEL C: SALES <350 METERS OF BOUNDARY -</u>					
Δ DNB SCORE (2004)	0.015**	0.033**	0.022**	0.003	0.005
Cluster robust standard error	(0.007)	(0.015)	(0.009)	(0.008)	(0.010)
Δ GENERAL CURRICULUM (1997-2004)	0.021***	0.031***	0.028***	0.012**	0.014**
Cluster robust standard error	(0.005)	(0.011)	(0.007)	(0.005)	(0.007)
Δ PRIVILEGED (1997-2004)	0.027***	0.047***	0.033**	0.014**	0.014*
Cluster robust standard error	(0.008)	(0.018)	(0.009)	(0.007)	(0.008)
Number of observations	124,608	31,143	31,152	31,160	31,153

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). The indicator of the proximity of private middle schools is constructed as follows: for each “reference” sale, we calculate the inverse of its distance to the closest private middle school; we then split this indicator into four quartiles, from the most distant to the closest local private school. Details on the matching of sales across public middle school attendance boundaries are given in section 3.2.4. The regressions are weighted by the inverse of the distance between each match’s components (*i.e.* the reference sale and the constructed counterfactual sale). Standard errors are clustered at the public middle school attendance boundary level.

private school density is sharper: for sales located within 250 meters of a public school attendance boundary (panel A), the impact of a standard deviation increase in the local public school's average *DNB* score raises housing prices by 3.4% in the bottom quartile of private school density but is insignificant and close to zero in the top quartile (column 5). Similar conclusions can be drawn when choosing different values of the distance to boundary and alternative indexes of school performance.

On the whole, these results support the theoretical prediction that the impact public school performance on housing prices exhibits spatial heterogeneity. When parents have the opportunity to send their children to local private schools, then housing prices do not seem to depend on the performance of the local public middle school; on the contrary, when there are few private schools available in the neighborhood, then the local public middle school performance appears to be capitalized into housing prices.

These results are also interesting as they suggest that our previous estimates of the impact of school quality on housing prices are unlikely to be driven by differences in neighborhood characteristics that are unrelated to school quality. The finding that the housing price premium shrinks and becomes insignificant in areas with a dense network of private schools is rather reassuring, as it shows that the housing price differential disappears when we do not expect to find a price premium attached to public school quality.

Given these estimates, one can perform the simple exercise that consists in comparing the cost of attending a private school with the cost of moving into the catchment area of a better-performing public school. For an average private school tuition fee of 1,000 euros per year in Paris, four years of private middle schooling cost about 4,000 euros to parents. In areas belonging to the top quartile of private school density, the average *DNB* exam score is 9.59 for public schools and 10.85 for private schools (see table 3.15). This difference of 1.26 point is equal to two thirds of a standard deviation in the school average *DNB* exam score. According to our estimates, the housing price premium to be paid for a similar increase in public school performance in areas belonging to the bottom quartile of private school density would be equal to about 2.5%, which is about 4,500 euros at the average flat price. The housing

Table 3.13: *Regression of cross-boundary housing price differentials on corresponding school performance differentials by quartile of private school density. Sales are matched across public middle school attendance boundaries. School years 1997-2004.*

Dependent variable: Cross-boundary differential in the log of housing price (in 2004 euros)					
	All sales (1)	Quartiles of private middle school density			
		Lower Quartile (2)	Middle Lower Quartile (3)	Middle Upper Quartile (4)	Upper Quartile (5)
<u>PANEL A: SALES <250 METERS OF BOUNDARY</u>					
Δ DNB SCORE (2004)	0.015**	0.034**	0.015*	0.007	0.002
Cluster robust standard error	(0.007)	(0.016)	(0.008)	(0.007)	(0.010)
Δ GENERAL CURRICULUM (1997-2004)	0.015**	0.028***	0.028***	0.010**	0.008
Cluster robust standard error	(0.007)	(0.010)	(0.007)	(0.004)	(0.007)
Δ PRIVILEGED (1997-2004)	0.015**	0.043***	0.029***	0.010*	0.005
Cluster robust standard error	(0.007)	(0.017)	(0.008)	(0.006)	(0.009)
Number of observations	99,917	24,979	24,964	24,994	24,980
<u>PANEL B: SALES <300 METERS OF BOUNDARY</u>					
Δ DNB SCORE (2004)	0.019***	0.035**	0.015*	0.006	0.004
Cluster robust standard error	(0.005)	(0.015)	(0.009)	(0.008)	(0.010)
Δ GENERAL CURRICULUM (1997-2004)	0.021***	0.030***	0.029***	0.012**	0.011*
Cluster robust standard error	(0.005)	(0.011)	(0.007)	(0.004)	(0.006)
Δ PRIVILEGED (1997-2004)	0.021***	0.044***	0.031***	0.012**	0.010
Cluster robust standard error	(0.005)	(0.016)	(0.009)	(0.006)	(0.008)
Number of observations	113,531	28,380	28,385	28,381	28,385
<u>PANEL C: SALES <350 METERS OF BOUNDARY</u>					
Δ DNB SCORE (2004)	0.024***	0.038***	0.017*	0.005	-0.001
Cluster robust standard error	(0.008)	(0.013)	(0.009)	(0.008)	(0.010)
Δ GENERAL CURRICULUM (1997-2004)	0.026***	0.032***	0.031***	0.011**	0.011*
Cluster robust standard error	(0.008)	(0.010)	(0.007)	(0.004)	(0.006)
Δ PRIVILEGED (1997-2004)	0.027***	0.046***	0.034***	0.011*	0.011
Cluster robust standard error	(0.008)	(0.016)	(0.009)	(0.007)	(0.008)
Number of observations	124,608	31,152	31,146	31,156	31,154

Notes: *: significant at the 10% level; **: significant at the 5% level; ***: significant at the 1% level. The average 2004 *Diplôme national du Brevet (DNB)* score at the school level (DNB SCORE) is computed from the *Organisation des Concours et Examens Académiques et Nationaux (OCEAN)* national examinations database. The proportion of middle school pupils entering general curriculum high schools (GENERAL CURRICULUM) and the fraction of pupils coming from privileged socio-economic backgrounds (PRIVILEGED BACKGROUND), averaged over school years 1997-2004, are computed from the annual *SCOLARITE* dataset of French pupils (1997-2004). The indicator of the density of private middle schools is constructed as follows: for each “reference” sale, we calculate the average of the inverse of its distance to every private middle school in Paris squared; we then split our indicator into four quarters, from the lowest density of private schools to the highest. Details on the matching of sales across public middle school attendance boundaries are given in section 3.2.4. The regressions are weighted by the inverse of the distance between each match’s components (*i.e.* the reference sale and the constructed counterfactual sale). Standard errors are clustered at the public middle school attendance boundary level.

price premium attached to public school performance in neighborhoods where private schools are scarce is therefore in the same order of magnitude as the individual cost of a private school four-year tuition fee in areas where they are many. These figures seem to indicate that the valuation of public school performance is at least partially determined by the cost of the outside option provided by private education.

3.7.4 Robustness of findings to alternative interpretations

For our estimation strategy to credibly identify the mitigating effect of private schools on the capitalization of public school performance in housing prices, we need to show that our results are not driven by other confounding factors.

The first concern that needs to be addressed is that the availability of private schools could be correlated with a number of socio-demographic characteristics that tend to lower the taste for public school performance in a given area. For instance, if private schools are on average concentrated in poorer areas, the apparent insignificant impact of public school performance on housing prices could derive from the weaker willingness of economically disadvantaged households to pay for better schools. To investigate this issue, we use the 1999 Census to compare households across the quartiles of private school availability. To save space, we present our results for the sample of sales located within 250 meters of a school attendance boundary only⁵², using successively proximity (table 3.14) and density (table 3.15) as our measures of private school availability. Overall, these figures tend to show that the neighborhoods included in each quartile of private school availability share very similar socio-demographic characteristics. Most census variables, including the distribution of occupations, have almost the exact same values across the different areas, especially when we use proximity to measure private school availability. Table 3.15 shows that small differences appear when use private school density instead of proximity, as areas belonging to the upper quartiles of private school density are slightly more wealthy than those belonging to the lower quartiles (more graduates, less public housing, etc.). However, these differences seem too small to explain our results and, in

⁵²The tables corresponding to the other samples of sales yield similar results and are available upon request.

Table 3.14: Comparison of the socio-demographic characteristics of the four quartiles of private school proximity for sales located within 250 meters of a public middle school attendance boundary. School years 1997-2004.

Variables	Quartiles of proximity to the closest private middle school			
	Lower Quartile	Middle Lower Quartile	Middle Upper Quartile	Upper Quartile
<u>CENSUS VARIABLES</u>				
Nb of census districts	268	361	397	284
Nb of individuals per district	2,271	2,393	2,445	2,447
Nb of households per district	1,223	1,275	1,287	1,286
<u>All households</u>				
% families with children under 25	0.21	0.22	0.22	0.22
% female-headed households	0.06	0.06	0.06	0.06
% owners	0.32	0.31	0.32	0.33
% public housing	0.11	0.13	0.13	0.11
% with graduate degree	0.41	0.40	0.41	0.41
% foreigners	0.23	0.22	0.22	0.22
% unemployed	0.11	0.11	0.11	0.11
Occupation:				
% managers	0.23	0.23	0.23	0.23
% self-employed workers	0.04	0.04	0.04	0.04
% intermediary occupation	0.14	0.14	0.14	0.14
% employees	0.14	0.14	0.14	0.14
% manual workers	0.06	0.06	0.06	0.06
% retired	0.17	0.18	0.18	0.18
% economically inactive	0.21	0.21	0.21	0.22
<u>PUBLIC SCHOOL PERFORMANCE</u>				
Average score at the 2004 DNB exam	9.70	9.63	9.45	9.55
(s.d.)	(1.46)	(1.50)	(1.54)	(1.55)
% entering general curriculum high schools	0.68	0.68	0.67	0.68
(s.d.)	(0.11)	(0.11)	(0.12)	(0.12)
% with privileged backgrounds	0.47	0.46	0.45	0.46
(s.d.)	(0.19)	(0.20)	(0.20)	(0.20)
PUBLIC PREMIUM: Average cross-boundary DNB premium of the “good” over the “bad” public middle school	0.11	0.12	0.11	0.10
(s.d.)	(0.07)	(0.08)	(0.07)	(0.07)
<u>CLOSEST PRIVATE SCHOOL PERFORMANCE</u>				
Average DNB exam score of the closest private school	11.19	10.80	10.79	10.79
(s.d.)	(1.25)	(1.48)	(1.56)	(1.86)
PRIVATE PREMIUM: Average DNB premium of the closest private school over local public schools	0.16	0.13	0.15	0.14
(s.d.)	(0.19)	(0.20)	(0.21)	(0.23)
Number of sales	24,978	24,973	24,985	24,981

Notes: Socio-demographic neighborhood characteristics are taken from the 1999 French Census and are available at the district level. Public middle school performance indexes are computed from the 2004 OCEAN national examinations database and the annual *SCOLARITE* dataset of French pupils (1997-2004). The indicator of the proximity of private middle schools is constructed as follows: for each “reference” sale, we calculate the inverse of its distance to the closest private middle school; we then split this indicator into four quartiles, from the most distant to the closest local private school. The average DNB score of the local private school is the score the private school which is closest to the “reference” sale. For details on the computation of the DNB premium of the closest private school over local public schools, see section 3.7.4.

Table 3.15: *Comparison of the socio-demographic characteristics of the four quartiles of private school density for sales located within 250 meters of a public middle school attendance boundary. School years 1997-2004.*

Variables	Quartiles of local private middle school density			
	Lower Quartile	Middle Lower Quartile	Middle Upper Quartile	Upper Quartile
<u>CENSUS VARIABLES</u>				
Nb of census districts	254	330	342	269
Nb of individuals per district	2,289	2,418	2,440	2,446
Nb of households per district	1,207	1,300	1,292	1,285
<u>All households</u>				
% families with children under 25	0.22	0.21	0.22	0.22
% female-headed households	0.06	0.06	0.06	0.06
% owners	0.30	0.31	0.32	0.33
% public housing	0.16	0.12	0.12	0.10
% with graduate degree	0.39	0.41	0.41	0.42
% foreigners	0.22	0.22	0.22	0.22
% unemployed	0.12	0.12	0.11	0.11
Occupation:				
% managers	0.22	0.23	0.23	0.23
% self-employed workers	0.04	0.04	0.04	0.04
% intermediary occupation	0.16	0.14	0.14	0.13
% employees	0.14	0.15	0.14	0.13
% manual workers	0.07	0.06	0.06	0.05
% retired	0.18	0.18	0.18	0.18
% economically inactive	0.20	0.20	0.21	0.23
<u>PUBLIC SCHOOL PERFORMANCE</u>				
Average score at the 2004 <i>DNB</i> exam	9.48	9.60	9.57	9.59
(s.d.)	(1.40)	(1.48)	(1.59)	(1.61)
% entering general curriculum high schools	0.66	0.67	0.68	0.68
(s.d.)	(0.10)	(0.12)	(0.12)	(0.12)
% with privileged backgrounds	0.44	0.46	0.46	0.47
(s.d.)	(0.18)	(0.20)	(0.20)	(0.21)
PUBLIC PREMIUM: Average cross-boundary <i>DNB</i> premium of the “good” over the “bad” public middle school	0.12	0.11	0.11	0.11
(s.d.)	(0.09)	(0.08)	(0.07)	(0.07)
<u>LOCAL PRIVATE SCHOOL PERFORMANCE</u>				
Average <i>DNB</i> exam score of private schools	10.94	10.88	10.83	10.85
(s.d.)	(0.43)	(0.58)	(0.78)	(1.32)
PRIVATE PREMIUM: Average <i>DNB</i> premium of local private schools over local public schools	0.16	0.14	0.14	0.14
(s.d.)	(0.14)	(0.15)	(0.17)	(0.20)
Number of sales	24,979	24,964	24,994	24,980

Notes: Socio-demographic neighborhood characteristics are taken from the 1999 French Census and are available at the district level. Public middle school performance indexes are computed from the 2004 OCEAN national examinations database and the annual *SCOLARITE* dataset of French pupils (1997-2004). The indicator of the density of private middle schools is constructed as follows: for each “reference” sale, we calculate the average of the inverse of its distance to every private middle school in Paris squared; we then split our indicator into four quarters, from the lowest density of private schools to the highest. The average *DNB* score of the locals private schools is a weighted average of the *DNB* score of all Parisian private schools, the weights being equal to the inverse of their distance (squared) to the “reference” sale. For details on the computation of the *DNB* premium of local private schools over local public schools, see section 3.7.4.

any case, cannot account for the estimates reported in table 3.12 since (as shown in table 3.14) socio-demographic characteristics are essentially identical across the four quartiles of private school proximity.

A second concern is that the mitigating effect of private schools could be driven by local variations in the average cross-boundary public school performance differential. This would be the case in particular if public school differentials are lower in areas where a large number of private schools operate and if small differentials tend to have no impact on housing prices. We examine this issue by comparing average cross-boundary performance differentials across the different quartiles of private school availability. To do so, we compute the average *DNB* score premium of the public school located on the “good” side of the boundary over the the *DNB* score of the school located on the “bad” side⁵³. The results are reported at the bottom of tables 3.14 and 3.15. Average cross-boundary school performance premiums show no particular association with private school availability: in table 3.14, the average premium is 11% for the bottom quartile of private school proximity, 12% for the middle lower quartile, 11% for the middle upper quartile and 10% for the upper quartile. Taken together, these results do not support the idea that our results could be capturing an effect of local variations in the average cross-boundary public school performance differential.

A final concern is that our estimates could in fact reflect the impact of local public-private performance differentials rather than the availability of private schools *per se*. This would be the case if areas where a large number of private schools operate are also those that experience the largest performance premiums of private schools over public schools. To test this alternative explanation, we compare the average performance premium of local private schools in each of our four quartiles of private school availability. When private school availability is measured using proximity to

⁵³For a particular school attendance boundary, the performance premium of the best of the two public middle schools is computed as:

$$\text{PUBLIC PREMIUM} = \frac{\text{DNB1} - \text{DNB2}}{\text{DNB2}}$$

where *DNB1* is the *DNB* exam score of the school located on the “good” side of the boundary and *DNB2* is the *DNB* score of the school located on the “bad” side of the boundary.

the closest private school, the premium is computed as follows:

$$\text{PRIVATE PREMIUM} = \frac{\text{DNBP} - 0.5(\text{DNB1} + \text{DNB2})}{0.5(\text{DNB1} + \text{DNB2})}$$

where DNBP, DNB1 and DNB2 denote the *DNB* exam scores of respectively the closest private middle school, the public middle school assigned to the “reference” sale and the public middle school assigned to the “counterfactual” sale. The formula is the same when we use private school density instead of proximity, except that DNBP is now computed as a weighted average of all private school *DNB* scores, the weights being equal to the inverse of their distance (squared) to the reference sale⁵⁴. The results are reported at the bottom of table 3.14 (using private school proximity) and table 3.15 (using private school density). In both tables, the average *DNB* premium of local private schools over local public schools appears remarkably stable (around 15%) across quartiles and shows no particular association with the level of private school availability. Thus, these results tend to rule out the idea that the mitigating effect of private schools could be essentially driven by local variations in the performance premium of private schools over public schools rather than by the availability of these private schools.

Overall, these investigations suggest that our results are not driven by the confounding factors reviewed in this section and that the outside option provided by private education appears to be the main reason why public middle school cross-boundary price differentials tend to disappear when the number of private schools operating in a particular area increases.

⁵⁴For a particular reference sale i , the local private school performance is therefore computed as:

$$\text{DNBP} = \sum_{j=1}^J w_{i,j} \text{DNBP}_j$$

where J denotes the number of private middle schools operating in the city of Paris, DNBP_j is the *DNB* score of each private schools and the weights $w_{i,j}$ are defined as:

$$w_{i,j} = \frac{\frac{1}{(d_{i,j})^2}}{\sum_{i=1}^J \frac{1}{(d_{i,j})^2}}$$

where $d_{i,j}$ is the distance between sale i and private school j .

3.8 Conclusion

Using comprehensive data on school and housing sales in Paris over the period 1997 to 2004, we find that the performance of public schools has a significant impact on housing prices by comparing price and school performance differentials across school attendance boundaries. A standard deviation increase in the average exam score at the school level raises prices by 1.5 to 2.5% depending on the chosen index of school performance. The size of this effect is similar to existing estimates in the US and UK contexts and can explain roughly 5% of the observed in housing prices differences between adjacent school zones.

We also find evidence that, following the predictions of theoretical models of school choice, private schools tend to attenuate the capitalization of public school performance in housing prices by providing an advantageous outside option to parents. The estimated impact of school performance in neighborhoods belonging to the top quartile of private school density is twice the size of the average effect whereas the coefficient is close to zero in areas belonging to the bottom quartile.

Finally, our results suggest that the coexistence of public and private schools is an important dimension to take into account when designing school assignment policies. In particular, the effect of alternative public school admission rules (strict residence-based assignment, relaxed school zoning, school choice, etc.) on school and residential segregation, pupil performance and educational inequalities will crucially depend on how the housing market incorporates public and private school performance.

Appendix A: The computation of “hedonic” prices

To apply our estimation strategy, we need to compare the evolution of housing prices in adjacent school attendance zones. Unfortunately, it is not possible to compare prices of transactions directly, as two flats rarely share the exact same features.

Because we do not have enough transactions taking place in a given year and within a given zone to be able to match housing units that share the same features, we adopt a hedonic method to construct series of comparable housing prices. We follow the methodology defined by the Insee (the French national statistical agency) to compute a hedonic price index⁵⁵. First, we define large zones (18 in the case of Paris⁵⁶) where prices are known to be fairly homogenous. We then regress the log of housing prices on observable flat features separately for each zone and use the estimated coefficients to compute a “typical flat equivalent price” for each sale. The “typical flat” is defined as a two-room flat located on the ground floor, with one bathroom and medium-sized rooms, without a parking space nor a maid’s room, constructed between 1850 and 1914 and sold during the fourth quarter of the year. Once they are expressed in terms of the “typical flat”, housing prices can be compared across sales because price differences related to structural differences in observable features of the properties are now corrected for.

Within each zone, the estimated housing price equation is the following:

$$\ln p_{i,k,y,q} = a_k + X'_{i,k,y,q} b_k + Y_i \cdot c_k + Q_i \cdot d_k + u_{i,k,y,q} \quad (3.7)$$

where $\ln p_{i,k,y,q}$ is the log of the price per square meter of sale i located in zone k and completed during year y and quarter q ; $X_{i,k,y,q}$ is the vector of flat i ’s features, Y_i and Q_i denote the year and quarter when the sale was completed and $u_{i,k,y,q}$ is the error term.

Regressions are run separately for each zone, to allow the coefficients a_k , b_k , c_k and d_k to vary across different parts of the city. Table 3.16 shows the output of such a regression for zone 1 (which groups the first to fourth *arrondissements*).

⁵⁵The methodology is described in details in Laferrère (2005).

⁵⁶We used geographic and price criteria to define the 18 zones, from the 80 administrative districts of Paris. Zones do not necessarily correspond to the 20 Parisian *arrondissements*.

The “hedonic price” $\ln \tilde{p}_{i,k,y,q}$ of sale i is then calculated as:

$$\ln \tilde{p}_{i,k,y,q} = \ln p_{i,k,y,q} + (X'_0 - X'_{i,k,y,q})\hat{b}_k + (Q_0 - Q_i)\hat{c}_k$$

where $\ln p_{i,k,y,q}$ is the observed price and of sale i , $X_{i,k,y,q}$ is the vector of its observed features, X_0 is the vector of the “typical” flat’s features, Q_i is the quarter when sale i was completed and Q_0 is the quarter of reference for the “typical” flat (fourth quarter). This procedure “converts” the price of sale i into the price it would have been sold to had it shared the same features as the “typical” flat.

Table 3.16: *Example of a hedonic regression for zone 1 (arrondissements 1 to 4), used to convert observed housing prices into the “typical” flat equivalence prices. School years 1997-2004.*

Dependent variable: log of housing price per square meters		
Variable	Coefficient	(Standard error)
Intercept	8.458	(0.019)
AGE OF BUILDING		
Unknown	0.020	(0.009)
Before 1850	0.063	(0.007)
1850-1913	ref.	
1914-1947	-0.003	(0.012)
1948-1969	0.022	(0.015)
1970-1980	0.097	(0.023)
After 1981	0.061	(0.025)
NUMBER OF BATHROOMS		
No bathroom	-0.167	(0.007)
1 bathroom	ref.	
2 bathrooms	0.067	(0.017)
PARKING LOT		
No parking lot	ref.	
1 parking lot	0.182	(0.024)
2 parking lot	0.205	(0.085)
FLOOR		
Ground floor	ref.	
First	0.046	(0.015)
Second	0.077	(0.014)
Third	0.085	(0.014)
Fourth or more (with a lift)	0.081	(0.013)
Fourth or more (without a lift)	0.054	(0.018)
NUMBER OF ROOMS		
One	-0.064	(0.008)
Two	ref.	
Three	0.029	(0.009)
Four	0.073	(0.012)
Five or more	0.032	(0.015)
AVERAGE ROOM SIZE		
Small	0.016	(0.007)
Average	ref.	
Large	0.036	(0.009)
Maid's room	0.068	(0.020)
Number of sales		10,338

Notes: The features of flats sold during school years 1997 to 2004 are computed from the *BIEN (Base d'Informations Économiques Notariales)* dataset, which is managed by the Notary Chamber of Paris and the Île-de-France. The sample is restricted to all arm's-length sales of second hand Parisian flats that took place between September 1997 and August 2004. Sales located on the Île Saint-Louis and Île de la Cité islands were excluded from the sample because of their very specific location and pattern of housing prices. Log of housing prices are in 2004 euros. The Regression includes year and quarter fixed effects.

Troisième partie

REDOUBLEMENT, ORIENTATION ET INÉGALITÉS SELON LE MOIS DE NAISSANCE

L'unification du système éducatif français, initiée à la fin des années 1950, s'est rapidement heurtée à la difficulté de scolariser tous les élèves au sein d'un cursus scolaire indifférencié du primaire à la fin du lycée. C'est pourquoi les réformes qui ont succédé à la réforme Berthoin de 1959 ont instauré un cycle d'observation de deux, puis quatre ans au collège, qui débouche sur des orientations différenciées. En France, l'orientation scolaire peut difficilement être dissociée de la pratique redoublement, dans la mesure où le retard scolaire accumulé par les élèves constitue l'un des critères les plus communément utilisés par les équipes pédagogiques pour déterminer si un élève doit poursuivre ses études dans la voie générale ou dans la voie professionnelle. Dans la troisième partie de cette thèse, nous étudions les conséquences de l'une et l'autre de ces pratiques éducatives sur le destin scolaire et professionnel des individus, à travers le prisme d'un facteur d'inégalité aussi surprenant que sous-estimé : la date de naissance.

Le chapitre 4 propose d'évaluer l'impact de la date de naissance des individus sur leurs trajectoires scolaires et professionnelles. Pour réaliser ce travail, nous avons mobilisé différentes sources qui permettent de mesurer les effets du mois de naissance tout au long du parcours éducatif et de la vie active des individus. Celles-ci incluent des données scolaires issues des panels primaire et secondaire de l'Éducation nationale, de la base Scolarité et de la Base du Diplôme national du Brevet, ainsi que des données sur la population adulte qui proviennent de l'enquête Emploi et des recensements de la population.

Nous commençons par montrer que bien que le mois de naissance des individus ne puisse pas être considéré comme totalement indépendant de leurs caractéristiques socio-démographiques, la composante sociologique de cette variable demeure relativement modeste et présente un certain nombre de propriétés qui permettent d'en neutraliser l'influence dans l'estimation des effets du mois de naissance sur les trajectoires scolaires et professionnelles.

Le premier enseignement de notre étude est que le handicap scolaire subi par les élèves nés en fin d'année par rapport à leurs camarades nés en début d'année, lié au fait qu'ils sont les plus jeunes au sein de leur cohorte scolaire, est plus important et persiste davantage au cours de la scolarité que ne le suggèrent les études menées en

France sur ce thème (Ferrier, 2003 ; Florin et al., 2004 ; Caille et Rosenwald, 2006). Ces dernières sous-estiment en effet l'impact de l'âge relatif des élèves sur leur résultats scolaires, parce qu'elles ne prennent pas suffisamment en compte les biais de sélection induits par le plus fort taux de redoublement des élèves nés en fin d'année. En nous inspirant de la stratégie développée par Bedard et Dhuey (2006), nous évaluons l'impact de l'âge de passage des évaluations sur les performances scolaires en utilisant le mois de naissance comme variable instrumentale. Le mois de naissance d'un élève est en effet directement corrélé à son âge absolu au moment des évaluations (cette corrélation étant égale à un pour les élèves qui n'ont ni redoublé, ni sauté une classe), mais n'exerce pas *a priori* d'influence autonome sur sa réussite scolaire. Les résultats obtenus en appliquant cette stratégie d'estimation indiquent que les écarts de performances liés aux différences d'âge sont particulièrement forts à l'entrée en primaire (tout particulièrement pour les élèves issus de milieux sociaux défavorisés) et tendent à diminuer ensuite, sans pour autant disparaître complètement : en CP, les élèves nés en fin d'année obtiennent en moyenne des scores inférieurs de 70 % d'un écart-type à ceux de leurs camarades nés en janvier ; lorsqu'ils passent les épreuves du Brevet, l'écart reste significatif et égal à 17 % d'un écart-type.

Le second enseignement de ce travail empirique est que le mois de naissance exerce une forte influence tant sur la probabilité de redoublement que sur l'orientation des élèves. Nos estimations indiquent que l'essentiel des effets du mois de naissance sur le redoublement s'exercent pendant la scolarité primaire : à l'âge de 11 ans, un élève né en décembre sur trois a redoublé contre seulement un sur six parmi les élèves nés en janvier. Nous montrons également que le mois de naissance influence significativement l'orientation scolaire : en fin de troisième, le fait d'être né en décembre plutôt qu'en janvier augmente de trois points la probabilité d'être orienté en lycée professionnel.

Le troisième enseignement de cette étude est qu'en dépit de cette différenciation des trajectoires scolaires en fonction du mois de naissance, les individus nés en fin d'années ne paraissent pas fortement pénalisés dans leur vie professionnelle par rapport aux individus nés en début d'année. Nous montrons en effet que leur situation d'emploi n'est pas significativement affectée et que l'écart de rémunération séparant les salariés nés en décembre des salariés nés en janvier est faible puisque compris

entre 1,5 et 3 %. Nous interprétons ces faibles écarts salariaux, qui proviennent en grande partie de l'entrée plus tardive sur le marché du travail des personnes nées en fin d'année, comme la conséquence du fait que le mois de naissance influence moins le niveau de diplôme que le type de formation suivie (générale ou professionnelle). Nos estimations révèlent néanmoins qu'il existe une pénalité salariale spécifiquement liée au plus faible niveau de qualification des hommes nés en fin d'année. Celle-ci se traduit par un écart de salaire égal à 1,5 % entre ceux qui sont nés en décembre et ceux qui sont nés en janvier.

Ces résultats suggèrent que si le mois de naissance des individus n'exerce pas une influence déterminante sur leur destin professionnel, il pénalise néanmoins le parcours scolaire de ceux qui sont nés en fin d'année. Dans la mesure où l'on ne choisit pas sa date de naissance, il paraîtrait souhaitable que le système éducatif prévoie un certain nombre de mécanismes institutionnels pour corriger les inégalités scolaires liés à ce facteur.

Chapitre 4

L'impact du mois de naissance sur les trajectoires scolaires et la vie professionnelle

4.1 Introduction

Phénomène longtemps méconnu, l'impact économique de la date de naissance est devenu un thème de recherche particulièrement actif depuis les travaux pionniers d'Angrist et Krueger (1991) qui ont pour la première fois mis en évidence l'existence d'une relation entre mois de naissance, niveau d'éducation et salaires aux États-Unis. Les recherches empiriques consacrées à ce sujet ont montré que selon leur mois de naissance, les individus n'ont pas le même destin scolaire et professionnel. Pour étonnante qu'elle puisse paraître au premier abord, cette constatation trouve son explication dans une série de facteurs qui renvoient tout autant aux effets de l'âge sur le développement intellectuel des élèves qu'aux spécificités institutionnelles des systèmes éducatifs.

4.1.1 Les effets potentiels du mois de naissance

Les études qui se sont intéressées à l'impact de la date de naissance sur le destin scolaire et professionnel des individus ont montré que les effets de cette variable

étaient susceptibles de transiter par cinq canaux distincts.

Effet de l'âge d'entrée à l'école. Dans la plupart des pays, les élèves sont supposés entrer à l'école élémentaire soit au mois de septembre de l'année civile au cours de laquelle il vont avoir l'âge requis¹ (cas de la France, où l'âge d'entrée à l'école est fixé à 6 ans), soit au mois de septembre de l'année scolaire au cours de laquelle ils vont avoir l'âge requis² (cas notamment des pays anglo-saxons). Par conséquent, selon leur date de naissance, les enfants ne commencent pas tous l'école au même âge, l'écart d'âge d'entrée à l'école pouvant en théorie aller jusqu'à 11 mois. Si l'âge auquel on entre à l'école influence le degré de préparation pour les études, il est possible que ce facteur influence durablement la réussite scolaire.

Effet de l'âge absolu. Dans la mesure où les épreuves d'évaluation ont lieu à la même date pour tous les élèves, ceux qui sont nés en fin d'année sont plus jeunes lorsqu'ils les passent que ceux qui leurs camarades nés en début d'année. On peut donc supposer que leur maturité intellectuelle est moins développée que celle des élèves plus âgés, ce qui risque d'affecter négativement leurs résultats.

Effet de l'âge relatif. Le handicap subi par les élèves nés en fin d'année pourrait s'expliquer non seulement pas le fait qu'il sont plus jeunes au moment des évaluations (effet d'âge absolu), mais également parce qu'ils sont les plus jeunes dans leur classe (en termes relatifs), cette forme d'« infériorité » pouvant pénaliser leur capacité d'apprentissage et leurs performances.

Effet de la durée des études. Un certain nombre de caractéristiques institutionnelles sont susceptibles de modifier la durée des études en fonction du mois de naissance. Un premier facteur tient à l'existence dans certains pays de rentrées décalées en fonction du mois de naissance : au Royaume-Uni, par exemple, certaines académies (*Local Education Authorities*) prévoient deux ou trois points d'entrée à

¹Ce qui signifie qu'une cohorte scolaire comprend l'ensemble des élèves nés entre le 1^{er} janvier et le 31 décembre d'une même année civile.

²Au Royaume-Uni, les enfants entrent à l'école s'ils vont avoir 5 ans au cours de l'année scolaire, si bien qu'une cohorte scolaire comprend l'ensemble des élèves nés entre le 1^{er} septembre de l'année n et le 31 août de l'année $n + 1$.

l'école selon la date de naissance des enfants. En général, les élèves les plus jeunes au sein de leur cohorte scolaire commencent l'école plus tôt (au premier semestre) que ceux qui sont les plus vieux et qui commencent au second ou au troisième trimestre. Le phénomène des rentrées décalées implique que les élèves nés en fin d'année ont passé moins de temps à l'école que leurs camarades nés en début d'année. Un second facteur institutionnel tend à avoir l'effet opposé : dans les pays qui autorisent les élèves à quitter l'école à la date à laquelle ils atteignent l'âge légal de fin d'études (16 ans en France), les élèves les plus jeunes dans leur cohorte scolaire doivent rester en moyenne plus longtemps à l'école que leur camarades plus âgés.

Effet des trajectoires scolaires différenciées. Les élèves nés en fin d'année ont tendance à redoubler davantage que ceux qui sont nés en début d'année. La date de naissance des individus peut également peser sur leur orientation scolaire, si les élèves les plus jeunes dans leur cohorte scolaire ont davantage tendance à être orientés vers des études plus courtes que les autres.

Ces différents effets sont dus tantôt à des facteurs ponctuels (l'orientation, l'obligation scolaire), tantôt à des facteurs permanents (les différences d'âge). Par ailleurs, il ne s'exercent pas *a priori* avec la même intensité tout au long du parcours scolaire : ainsi, on peut raisonnablement supposer que les effets d'âge absolu et d'âge relatif tendent à diminuer à mesure que le nombre de mois qui séparent les élèves les plus jeunes des élèves les plus vieux au sein d'une même classe devient négligeable par rapport à leur durée de vie écoulée. Il faut noter enfin que les inégalités scolaires engendrées par ces facteurs sont susceptibles d'affecter durablement le destin professionnel des individus.

4.1.2 Difficultés méthodologiques

Si les effets potentiels du mois de naissance sur les trajectoires scolaires et professionnelles sont bien identifiés, leur quantification pose d'importantes difficultés méthodologiques.

Démêler les différents effets. Le principal obstacle tient à la difficulté d'estimer séparément les différents effets mentionnés précédemment. Lorsqu'on cherche à évaluer l'impact du mois de naissance sur les performances des élèves en cours de scolarité, les effets de l'âge d'entrée à l'école ne peuvent pas être en général distingués des effets de l'âge absolu de passage des évaluations d'une part, et des effets d'âge relatif d'autre part, dans la mesure où ces trois variables sont parfaitement colinéaires : les élèves nés en fin d'année sont à la fois moins âgés lorsqu'ils rentrent à l'école, sont plus jeunes en termes absolus lorsqu'ils passent les évaluations et sont plus jeunes en termes relatifs par rapport à leurs camarades de classe. Déterminer la part qui revient aux différents effets du mois de naissance dans le destin professionnel des individus est encore plus délicat, dans la mesure où aux effets de maturité peuvent s'être ajoutés les effets de la durée des études et des trajectoires scolaires différenciées.

Biais de sélection. Une seconde difficulté méthodologique tient aux biais de sélection associés au mois de naissance.

Un premier biais de sélection provient du fait que tous les élèves n'entrent pas à l'école à l'âge théorique. Certains parents choisissent de retarder d'un ou deux ans l'inscription de leur enfant en première année d'école primaire, alors que d'autres au contraire choisissent de les inscrire avec un an d'avance. Or le fait de ne pas entrer à l'école à l'âge normal est fortement corrélé avec le mois de naissance : les enfants qui entrent à l'école avec un an d'avance ont plus souvent tendance à être nés en début d'année alors que ceux qui entrent en retard sont plus souvent nés en fin d'année. Il paraît en effet plus raisonnable d'inscrire son enfant en avance si celui-ci n'a pas une trop grande différence d'âge par rapport à ses camarades de classe ; à l'inverse, le choix d'une inscription différée peut s'expliquer par la volonté de certains parents d'éviter à leur enfant né en fin d'année de se retrouver parmi les plus jeunes de sa classe. Ce phénomène tend à réduire l'écart d'âge réel qui sépare les enfants d'une même classe par rapport à l'écart d'âge théorique lié uniquement au mois de naissance. Le fait que les élèves nés en janvier tendent à entrer un peu plus tôt à l'école et les élèves nés en décembre un peu plus tard implique qu'en moyenne, la différence entre l'âge moyen de ces deux groupes d'élèves est inférieure à 11 mois. Dans ces conditions,

régresser les scores individuels sur le mois de naissance tendra à sous-estimer l'effet réel de l'âge sur les performances scolaires.

Un second biais de sélection est lié au fait que les élèves nés en fin d'année ont une plus forte probabilité de redoubler que les élèves nés en début d'année. La conséquence de ce phénomène est que dans une classe donnée, le mois de naissance est fortement corrélé avec la probabilité d'être à l'âge normal. Si l'on se contente de régresser les scores obtenus par les élèves qui n'ont pas redoublé sur leur mois de naissance, on risque donc de fortement sous-évaluer l'impact de l'âge sur les performances scolaires, les élèves les plus jeunes dans la classe ayant tendance à être « sur-sélectionnés » par rapport à leurs camarades nés en début d'année. Le fait de prendre en compte tous les élèves, qu'ils aient redoublé ou non, ne résout que partiellement cette difficulté, dans la mesure où les différences d'âges réelles séparant les élèves d'une même classe seront alors plus faibles que les différences d'âge théoriques liées au mois de naissance³, ce qui induit le même type de sous-estimation des effets de l'âge sur les performances scolaires que le biais lié à l'âge d'entrée à l'école. Ce problème est encore plus prononcé lorsque les élèves ne suivent plus le même cursus éducatif. En effet, si le mois de naissance influence l'orientation scolaire, les effets mesurés au-delà du cursus général sont potentiellement biaisés par le fait qu'en fonction de leur date de naissance, tous les élèves n'ont pas la même probabilité d'être scolarisés dans une filière donnée.

Biais d'endogénéité. Compte-tenu des biais de sélection mentionnés précédemment, il peut sembler préférable d'estimer les effets de l'âge sur les performances scolaires en utilisant non pas le mois de naissance des élèves, mais directement leur âge absolu (en mois) au moment du passage des évaluations. Cette stratégie d'estimation se heurte néanmoins au fait que l'âge auquel on mesure les performances scolaires des élèves est endogène, ce qui signifie qu'il est corrélé avec un certain nombre de caractéristiques individuelles inobservables qui peuvent influencer les performances des élèves indépendamment de leur âge.

Un premier biais d'endogénéité provient du fait que l'âge auquel les élèves passent les évaluations varie selon qu'ils ont commencé l'école à l'âge normal ou non. Un élève

³Parce qu'ils ont redoublé plus souvent, les élèves nés en fin d'année sont en moyenne plus « vieux » que les élèves nés en début d'année.

qui est entré en CP à l'âge de 5 ans passera les épreuves de CP un an plus jeune que ses camarades qui sont nés le même mois mais sont entrés à l'âge normal. Or le fait de ne pas entrer à l'école à l'âge normal est *a priori* corrélé avec les capacités de réussite scolaire et les caractéristiques des parents : les enfants qui entrent à l'école avec un an d'avance sont souvent considérés comme plus doués pour les études que les autres et ont des parents qui les « poussent » davantage à la réussite scolaire. À l'inverse, une entrée retardée à l'école est généralement choisie parce que l'enfant présente un certain retard de maturité intellectuelle. Dans ces conditions, une simple régression des scores obtenus par les élèves sur leur âge au moment des évaluations tendra à sous-estimer l'impact réel de cette variable sur leurs performances scolaires.

Un second biais d'endogénéité tient au fait que l'âge d'un élève au moment des évaluations est corrélé avec le fait d'avoir redoublé ou sauté une classe au cours de sa scolarité. En effet, les élèves les plus vieux au moment des épreuves sont généralement des redoublants, donc *a priori* moins doués pour les études que les autres. Inversement, les élèves les plus jeunes au moment du passage des évaluations ont souvent sauté une classe, ce qui semble indiquer qu'ils manifestent de meilleures dispositions pour la réussite scolaire que leurs camarades. La combinaison de ces deux phénomènes induit mécaniquement une forte corrélation négative entre l'âge absolu de passage de l'épreuve et le terme d'erreur de la régression, ce qui conduit à sous-estimer l'impact causal de l'âge sur les performances scolaires, voire à en inverser le signe.

Le mois de naissance est-il exogène ? Une ultime difficulté liée à la mesure des effets du mois de naissance tient à la possible corrélation de cette variable avec un certain nombre de caractéristiques socio-démographiques qui peuvent influencer le destin scolaire et professionnel des individus indépendamment des différents effets mentionnés plus haut.

Le mois (ou le trimestre) de naissance des individus a été abondamment utilisé dans la littérature comme une source d'identification pour étudier les rendements de l'éducation (Angrist et Krueger, 1991) ou l'offre de travail des femmes (Gelbach, 2002). Cette variable semble en effet présenter les deux propriétés d'un bon instrument : la date de naissance d'un individu agit sur un certain nombre de va-

riables explicatives (la durée des études, par exemple) sans avoir *a priori* d'influence propre sur la variable dépendante (le salaire individuel, l'offre de travail des mères, etc.).

Un certain nombre d'auteurs ont néanmoins questionné la validité de cet instrument. Une première critique porte sur l'intensité de la corrélation qui unit le trimestre de naissance aux variables que l'on cherche à instrumenter. Bound et al. (1995) ont montré en effet qu'une relation trop faible peut créer un biais important dans l'estimation. Une seconde critique, plus radicale, porte sur le degré d'exogénéité du mois de naissance par rapport aux variables d'intérêt. Bound et Jaeger (2000) montrent par exemple que la corrélation identifiée par Angrist et Krueger (1991) entre le trimestre de naissance et le salaire n'est pas nécessairement due aux lois sur l'obligation scolaire, mais peut transiter par d'autres canaux : outre les effets de l'âge sur les performances scolaires mentionnés précédemment, la date de naissance d'un individu est potentiellement corrélée avec les caractéristiques socio-démographiques de ses parents, si la saisonnalité des naissances varie d'un groupe social à un autre. Il serait donc possible que les effets identifiés en utilisant cette variable traduisent en réalité l'influence d'autres déterminants du salaire que la seule durée des études.

4.1.3 Que sait-on sur la part des différents effets du mois de naissance ?

Les études consacrées à l'impact de la date de naissance des individus sur leur destin scolaire et professionnel s'efforcent de proposer des stratégies d'estimation qui permettent de contourner en partie les difficultés méthodologiques évoquées dans la section précédente. Si les résultats de ces travaux empiriques fournissent de précieux renseignements sur l'ampleur des différents effets du mois de naissance, peu d'entre eux parviennent réellement à les démêler.

Effet de l'âge d'entrée à l'école. L'intérêt des économistes s'est porté en priorité sur les effets de l'âge d'entrée à l'école sur les performances scolaires et la vie professionnelle. Cette question revêt en effet une importance particulière du point de vue de l'efficacité des politiques éducatives, dans la mesure où retarder ou au contraire

réduire l'âge auquel les enfants entrent à l'école pourrait améliorer à moindre coût les performances globales du système scolaire.

Les premières études consacrées à ce thème se contentent soit de comparer les performances des élèves entrés à l'école primaire à l'âge normal (généralement dans l'année de leur 6 ans) avec celles des élèves entrés avec un an d'avance ou un an de retard⁴, soit de comparer les scores des élèves nés la même année en utilisant le fait qu'en raison de leur mois de naissance, ils ne sont pas entrés à l'école au même âge⁵. Parce qu'elles ne contrôlent pas pour les biais de sélection mentionnés plus haut, ces études ne permettent pas en réalité de mesurer de manière convaincante l'impact causal de l'âge d'entrée à l'école sur les performances scolaires.

Les limites méthodologiques de ces deux types d'approches ont été en partie résolues par l'utilisation de variables instrumentales. Angrist et Krueger (1992) ont été les premiers à montrer que le mois de naissance des individus pouvait constituer un instrument crédible pour estimer l'impact de l'âge d'entrée à l'école sur le destin scolaire et professionnel des individus. En effet, du fait de la définition des cohortes scolaires, le mois de naissance est fortement corrélé avec l'âge auquel les élèves entrent à l'école mais ne paraît pas *a priori* susceptible d'influencer de manière indépendante leurs performances scolaires.

Les résultats produits par les très nombreuses études qui ont utilisé cette stratégie d'estimation sont loin d'être univoques : alors que certains auteurs identifient un impact légèrement positif d'une entrée plus tardive à l'école sur les performances scolaires (Puhani et Weber 2005 ; Datar, 2006), d'autres ne trouvent aucun effet (Mayer et Knutson, 1998 ; Fertig et Kluge 2005). La principale limite de ces travaux est que les écarts de scores éventuellement observés en fonction de l'âge d'entrée à l'école peuvent tout aussi bien provenir des différences de maturité au moment où les élèves sont soumis aux tests de connaissances. En effet, les évaluations à partir desquelles sont mesurées leurs performances scolaires ont généralement lieu à une date fixe, ce qui implique que les élèves les plus jeunes au sein de leur cohorte scolaire sont

⁴Pour des travaux de ce type menés sur données américaines, on pourra se reporter à Langer et al. (1984), Cameron et Wilson (1990), May et al. (1995) ainsi qu'à Graue et di Perna (2002).

⁵Cf. Langer et al. (1984), Cahan et Davis (1987), Sweetland et de Simone (1987), Cameron et Wilson (1990), Jones et Mandeville (1990), Bickel et al. (1991), Crosser (1991), Morrison et al. (1997), McClelland et al. (2000), Stipek et Byler (2001).

également plus jeunes que les autres lorsqu'ils passent ces épreuves.

Pour éviter que les estimations ne soient contaminées par les effets de maturité, un certain nombre d'études ont cherché à mesurer l'impact de l'âge d'entrée à l'école une fois que les individus sont entrés dans la vie active. Cette approche repose sur l'idée que les écarts de maturité observés à l'école en fonction de la date de naissance des individus n'exercent plus d'influence à l'âge adulte, si bien les écarts constatés dans le niveau d'éducation ou le salaire des membres d'une même cohorte scolaire peuvent être attribués aux effets de l'âge auquel ils ont commencé les études. Sur données américaines, Angrist et Krueger (1992) ne trouvent pas d'impact significatif de l'âge d'entrée à l'école sur le nombre total d'années d'études ni sur la probabilité d'obtenir un diplôme universitaire. Sur données administratives suédoises, Fredriksson et Ockert (2006) montrent que l'âge d'entrée à l'école a un impact faiblement positif sur le nombre d'années d'études, mais que l'effet sur les salaires est nul, en raison de la perte d'expérience professionnelle associée à une entrée plus tardive sur le marché du travail. La principale limite de cette approche est qu'elle ne permet pas d'affirmer avec certitude que la persistance des effets du mois de naissance à l'âge adulte est due principalement à l'âge d'entrée à l'école. Il est possible en effet que les moins bons résultats obtenus par les élèves les plus jeunes au sein de leur cohorte scolaire aient pu influencer leurs trajectoires éducatives indépendamment de l'âge auquel ils ont entamé leur scolarité.

Dans l'ensemble, ces études ne permettent pas d'aboutir à une conclusion tranchée concernant le rôle exact joué par l'âge d'entrée à l'école dans le destin scolaire et professionnel des individus. Les travaux existants suggèrent néanmoins que s'ils existent, ces effets sont relativement limités et qu'ils jouent un rôle beaucoup moins important que les effets de maturité pour expliquer les écarts de performances constatés entre élèves d'une même classe.

Effets de l'âge sur les performances scolaires. L'âge d'entrée à l'école ne constitue sans doute pas la principale explication des moins bons résultats scolaires obtenus par élèves les plus jeunes dans leur cohorte scolaire. De nombreux travaux de recherche ont montré en effet qu'il existe un handicap propre lié au fait pour un élève

d'être plus jeune au moment du passage des évaluations et que la prise en compte de ce phénomène par les politiques éducatives dépasse largement la question de l'âge optimal d'entrée à l'école.

Les moindres performances scolaires obtenues par les élèves les plus jeunes dans leur classe est une réalité qui a été documentée dans un certain nombre de pays. La plupart des études mesurent cet effet en comparant les scores obtenus par les élèves aux évaluations des compétences en fonction de leur mois de naissance ou en fonction de leur âge à la date de passage des épreuves⁶. La conclusion à laquelle elles aboutissent le plus souvent est que la pénalité subie par les élèves les plus jeunes est relativement forte à l'école primaire, mais qu'elle est transitoire et disparaît presque totalement au-delà. On peut néanmoins suspecter ces travaux de sous-évaluer dans une assez large mesure les effets de l'âge sur les performances scolaires, dès lors qu'ils ne tiennent pas compte des biais de sélection et des biais d'endogénéité mentionnés plus haut.

Pour contourner ces difficultés méthodologiques, Bedard et Dhuey (2006) proposent d'instrumenter l'âge des élèves à la date des évaluations par leur mois de naissance. Ces auteurs utilisent les données de l'enquête TIMSS⁷ qui mesure les compétences en mathématiques et en sciences d'élèves scolarisés dans pays de l'OCDE et âgés de 9 et 13 ans. La manière différente dont les cohortes scolaires sont définies d'un pays à l'autre leur permettent de contrôler pour d'éventuels biais liés à la composante sociologique du mois de naissance. Leurs résultats indiquent que l'impact de l'âge sur les performances des élèves est fort et persistant, dans la mesure où la pénalité associée à 11 mois de différence d'âge est comprise entre 12 et 35 % d'un écart-type à l'âge de 9 ans et entre 8 et 26 % d'un écart-type à l'âge de 13 ans. La principale limite de cette étude est que les coefficients qui sont estimés peuvent refléter également les effets de l'âge d'entrée à l'école et de l'âge relatif des élèves, en plus de l'effet de l'âge absolu au moment du passage des évaluations.

Une autre stratégie proposée pour neutraliser les biais de sélection et d'endogénéité dans l'estimation des effets d'âge sur les performances scolaires consiste à utiliser des échantillons d'élèves sélectionnés sur le seul critère de l'âge et indépendamment

⁶Cf. Bells et Daniels (1990), Thomas (1995), Sharp (1995).

⁷Trends in International Mathematics and Science Study.

de la classe qu'ils fréquentent. De tels données permettent de mesurer l'impact de l'âge au moment du passage des évaluations sans avoir recours à des instruments. En utilisant les résultats obtenus à l'âge de 15 ans par les élèves norvégiens aux évaluations conduites dans le cadre de l'enquête internationale PISA⁸, Strøm (2004) estime que le fait d'être né en décembre plutôt qu'en janvier réduit en moyenne les scores de 20 % d'un écart-type. Ces résultats souffrent néanmoins d'un certain nombre de limites. D'une part, l'enquête PISA ne concerne que des élèves relativement vieux, donc *a priori* moins sensibles aux effets d'âge que les élèves plus jeunes. D'autre part, cette enquête ne contient que des échantillons nationaux relativement réduits (de l'ordre de quelques milliers par pays), ce qui nuit à la précision des estimations. Enfin, l'approche empirique utilisée par Strøm ne permet pas de mesurer les effets de maturité indépendamment des autres effets du mois de naissance.

Crawford et al. (2007) tirent avantage des spécificités institutionnelles du système éducatif britannique pour estimer séparément les effets de l'âge absolu, de l'âge relatif et de l'âge d'entrée à l'école sur les performances scolaires. L'absence quasi totale de redoublement au Royaume-Uni réduit considérablement les biais de sélection et les biais d'endogénéité qui compliquent traditionnellement l'estimation de ces effets. Par ailleurs, le fait que les règles qui gouvernent l'âge d'entrée à l'école primaire varient d'une académie à une autre permet d'isoler l'impact propre de cette variable sur les résultats aux évaluations. Pour estimer l'effet de l'âge relatif relatif sur les performances scolaires, les auteurs exploitent le fait que la position d'un élève dans la hiérarchie des âges ne dépend pas uniquement de sa date de naissance, mais également de celle de ses camarades. Ils proposent donc de mesurer l'âge relatif en calculant la proportion d'enfants plus âgés qu'un élève donné, parmi ceux qui sont scolarisés

⁸Programme international pour le suivi des acquis des élèves. Réalisée tous les trois ans par l'Organisation de coopération et de développement économiques (OCDE), cette vaste enquête a pour objectif de mesurer les acquis des élèves de 15 ans dans une soixantaine de pays (pour une présentation détaillée des méthodes et des principaux enseignements du programme PISA pour la France, voir Grenet (2008)). Les compétences mesurées par PISA relèvent de trois domaines de compétences distincts : la compréhension de l'écrit, la culture mathématique et la culture scientifique. Ces trois domaines ont été abordés successivement lors des trois vagues de l'enquête (2000, 2003, 2006). Les scores attribués aux élèves dans chacun des trois domaines évalués sont calculés à partir de leurs réponses à une série d'items mis au point par des groupes d'experts internationaux à partir d'un cadre de référence établi au niveau international. L'échelle de ces scores a été conçue de manière à ce que le score moyen des pays de l'OCDE soit égal à 500 points et que deux tiers environ des élèves obtiennent entre 400 et 600 points (ce qui correspond à un écart-type de 100).

dans la même section et la même école. Leurs résultats indiquent que la pénalité associée à une différence d'âge de 11 mois est égale à 80 % de l'écart-type de la distribution des scores à l'âge de 5 ans et décroît à mesure que les élèves avancent dans le cursus éducatif, sans pour autant disparaître totalement : à l'âge de 15 ans, l'écart de performance représente encore 12 % d'un écart-type. Leurs analyses empiriques révèlent également que la quasi-totalité des effets estimés provient de la maturité différente des élèves au moment du passage des évaluations et que les effets de l'âge d'entrée à l'école (qui, en raison du phénomène des rentrées décalées, induisent des durées d'études différentes en fonction du mois de naissance) et d'âge relatif sont négligeables.

Effets du mois de naissance sur la durée des études. Un certain nombre de travaux empiriques ont montré qu'en raison des règles qui régissent l'âge auquel les élèves sont autorisés à commencer puis à interrompre leurs études, le mois de naissance peut influencer significativement la durée totale de la scolarité des individus.

L'impact du mois de naissance sur la durée des études peut provenir des modalités d'admission à l'école primaire. Aux Pays-Bas, par exemple, les élèves doivent obligatoirement être inscrits au cours préparatoire le jour de la rentrée, s'ils vont avoir 5 ans au cours de l'année scolaire. Les parents ont néanmoins la possibilité d'y inscrire leur enfant dès que celui-ci atteint l'âge de 4 ans et 98 % des élèves entrent à l'école avant l'âge légal. Les enfants qui entrent à l'école à l'âge de 4 ans peuvent donc passer jusqu'à deux ans en cours préparatoire. Cette spécificité du système néerlandais implique qu'en fonction de leur mois de naissance, les élèves d'une même cohorte scolaire n'ont pas la même durée moyenne de scolarité. Leuven et al. (2006) exploitent le fait que cette corrélation entre durée effective de scolarité et mois de naissance n'existe pas pour les élèves nés pendant l'été (puisque une inscription anticipée ne peut intervenir pendant les vacances scolaires) pour évaluer séparément les effets de la durée de scolarité et les effets de l'âge sur les performances scolaires. Les résultats de leurs estimations indiquent que si l'effet de la scolarité anticipée sur le score des élèves à l'âge de 6 ans est beaucoup plus faible que l'effet de leur âge au moment des évaluations, il est néanmoins positif (mais non significatif) : chaque mois

supplémentaire de scolarité anticipée augmenterait le score des élèves d'environ 2,5 % d'un écart-type.

L'effet du mois de naissance sur la durée des études peut également résulter du fonctionnement de l'obligation scolaire, qui influence la date à laquelle les élèves sont autorisés à quitter l'école. Dans un article célèbre, Angrist et Krueger (1991) montrent qu'aux États-Unis, les individus nés au premier trimestre ont durée moyenne d'études légèrement plus faible que ceux qui sont nés au dernier trimestre. D'après ces auteurs, un tel phénomène provient des lois qui régissent l'obligation scolaire et autorisent les élèves à quitter l'école au moment où ils atteignent l'âge légal de fin d'études. Dans ces conditions, Angrist et Krueger interprètent la croissance du salaire moyen en fonction du trimestre de naissance comme la conséquence directe de cette variation exogène dans la durée d'études.

Outre les réserves émises par Bound et al (1995) à propos de l'exogénéité du trimestre de naissance, l'étude d'Angrist et Krueger a été critiquée du point de vue de l'interprétation qu'ils donnent de la corrélation entre trimestre de naissance et salaires. En effet, L'obligation scolaire ne constitue pas le seul mécanisme par lequel la date de naissance des individus est susceptible d'avoir des effets de long terme sur leur vie professionnelle. Plug (2001) montre par exemple qu'aux Pays-Bas, le mois de naissance influence l'âge de fin d'études indépendamment de l'obligation scolaire et que les individus les plus âgés au sein de leur cohorte scolaire tendent à avoir des niveaux d'éducation et des salaires plus élevés que les individus plus jeunes dans la même cohorte. L'auteur interprète ce phénomène comme une indication de la persistance à l'âge adulte des effets de maturité liés aux différences d'âge séparant les élèves d'une même classe. Sur données suédoises, Fredriksson et Ockert (2006) trouvent également que le nombre d'années d'études est influencé par le mois de naissance (les élèves les plus âgés dans leur cohorte restant plus longtemps à l'école que les élèves plus jeunes) et que cette relation n'a aucun lien avec les lois qui régissent l'obligation scolaire, dans la mesure où en Suède, les élèves ne sont pas autorisés à arrêter leurs études avant la fin de l'année scolaire. Cette corrélation est interprétée par les auteurs comme traduisant l'impact positif de l'âge d'entrée à l'école sur le niveau d'éducation atteint à la fin de la scolarité, bien qu'en réalité, ce phénomène ne puisse être distingué

des effets de maturité⁹.

Effets du mois de naissance sur les trajectoires scolaires. Assez peu d'études se sont intéressées aux effets du mois de naissance sur les trajectoires scolaires. Si l'influence du mois de naissance sur la probabilité de redoubler est assez bien documentée (Eide et Showalter, 2001 ; Martin et al., 2004), la possibilité que ce facteur puisse influencer l'orientation scolaire des individus a été peu analysée. Cette lacune s'explique en grande partie par le fait que l'orientation intervient relativement tard dans la plupart des systèmes éducatifs, à un âge où l'on suppose généralement que les effets de maturité ne sont plus suffisamment forts pour peser significativement sur les trajectoires scolaires.

Le cas de l'Allemagne a néanmoins fait l'objet d'une étude menée par Jürges et Schneider (2004) à partir des données de l'enquête PISA. Le système éducatif allemand présente en effet la particularité d'orienter les élèves très tôt dans leur scolarité : à l'âge de 10 ans, les écoliers allemands sont envoyés soit dans un *Gymnasium* pour y poursuivre des études secondaires généralistes, soit dans une *Hauptschule* ou une *Realschule* pour y acquérir une formation professionnelle. Les estimations des auteurs indiquent que la probabilité d'être orienté dans la voie professionnelle est supérieure de 6 points pour les élèves les plus jeunes dans leur cohorte scolaire par rapport aux élèves les plus âgés.

Quels sont les principaux enseignements qu'il faut retenir de cette vaste littérature consacrée aux effets du mois de naissance ? Le premier est que du point de vue des performances scolaires, l'effet de l'âge absolu domine largement les autres effets (âge relatif, âge d'entrée à l'école, durée de scolarisation). Le second enseignement est que les effets de maturité du mois de naissance tendent à s'atténuer rapidement à mesure que les élèves progressent dans leur parcours éducatif. Le troisième enseignement est que trois mécanismes distincts peuvent expliquer la persistance des effets du mois de naissance à l'âge adulte : le redoublement ou l'orientation différenciée des

⁹Les élèves qui commencent l'école à un âge plus avancé parce qu'ils sont nés juste après la date séparant deux cohortes scolaires sont également les plus âgés au moment du passage des évaluations. Leur meilleure réussite scolaire aux examens, liée à leur plus grande maturité, pourrait expliquer qu'ils prolongent leurs études plus longtemps que les élèves les plus jeunes dans leur cohorte.

élèves, l'influence des effets de maturité sur le nombre d'années d'études et l'effet de l'obligation scolaire sur la durée totale de scolarisation.

4.1.4 Études sur données françaises

Parce qu'une grande partie des données permettant d'étudier cette question ont longtemps été inaccessibles aux chercheurs, les effets du mois de naissance ont été relativement peu analysés en France. Le système éducatif français présente pourtant un certain nombre de spécificités institutionnelles qui en font un terrain d'étude privilégié pour analyser ces effets.

Caractéristiques institutionnelles du système éducatif français. En France, l'instruction est obligatoire à partir de 6 ans. Les enfants sont inscrits à l'école élémentaire à la rentrée scolaire de l'année civile au cours de laquelle ils atteignent l'âge de 6 ans, ce qui signifie qu'une cohorte scolaire est en principe composée de tous les élèves nés entre le 1^{er} janvier et le 31 décembre d'une même année. Les parents ont la possibilité d'envoyer leurs enfants à l'école maternelle à partir de l'âge de 3 ans ou de 2 ans si leur enfant a atteint cet âge le jour de la rentrée scolaire et que le nombre de places disponibles le permet. L'entrée à l'école maternelle a lieu en principe au mois de septembre, mais certains établissements autorisent les enfants à entamer leur scolarité pré-élémentaire en cours d'année. L'admission des élèves à l'âge de 2 ans étant généralement réservée en priorité aux enfants les plus âgés, ceux qui sont nés en début d'année bénéficient en moyenne d'une scolarité pré-élémentaire plus longue que leurs camarades nés en fin d'année.

La scolarité est obligatoire jusqu'à l'âge de 16 ans révolus, ce qui signifie qu'un élève est en principe autorisé à quitter l'école le jour de ses 16 ans. Le fonctionnement de l'obligation scolaire en France est donc assez proche de la pratique américaine. Aussi peut-on s'attendre à ce qu'en moyenne, les élèves nés en fin d'année quittent l'école un peu plus tôt que les autres.

Une spécificité importante du système français tient à la place importante qu'y occupent le redoublement et l'orientation. Le redoublement est massivement pratiqué tant à l'école primaire qu'au collège et on peut s'attendre à ce qu'il touche davantage

les élèves nés en fin d'année que les élèves nés en début d'année. L'orientation scolaire, quant à elle, intervient à la fin de la classe de troisième : le conseil de classe du dernier trimestre propose, en réponse aux vœux des familles, soit une orientation en seconde générale et technologique, soit une orientation en seconde professionnelle ou en première année de CAP. En cas de désaccord entre les vœux de la famille et la proposition du conseil de classe, la décision d'orientation intervient après un entretien obligatoire avec le chef d'établissement. Si le désaccord persiste, les parents peuvent obtenir le maintien de l'élève dans sa classe d'origine pour une durée d'un an au maximum¹⁰. Les critères les plus couramment utilisés pour décider de l'orientation d'un élève étant les résultats obtenus en troisième et le retard scolaire accumulé, il est possible que le mois de naissance influence la probabilité d'être orienté dans l'une ou l'autre voie.

Études empiriques sur données françaises. Les rares études consacrées en France aux effets du mois de naissance se concentrent essentiellement sur la scolarité pré-élémentaire et élémentaire.

L'impact du mois de naissance sur la durée de la scolarité en maternelle a été analysé par Goux et Maurin (2008) dans le cadre d'une étude plus large sur l'impact de la scolarisation à 2 ans sur l'offre de travail des femmes. Ces auteurs montrent que la probabilité d'être admis à la maternelle à l'âge de 2 ans est fortement corrélée avec le mois de naissance : la priorité étant généralement donnée aux enfants les plus âgés au sein de leur cohorte scolaire, on constate que la durée moyenne de scolarité pré-élémentaire décroît avec le mois de naissance. Goux et Maurin utilisent le fait que le taux d'admission des élèves nés pendant l'été varie fortement d'une région à l'autre en raison de l'hétérogénéité des capacités d'accueil des écoles maternelles, pour évaluer l'impact spécifique de la durée de scolarité pré-élémentaire sur les résultats scolaires en début de CE2 et sur la probabilité de quitter l'école précocement (à l'âge de 17, 18 ou 19 ans). Leurs estimations indiquent que le fait d'avoir passé plus de temps à l'école maternelle n'a d'impact significatif ni sur les performances scolaires des élèves en CE2, ni sur leur probabilité d'interrompre précocement leurs études. Les

¹⁰Article D 331-37 du Code de l'Éducation (décret du 23 mai 2006).

auteurs interprètent ce résultat comme traduisant le fait que la scolarisation précoce en maternelle ne bénéficie pas davantage aux enfants que la garde à domicile ou l'inscription en crèche.

Les effets de maturité du mois de naissance sur les performances scolaires ont été documentés en France pour les élèves scolarisés en maternelle et à l'école primaire. Une étude menée par Ferrier (2003) sur un échantillon de 40 000 élèves de l'académie de Poitiers montre que les écarts de réussite en fonction des différences d'âge sont importants : à l'évaluation nationale de CE2, les élèves nés en début d'année obtiennent en moyenne 6,4 points de plus que ceux de la fin de l'année, et en mathématique cette différence s'élève à 7,6 points. En sixième, le nombre d'élèves ayant une ou deux années de retard scolaire est deux fois plus élevé parmi ceux qui sont nés en décembre que parmi ceux qui sont nés en janvier. En utilisant les données du panel primaire de l'Éducation nationale, Florin et al. (2004) trouvent également que les élèves nés en début d'année civile obtiennent de meilleures performances scolaires que ceux nés en fin d'année et ont une probabilité plus élevée de suivre une scolarité sans redoubler. Caille et Rosenwald (2006) utilisent les mêmes données que Florin et al. et montrent que toutes choses égales par ailleurs¹¹, les élèves nés au quatrième trimestre réussissent en moyenne 6,3 items de plus aux évaluations de CP que les élèves du premier trimestre. Caille et Rosenwald trouvent par ailleurs qu'à l'entrée en sixième, seuls 12 % des élèves nés au premier trimestre ont redoublé au cours de leur scolarité primaire, contre 21 % des élèves nés au quatrième trimestre. D'après ces auteurs, la meilleure réussite scolaire des élèves nés en début d'année est liée exclusivement à des acquis plus élevés au sortir de l'école maternelle et les écarts de performance tendent à disparaître complètement à la fin de l'école primaire : leurs estimations indiquent même qu'en sixième, l'écart de réussite s'inverse au bénéfice des élèves de la fin d'année, puisque tant en mathématiques qu'en français, ils réussissent un item de plus en moyenne que les élèves du premier trimestre.

On peut craindre cependant que ces différentes évaluations ne sous-estiment l'impact réel du trimestre de naissance sur les performances scolaires, parce qu'elles ne

¹¹Les contrôles inclus dans la régression sont, outre l'année et le trimestre de naissance, le sexe de l'élève, son rang dans la fratrie, la PCS de la personne de référence, le niveau de diplôme du père et de la mère, l'activité de la mère, la taille de la famille, la structure parentale, le rapport des parents à l'immigration et l'âge d'entrée à l'école maternelle.

prennent pas suffisamment en compte les biais de sélection mentionnés précédemment. Caille et Rosenwald mesurent par exemple les écarts de performance en sixième en comparant les scores des élèves nés au premier trimestre de l'année 1991 à celui des élèves nés au quatrième trimestre de la même année. Le problème est que tous les élèves ne passent pas les épreuves d'évaluations de sixième au même âge, selon qu'ils sont en avance ou ont redoublé. Or, parce que les élèves de la fin de l'année ont plus souvent redoublé que les élèves du début de l'année, le différentiel d'âge réel séparant les élèves du premier et du quatrième trimestre au moment du passage des évaluations est inférieur au différentiel d'âge théorique (11 mois) que l'on aurait constaté si tous les élèves redoublaient avec la même probabilité quel que soit leur mois de naissance. Ce problème se pose de manière symétrique pour les élèves qui sont entrés en sixième avec un an d'avance et qui ont donc passé les évaluations à un âge plus jeune que les autres, dans la mesure où ces élèves sont généralement nés pendant les premiers mois de l'année. Pour les raisons mentionnées dans la sous-section 4.1.2, de tels biais conduisent à sous-estimer l'impact réel de l'âge des élèves sur leurs performances scolaires.

4.1.5 Contribution

L'objectif de la présente étude est de produire une évaluation la plus systématique possible des effets du mois de naissance dans le contexte éducatif français, en mobilisant un grand nombre de sources permettant de suivre les individus tout au long de leur parcours éducatif et de leur vie professionnelle.

Par rapport aux études existantes, notre travail innove en cherchant à corriger les biais d'estimation qui pourraient provenir de la composante sociologique du mois de naissance. Nous montrons en effet que la relation qui unit la date de naissance aux caractéristiques socio-démographiques des individus présente un certain nombre de propriétés qui permettent d'en limiter l'influence dans l'estimation des effets du mois de naissance sur les trajectoires scolaires et professionnelles.

Une contribution importante de cette étude concerne l'évaluation de l'impact des différences de maturité associées au mois de naissance sur les performances scolaires des élèves. Nous montrons que les effets de maturité sont à la fois plus forts et

persistent plus longtemps que ne le laissent entendre les études françaises sur le sujet : en CP, la pénalité associée à 11 mois d'écart d'âge au moment du passage des évaluations est égale à 70 % d'un écart type ; lorsque les élèves passent le diplôme du Brevet, l'écart de performance est significatif et égal à 17 % d'un écart-type. Nous parvenons à cette conclusion en adoptant une stratégie d'estimation qui permet de neutraliser les biais de sélection et d'endogénéité associés au mois de naissance. Nous montrons également qu'à l'école primaire, les élèves souffrent d'autant plus des effets de maturité qu'ils sont issus de milieux sociaux moins favorisés.

Une seconde contribution de cette étude concerne la mise en évidence des effets du mois de naissance sur les trajectoires scolaires, un sujet qui a été relativement peu abordé dans la littérature. Nos estimations confirment que l'influence du mois de naissance sur la probabilité de redoubler s'exerce essentiellement à l'école primaire : à l'âge de 11 ans, les élèves nés en décembre ont redoublé deux fois plus souvent que les élèves nés en janvier. Nous montrons par ailleurs que les effets du redoublement ne semblent pas avoir d'influence propre sur les performances des élèves, dans la mesure où nos estimations des écarts de performances liés au mois de naissance aux différents stades du cursus éducatif sont quasiment identiques aux résultats obtenus par Crawford et al. (2007) pour le Royaume-Uni, où le redoublement n'est pas pratiqué. En revanche, le taux de redoublement plus élevé des élèves explique sans doute en partie que les élèves nés en fin d'année tendent à être davantage orientés dans la voie professionnelle que les autres : nos résultats indiquent que le fait d'être né en décembre plutôt qu'en janvier augmente de trois points la probabilité d'être orienté en lycée professionnel après la classe de troisième.

La troisième contribution de notre étude concerne l'analyse de la persistance des effets du mois de naissance à l'âge adulte. Nous montrons qu'en France, le mois de naissance influence significativement le niveau de formation des individus. Cependant, à la différence des autres pays où ce phénomène a pu être constaté (Angrist et Krueger, 1991 ; Plug, 2001 ; Fredriksson et Ockert, 2006), la différenciation du niveau de formation en fonction du mois de naissance ne concerne pas la durée totale de scolarisation, mais plutôt le type de formation suivie. Nos résultats indiquent en effet que les individus nés en fin d'année ont une probabilité plus forte de quitter l'école

avec un diplôme de l'enseignement professionnel (CAP ou BEP) qu'avec un diplôme de l'enseignement général (BEPC ou Baccalauréat), la différence étant de l'ordre de 2 points. Nous interprétons ce phénomène comme la conséquence directe de l'impact du mois de naissance sur l'orientation en fin de troisième décrite précédemment. Nos analyses révèlent cependant qu'en dépit de son effet sur le niveau de formation, le mois de naissance ne semble pas exercer une influence déterminante sur le destin professionnel des individus. Nous montrons en effet que la situation d'emploi des individus nés en fin d'année n'est pas significativement différente de celle des individus nés en début d'année. Il existe cependant une faible pénalité salariale, de l'ordre de 1,5 à 3 % associée au fait d'être né en décembre plutôt qu'en janvier. Cette écart salarial s'explique en partie par l'entrée un peu plus tardive sur le marché du travail des individus nés en fin d'année mais également, dans le cas des hommes, par le niveau de qualification légèrement plus faible de ceux qui sont nés en fin d'année.

Au total, les résultats de nos recherches nous conduisent à réévaluer sensiblement le rôle du mois de naissance comme facteur d'inégalités scolaires. Il apparaît en effet que bien que n'ayant pas une forte incidence à long terme sur leur destin professionnel, le mois de naissance pénalise la réussite scolaire des individus nés en fin d'année plus lourdement qu'on ne le suppose habituellement. Ce phénomène invite à réfléchir aux mécanismes institutionnels qui pourraient être mis en place pour tenter de corriger cette source d'inégalités.

Le reste de cette étude est organisé comme suit : nous présentons dans la section 4.2 les données utilisées pour réaliser nos estimations ; la section 4.3 évalue l'importance de la composante sociologique du mois de naissance et discute les moyens qui peuvent être mis en œuvre pour en neutraliser l'influence dans les régressions ; les résultats de l'estimation des effets du mois de naissance sur les performances scolaires sont présentés dans la section 4.4 ; leur impact sur les trajectoires scolaires sont évalués dans la section 4.5 et leur persistance à l'âge adulte dans la section 4.6 ; la section 4.7 conclut cette étude.

4.2 Données

Les données permettant d'étudier l'influence de la date de naissance sur les performances scolaires, l'emploi et les salaires en France sont suffisamment nombreuses pour qu'on puisse en apprécier les effets aux différents étapes de la vie scolaire et professionnelle des individus. La date de naissance est en effet une variable fréquemment incluse tant dans les enquêtes de l'Insee que dans les sources statistiques du ministère de l'Éducation nationale.

Données éducatives. Les sources exploitables pour analyser et quantifier les effets de la date de naissance sur les performances scolaires sont particulièrement riches et permettent de couvrir l'ensemble du parcours éducatif des élèves.

Nos analyses pour l'enseignement primaire reposent exclusivement sur les données issues du panel primaire construit en 1997 par la Direction de l'évaluation, de la performance et de la prospective (DEPP) pour le compte du ministère de l'Éducation nationale. Ce panel permet de suivre la scolarité primaire complète d'un échantillon représentatif de 9 632 élèves entrés en CP à la rentrée 1997. Le principal intérêt de cette source statistique réside dans le fait qu'elle contient les résultats obtenus par les élèves du panel aux épreuves d'évaluation des compétences qui ont eu lieu à la rentrée en CP (c'est-à-dire en octobre 1997) et à la rentrée de CE2 (c'est-à-dire en octobre 1999 pour les élèves n'ayant pas redoublé et en octobre 2000 pour les élèves qui ont redoublé une fois). Les épreuves proposées étaient composées de séries d'items donnant lieu au calcul de scores individuels sur une échelle de 0 à 100. Le panel primaire comporte également un grand nombre d'informations sur la scolarité pré-élémentaire des élèves, leurs caractéristiques individuelles (date de naissance, sexe, catégorie socio-professionnelle des parents, etc.) et celles de l'école fréquenté (statut public ou privé de l'établissement, classement en ZEP, nombre d'élèves dans la classe, etc.).

Les données utilisables pour étudier les effets de la date de naissance dans l'enseignement secondaire sont plus nombreuses que dans le primaire. Une première source statistique est le panel secondaire réalisé en 1995 par la DEPP selon la même méthodologie que le panel primaire. Il permet de suivre la scolarité secondaire complète

d'un échantillon représentatif d'environ 17 800 élèves scolarisés en sixième à la rentrée scolaire 1995 dans un établissement public ou privé de France métropolitaine. Les informations fournies par le panel permettent de reconstituer le parcours éducatif antérieur des élèves et de connaître leurs résultats en français et en mathématiques aux épreuves nationales d'évaluation qui ont lieu à l'entrée en sixième et qui sont notées sur une échelle de 0 à 100. Le panel contient également des informations sur les résultats obtenus par les élèves au cours de leur scolarité secondaire : au collège, on connaît les notes obtenues au contrôle continu (mais pas à l'examen terminal) du diplôme national du Brevet en français, en mathématiques et en première langue vivante ; au lycée, on dispose de l'ensemble des notes obtenues au Baccalauréat. Les informations fournies sur les caractéristiques des élèves sont semblables à celles du panel primaire. Malheureusement, le plan de sondage a été établi de telle manière que le panel secondaire ne contient aucun élève né en mars, juillet ou octobre¹², ce qui introduit une lacune dommageable lorsqu'on s'intéresse aux effets du mois de naissance.

Bien que les panels primaire et secondaire représentent une source d'information très riche sur les performances scolaires des élèves français, la taille relativement petite des échantillons limite mécaniquement la précision de nos estimations. C'est la raison pour laquelle nous avons choisi de compléter notre analyse pour le collège en utilisant les données de la base du Diplôme National du Brevet (DNB), disponible uniquement pour la session de 2004¹³. Le principal intérêt de ce fichier statistique tient à son caractère exhaustif : en 2004, il permet de connaître les notes obtenues par près de 813 000 élèves au contrôle continu et aux épreuves de l'examen terminal (français, mathématiques, histoire-géographie). Par ailleurs, la base du DNB fournit, en plus de la date de naissance des élèves, un certain nombre d'informations socio-démographiques sommaires qui incluent le sexe, la PCS du chef de famille ou encore la nationalité de l'élève.

L'influence de la date de naissance sur le parcours éducatif des individus ne se

¹²Plus précisément, l'échantillon a été construit de manière à obtenir un taux de sondage proche de 1/40^e. N'ont par conséquent été retenus dans le panel que les élèves nés le 17 d'un mois, à l'exclusion de ceux qui sont nés en mars, juillet ou octobre.

¹³Les données correspondant aux autres examens nationaux (en particulier le Baccalauréat) existent (base OCÉAN), mais n'ont pas encore été mises à la disposition des chercheurs.

limite pas nécessairement à son impact sur les résultats scolaires, mais peut également transiter par un redoublement et une orientation différenciés selon la position occupée par l'élève au sein de sa cohorte scolaire. Pour étudier ce double phénomène, nous avons choisi d'utiliser la base Scolarité, qui recense l'ensemble des élèves inscrits dans les établissements d'enseignement secondaire appartenant au secteur public ou au secteur privé sous contrat. Ce fichier administratif, réalisé à partir d'extractions des Bases élèves académiques (BEA) gérées au niveau des rectorats, couvre la quasi-totalité des élèves scolarisés dans le secteur public depuis la rentrée 1993 mais n'atteint un taux de couverture satisfaisant pour le secteur privé sous contrat que depuis la rentrée 2000¹⁴. Parmi les informations disponibles pour chaque élève figurent, outre la date de naissance, le sexe, la PCS du chef de famille, la nationalité, le département et la commune de résidence, l'établissement fréquenté et la formation suivie au cours de l'année n et de l'année $n - 1$. Si la base Scolarité ne peut être utilisée pour constituer un panel statistique d'élèves suivis tout au long de leur scolarité dans l'enseignement secondaire, elle permet néanmoins de calculer la proportion d'élèves en retard au sein de chaque cohorte scolaire ainsi que leur ventilation entre les différents niveaux de formation. Ainsi, bien que ne contenant aucune évaluation directe des performances individuelles, ce fichier constitue une source d'information précieuse lorsqu'on s'intéresse aux trajectoires scolaires des élèves.

Autres données. La persistance des effets de la date de naissance à l'âge adulte ne peut être étudiée qu'à l'aide de sources statistiques contenant des informations sur la situation professionnelle des individus. Dans cette perspective, nous avons utilisé conjointement les données de l'enquête Emploi et les fichiers du recensement de la population.

L'enquête Emploi, réalisée chaque année par l'Insee sur un échantillon représentatif d'environ 60 000 ménages, contient un grand nombre de renseignements relatifs au niveau de formation et à la situation professionnelle des personnes enquêtées. Nous avons choisi de travailler à partir des enquêtes effectuées entre 1990 et 2002, dans la mesure où les précédentes éditions ne permettent pas de connaître le salaire des

¹⁴En 2000, le taux de couverture pour les établissements privés sous contrat s'élève à 86% et atteint 98% en 2002.

individus autrement que par tranches. Les informations sur le niveau de formation incluent notamment l'âge de fin d'études et le diplôme le plus élevé obtenu. La situation professionnelle des individus peut quant à elle être étudiée sous des angles divers : situation d'emploi, profession, catégorie socio-professionnelle, secteur d'activité, etc.

Pour les besoins de notre étude, nous avons également utilisé les extractions au quart des recensements de la population réalisés en 1982 et en 1999. Bien que ces fichiers contiennent beaucoup moins d'informations sur le niveau de formation et la situation professionnelle des individus que l'enquête Emploi, ils ont l'avantage de couvrir le quart de la population française, ce qui permet d'atteindre une très grande précision dans les estimations. Grâce aux informations fournies sur le statut professionnel et les diplômes des individus, les fichiers du recensement permettent d'analyser de manière très fine l'éventuelle persistance à l'âge adulte des effets de la date de naissance. Ces données sont également utiles pour tester l'existence d'une possible corrélation entre la date de naissance et un certain nombre de caractéristiques socio-démographiques susceptibles d'influencer de manière autonome les trajectoires éducatives et professionnelles des individus.

4.3 Le mois de naissance : un instrument à manier avec précaution

Bien que la date de naissance constitue un instrument particulièrement populaire dans la littérature économique, son utilisation ne va pas nécessairement de soi. En effet, il est possible que cette variable soit corrélée avec certaines caractéristiques individuelles observables ou inobservables qui pourraient biaiser les estimations.

Pour tenter de mesurer le degré d'exogénéité de la date de naissance par rapport aux caractéristiques socio-démographiques des individus dans le cas français, nous avons utilisé les recensements de la population réalisés en 1982 et 1998. Nos analyses indiquent qu'il existe deux séries de facteurs expliquant que les individus nés à différentes périodes de l'année ne soient pas parfaitement comparables : le biais du mois de janvier et la différenciation sociologique du mouvement saisonnier des naissances. Nous montrons cependant que les biais engendrés par ces deux phénomènes

peuvent être assez aisément neutralisés lorsqu'on cherche à estimer les effets du mois de naissance sur les trajectoires scolaires et la vie professionnelle.

4.3.1 Un biais de mesure : la surreprésentation du mois de janvier

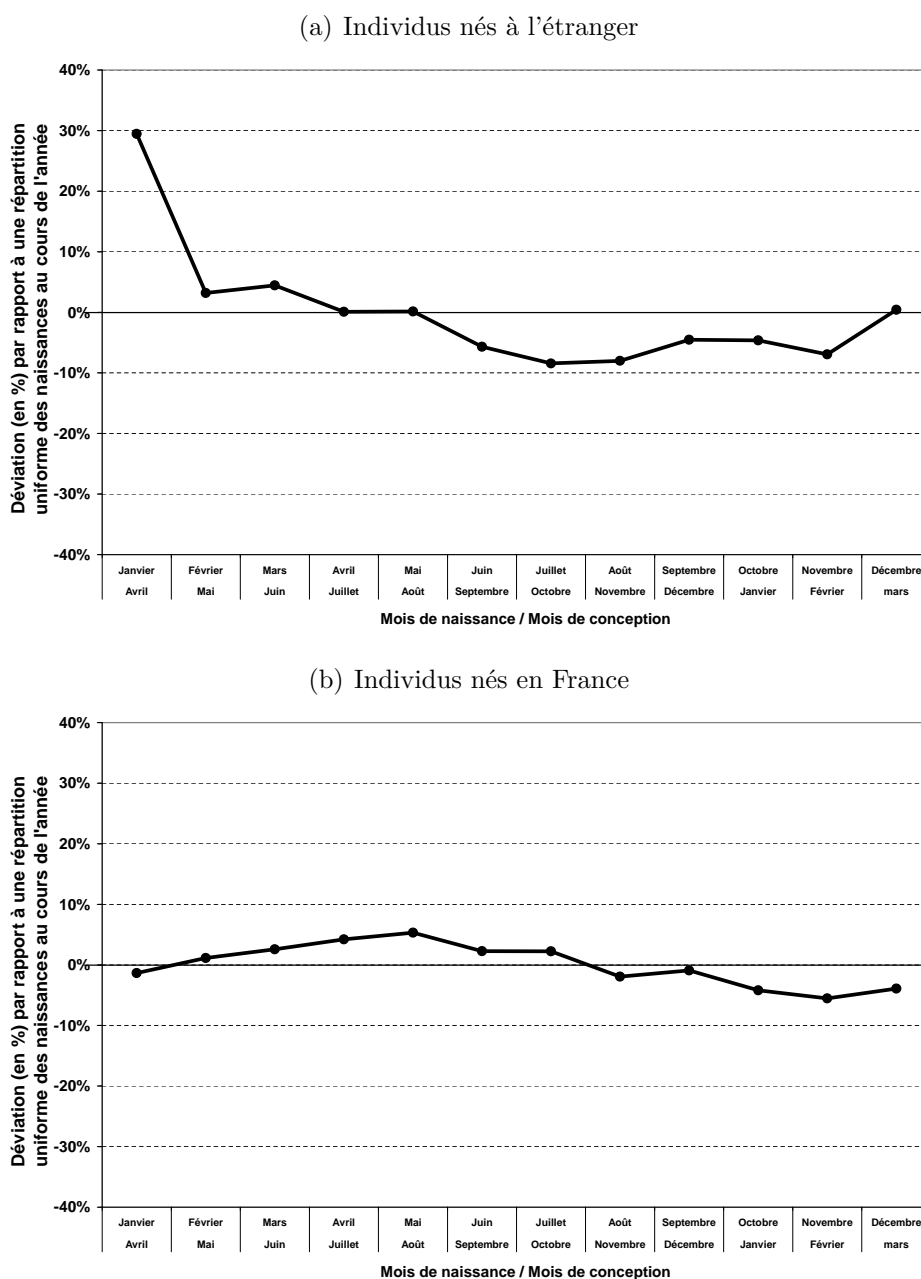
Une première source de corrélation entre date de naissance et caractéristiques socio-démographiques tient à la mesure-même de cette variable. Il est possible en effet que la date de naissance fournie par un individu ne corresponde pas à sa date de naissance réelle¹⁵. Cette erreur de mesure, qui porte essentiellement sur le jour et le mois de naissance, concerne notamment les personnes qui sont nées dans des pays dont les registres d'état civil sont imparfaitement tenus. Lorsque le jour et le mois de naissance d'un individu ne sont pas connus avec précision, une pratique courante consiste à leur attribuer une date de naissance fictive, qui correspond souvent au 1^{er} janvier.

Les données du recensement 1999 révèlent que ce phénomène est loin d'être négligeable en France. Lorsqu'on calcule la répartition des individus nés à l'étranger en fonction de leur mois de naissance (par rapport à une répartition théorique uniforme des naissances au cours de l'année), on constate une très forte surreprésentation du mois de janvier. Le graphique 4.1(a) de la figure 4.1 indique la déviation de la distribution des mois de naissance des individus nés à l'étranger entre le 1^{er} janvier 1900 et le 31 décembre 1998 par rapport à une distribution uniforme. Le « pic » du mois de janvier apparaît très clairement, avec un nombre de naissances supérieur de 30 % au taux attendu.

Cette surreprésentation du mois de janvier constitue une source de biais particulièrement importante lorsqu'on envisage d'utiliser la date de naissance comme instrument, dans la mesure où les individus nés à l'étranger ont en moyenne des caractéristiques socio-démographiques moins favorables que les individus nés en France. Un tel phénomène risque donc de gonfler artificiellement les effets du mois de naissance que l'on souhaite mesurer.

¹⁵Je tiens à remercier Roxanne Silberman, du centre Maurice Halbwachs, pour m'avoir signalé ce problème.

Figure 4.1 – *Déviaton de la distribution des mois de naissance des individus nés à l'étranger et des individus nés en France entre le 1^{er} janvier 1900 et le 31 décembre 1998 par rapport à une répartition uniforme des naissances au cours de l'année.*
Source : Recensement de la population 1999 (au 1/4).



Notes: Ce graphique présente la déviation de la distribution des mois de naissances des individus nés à l'étranger et des individus nés en France entre le 1^{er} janvier 1900 et le 31 décembre 1998, par rapport à la distribution que l'on aurait constatée si la répartition des naissances au cours de l'année était uniforme. En notant λ_m la déviation associée au mois m , on a : $\lambda_m = (J_m/365,2)(N_m/N)$ où J_m désigne le nombre de jours dans le mois m (28,2 pour le mois de février en raison des années bissextiles, 30 ou 31 pour les autres mois), N_m le nombre de naissances constatées au cours du mois m pour les individus nés à l'étranger (resp. nés en France) et N le nombre total d'individus nés à l'étranger (resp. nés en France).

Ce problème peut néanmoins être aisément contourné en restreignant l'échantillon d'étude aux personnes nées en France. Le graphique 4.1(b) montre en effet qu'on ne détecte pas dans le recensement de 1999 de prévalence anormale du mois de janvier dans cette partie de la population. Pour nous prémunir contre le biais de mesure du mois de janvier, nous n'utiliserons dans la suite de cette étude que des échantillons composés d'individus nés en France.

4.3.2 Une saisonnalité différenciée sociologiquement

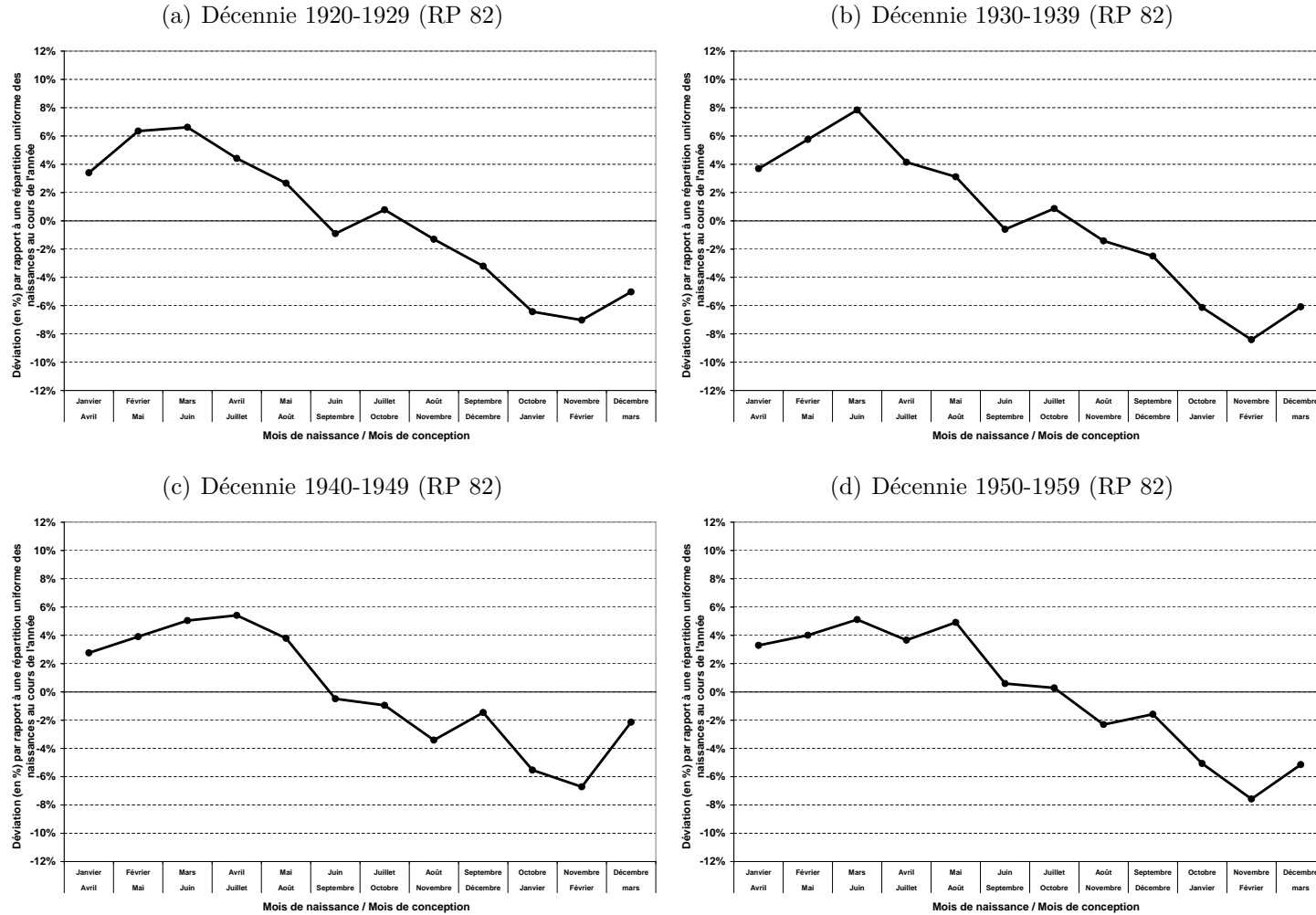
Au delà des erreurs de mesure portant sur la date de naissance des individus, la différenciation sociologique du mouvement saisonnier des naissances constitue un second phénomène susceptible d'engendrer une corrélation entre date de naissance et caractéristiques socio-démographiques.

Le fait que les naissances ne soient pas uniformément réparties au cours de l'année civile est un phénomène bien connu des démographes¹⁶. L'évolution du mouvement saisonnier des naissances au cours du XX^e siècle peut être suivie avec précision à partir des données des recensements de 1982 et de 1999. Il suffit pour cela de calculer pour chaque décennie l'écart entre la distribution observée des naissances, estimée à partir des individus présents dans le fichier, et une distribution théorique uniforme¹⁷. La figure 4.2 présente l'évolution de la saisonnalité des naissances en France pour chaque décennie comprise entre 1920 et 1980. On peut tirer deux leçons de ces différents graphiques. La première est que l'amplitude du mouvement saisonnier des naissances a toujours été relativement forte : la plus forte déviation du nombre de naissances constaté par rapport au nombre attendu s'établit en général à 6 % et atteint même 10 % au cours de la décennie 1970-1979. Le second enseignement que l'on peut tirer de ces graphiques est que l'allure globale du mouvement saisonnier des naissances a varié au cours du siècle : alors que jusqu'à la fin des années 1950, le nombre de naissances avait plutôt tendance à décroître au cours de l'année, les comportements

¹⁶Voir par exemple Dupaquier (1976), Saboulin (1978), Prioux (1988) et Reignier-Loilier (2004).

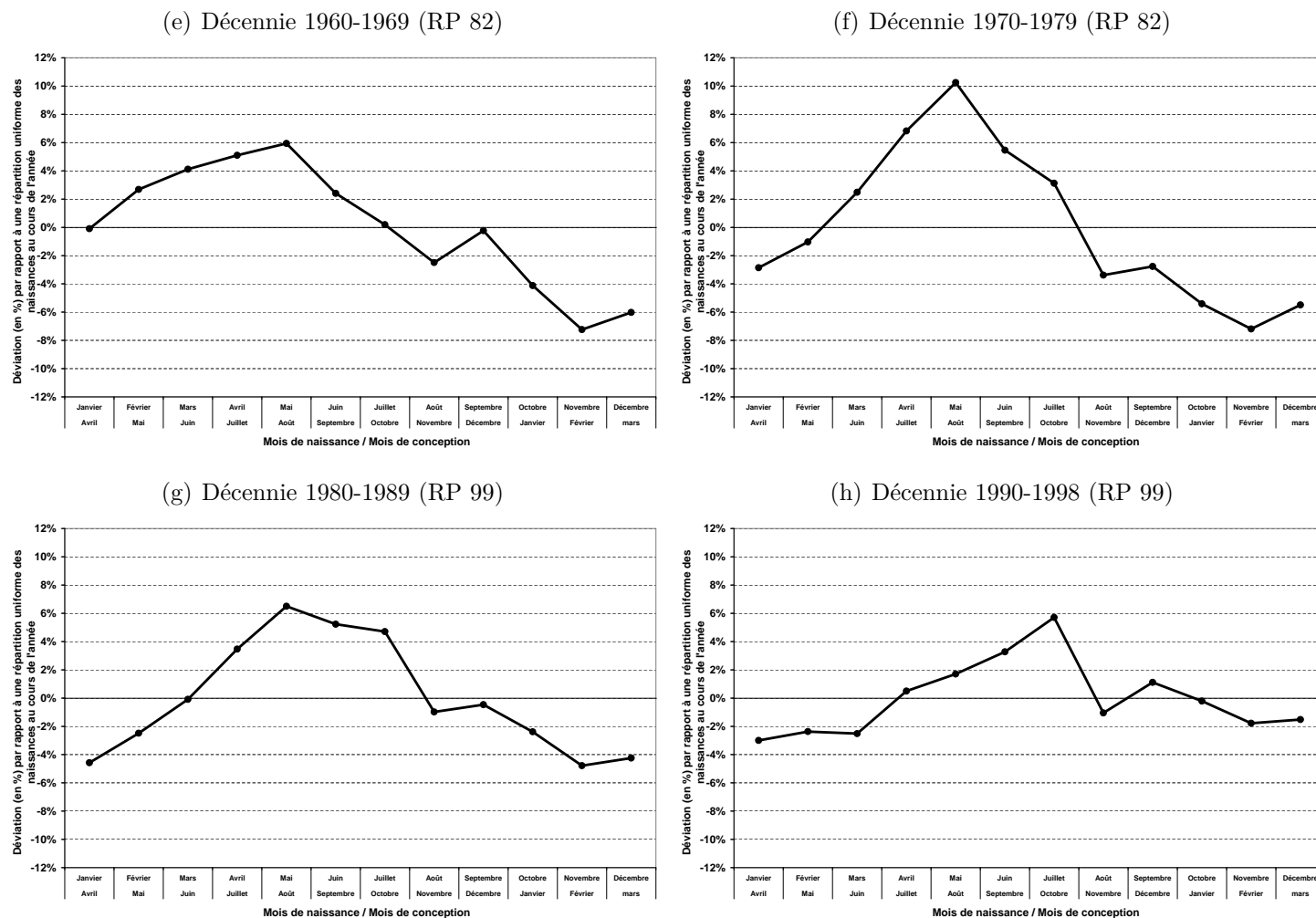
¹⁷Dans le cas des cohortes les plus anciennes, cette approche pourrait être éventuellement invalidée en présence d'une mortalité différenciée par mois de naissance (sur ce sujet, voir Doblhammer et Vaupel, 2001). Les biais potentiels associés à un tel phénomène étant relativement négligeables par rapport à l'amplitude du mouvement saisonnier des naissances, nous avons choisi de ne pas en tenir compte.

Figure 4.2 – *Mouvement saisonnier des naissances par décennie de naissance. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Décennies 1920-1959. Sources : Recensement de la population 1982 et 1999 (au 1/4).*



Notes: Chaque graphique correspond à une décennie de naissance. Un point représente la déviation du nombre de naissances constatées au cours d'un mois d'une décennie donnée par rapport au nombre que l'on aurait constaté si la répartition des naissances au cours de l'année était uniforme. En notant $\lambda_{m,d}$ la déviation associée au mois m pour la décennie d , on a : $\lambda_{m,d} = (J_m/365,2)(N_{m,d}/N_d)$ où J_m désigne le nombre de jours dans le mois m (28,2 pour le mois de février en raison des années bissextiles, 30 ou 31 pour les autres mois), $N_{m,d}$ le nombre de naissances constatées au cours du mois m de la décennie d et N_d le nombre total de naissances au cours de la décennie d .

Figure 4.2 – *Mouvement saisonnier des naissances par décennie de naissance. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Décennies 1960-1998. Sources : Recensement de la population 1982 et 1999 (au 1/4).*



Notes: Chaque graphique correspond à une décennie de naissance. Un point représente la déviation du nombre de naissances constatées au cours d'un mois d'une décennie donnée par rapport au nombre que l'on aurait constaté si la répartition des naissances au cours de l'année était uniforme. En notant $\lambda_{m,d}$ la déviation associée au mois m pour la décennie d , on a : $\lambda_{m,d} = (J_m/365, 2)(N_{m,d}/N_d)$ où J_m désigne le nombre de jours dans le mois m (28,2 pour le mois de février en raison des années bissextiles, 30 ou 31 pour les autres mois), $N_{m,d}$ le nombre de naissances constatées au cours du mois m de la décennie d et N_d le nombre total de naissances au cours de la décennie d .

démographiques se sont modifiés à partir des années 1960, à travers le déplacement du pic des naissances au deuxième trimestre (qui correspond à un pic des conceptions pendant l'été).

L'analyse des causes du mouvement saisonnier des naissances dépasse largement l'objet de cette étude et le champ de compétence de l'économiste, dans la mesure où elle fait intervenir des facteurs non seulement culturels, mais également biologiques, climatologiques, voire psychologiques¹⁸. Deux facteurs généralement invoqués pour expliquer la modification du profil de la saisonnalité des naissances au début des années 1960 méritent néanmoins être soulignés : la généralisation de la contraception, qui aurait accru la faculté de contrôler la date de naissance de son enfant, et la généralisation des congés payés, qui serait en partie responsable de l'apparition d'un pic des conceptions pendant l'été et d'un pic secondaire en décembre. Si ces deux phénomènes nous intéressent au premier chef, c'est qu'ils pourraient avoir contribué à accentuer la différenciation sociologique du mouvement saisonnier des naissances. En effet, la contraception permet aux couples de mieux accorder la date de conception et la date de naissance de leur enfant avec leurs préférences et leurs contraintes personnelles, qui peuvent dépendre en partie de leur catégorie socio-professionnelle. Les pratiques sociales relatives aux congés payés sont quant à elles suffisamment hétérogènes pour constituer un facteur important de différenciation de la saisonnalité des naissances d'un groupe social à l'autre.

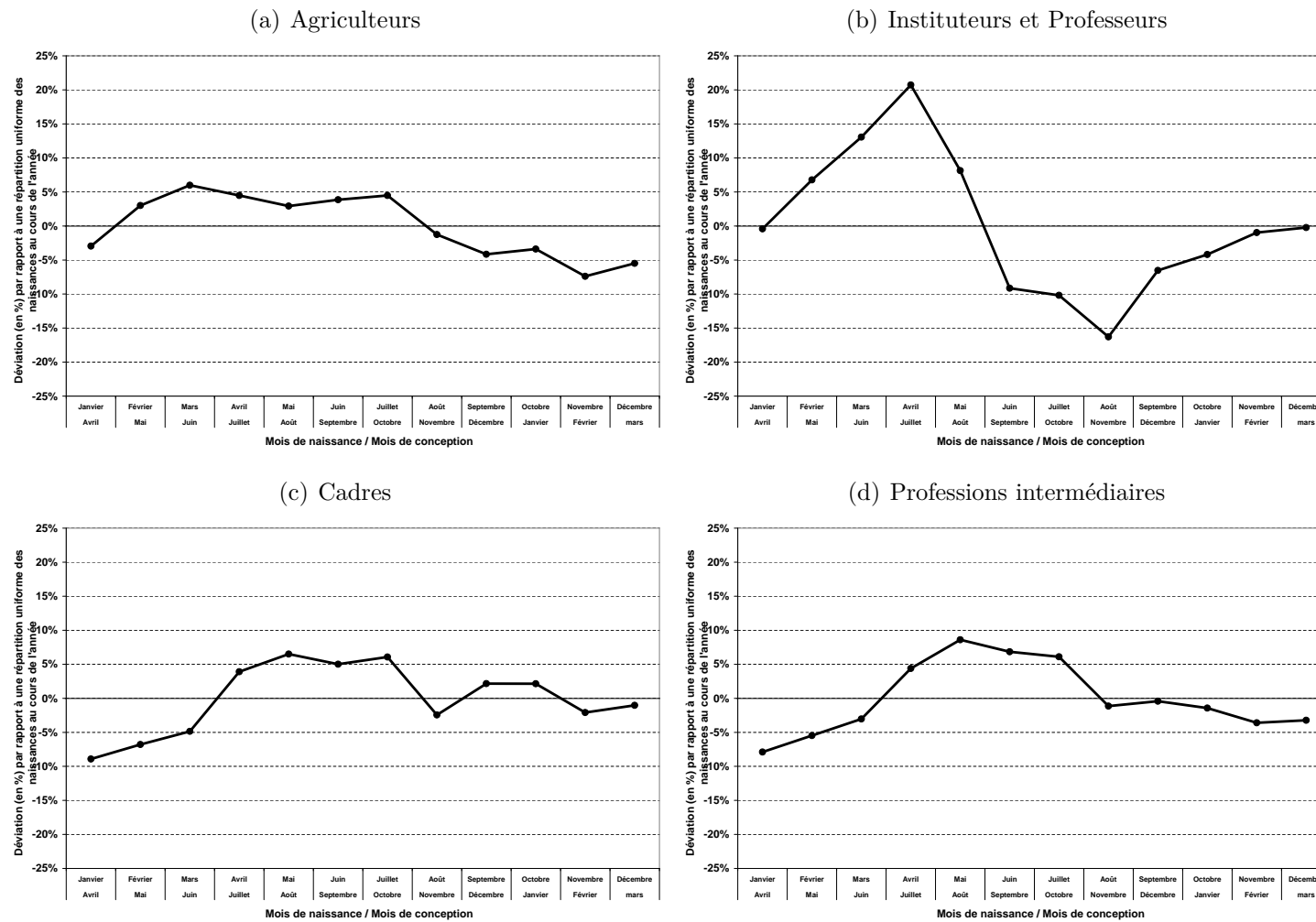
La composante sociologique de la date de naissance peut être étudiée pour les cohortes les plus récentes à partir des données du recensement de 1999. Ce fichier permet en effet, pour les enfants qui vivent chez leur parents, de connaître la catégorie socio-professionnelle du père et de la mère. Bien que la variable d'identification des membres d'un même ménage ne figure pas dans l'extraction au quart diffusée par l'Insee, le fichier du recensement de 1999 indique néanmoins, pour chaque individu, la catégorie socio-professionnelle et la profession (en 456 postes) du père et de la mère de famille. Afin de nous limiter aux enfants susceptibles de vivre encore chez leurs parents, nous avons restreint notre échantillon aux individus nés entre le 1^{er} janvier

¹⁸Sur ces différents aspects, on pourra se reporter entre autres aux études de Bumpass et al. (1978), Marini et Hodson (1981), Lutinié (1987), Sandron (1988), Leridon (1988), Oppenheimer (1988), Retherford et Sewell (1989), Bobak et Gjonca (2001) et Goldin et Katz (2002).

1980 et le 31 décembre 1998. En raison de la forte proportion d'enfants vivant dans des familles monoparentales, nous avons choisi d'étudier la saisonnalité différenciée des naissances à partir de la catégorie socio-professionnelle de la mère (la très grande majorité des enfants âgés de 0 à 18 ans présents dans l'échantillon vivant encore chez leur mère). La figure 4.3 indique comment la saisonnalité des naissances varie en fonction de la PCS de la mère (en 8 catégories : agriculteurs, instituteurs et professeurs, cadres, professions libérales, indépendants, employés, ouvriers, chômeurs et inactifs). Les différents graphiques montrent que la différenciation sociologique du mouvement saisonnier des naissances est relativement prononcée. Quatre modèles peuvent être schématiquement distingués : les agriculteurs avec des naissances plus nombreuses entre avril et juillet (qui correspondent à des conceptions entre mai et octobre) que pendant le reste de l'année ; les cadres et professions intermédiaires, avec un pic de conceptions pendant l'été et, dans le cas des cadres, un pic secondaire pendant les vacances d'hiver ; les indépendants, ouvriers, employés, chômeurs et inactifs, avec une saisonnalité des naissances moins marquée et un pic des naissances en juillet (conceptions en octobre). Les instituteurs et professeurs, enfin, qui constituent un groupe tout à fait à part, dans la mesure où ils se distinguent des autres catégories socio-professionnelles par l'amplitude et le profil du mouvement saisonnier des naissances. Dans leur cas, le pic des naissances s'établit en effet au mois d'avril, qui enregistre un nombre de naissances supérieur de 20 % au taux normal attendu. Bien que les causes du profil particulier des naissances des instituteurs et professeurs ne soient pas *a priori* aisées à identifier, on peut penser que le fonctionnement des congés maternité n'y est pas totalement étranger. Une naissance au mois d'avril permet en effet aux enseignantes de bénéficier de 6 semaines de congé prénatal et de 10 semaines de congé post-natal prenant fin au début des vacances d'été. Un tel choix dans la date de naissance de leur enfant donne aux femmes qui exercent le métier d'institutrice ou de professeur la possibilité d'interrompre leur activité professionnelle tout en étant rémunérées pendant une période totale de 6 mois.

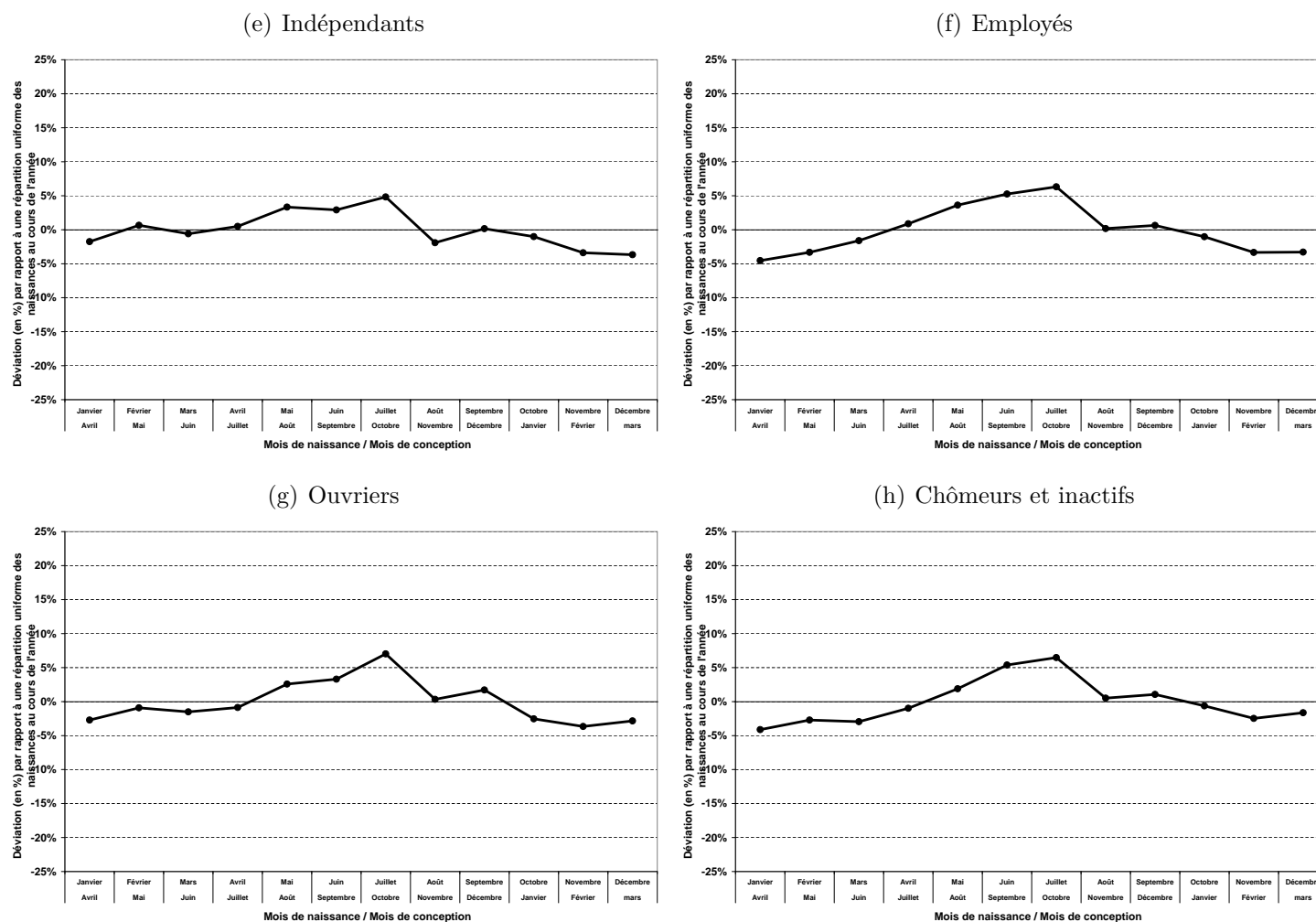
Les conséquences de cette différenciation sociologique du mouvement saisonnier des naissances du point de vue de l'étude des effets économiques de la date de naissance sont loin d'être anecdotiques. Le fait que les individus nés à différentes périodes

Figure 4.3 – *Mouvement saisonnier des naissances en fonction de la catégorie socio-professionnelle de la mère de famille. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Naissances survenues entre le 1^{er} janvier 1980 et le 31 décembre 1998. Source : Recensement de la population 1999 (au 1/4).*



Notes: Chaque graphique représente la saisonnalité des naissances en fonction de la catégorie socio-professionnelle de la mère pour les naissances intervenues entre le 1^{er} janvier 1980 et le 31 décembre 1998. Un point représente, pour cette catégorie socio-professionnelle, la déviation du nombre de naissances constatées au cours d'un mois donné par rapport au nombre que l'on aurait constaté si la répartition des naissances au cours de l'année était uniforme. En notant $\lambda_{m,p}$ la déviation associée au mois m pour les femmes appartenant à la catégorie socio-professionnelle p , on a : $\lambda_{m,p} = (J_m/365, 2)(N_{m,p}/N_p)$ où J_m désigne le nombre de jours dans le mois m (28,2 pour le mois de février en raison des années bissextiles, 30 ou 31 pour les autres mois), $N_{m,p}$ le nombre de naissances constatées au cours du mois m pour les femmes appartenant à la catégorie socio-professionnelle p et N_p le nombre total de naissances survenues pour ces femmes au cours de la période 1980-1998.

Figure 4.3 – *Mouvement saisonnier des naissances en fonction de la catégorie socio-professionnelle de la mère de famille. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Naissances survenues entre le 1^{er} janvier 1980 et le 31 décembre 1998. Source : Recensement de la population 1999 (au 1/4).*



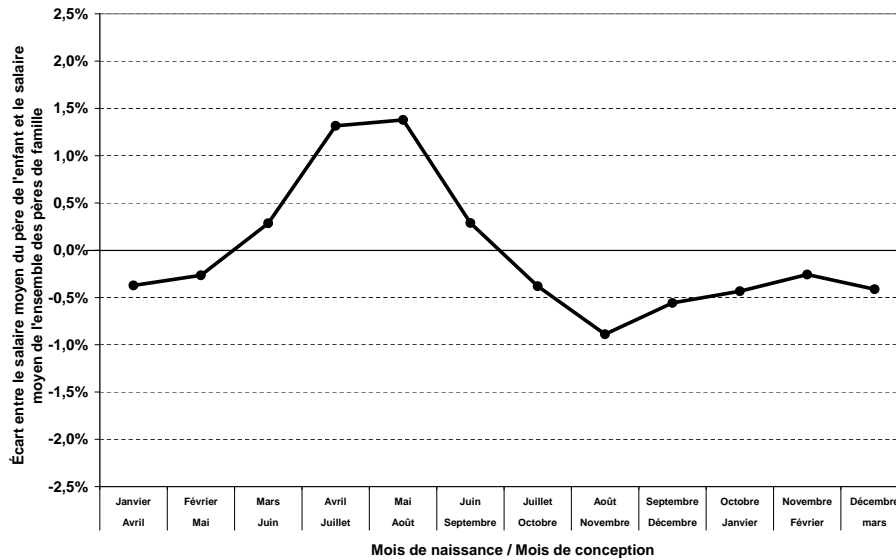
Notes: Chaque graphique représente la saisonnalité des naissances en fonction de la catégorie socio-professionnelle de la mère pour les naissances intervenues entre le 1^{er} janvier 1980 et le 31 décembre 1998. Un point représente, pour cette catégorie socio-professionnelle, la déviation du nombre de naissances constatées au cours d'un mois donné par rapport au nombre que l'on aurait constaté si la répartition des naissances au cours de l'année était uniforme. En notant $\lambda_{m,p}$ la déviation associée au mois m pour les femmes appartenant à la catégorie socio-professionnelle p , on a : $\lambda_{m,p} = (J_m/365, 2)(N_{m,p}/N_p)$ où J_m désigne le nombre de jours dans le mois m (28,2 pour le mois de février en raison des années bissextiles, 30 ou 31 pour les autres mois), $N_{m,p}$ le nombre de naissances constatées au cours du mois m pour les femmes appartenant à la catégorie socio-professionnelle p et N_p le nombre total de naissances survenues pour ces femmes au cours de la période 1980-1998.

de l'année ne soient pas issus de milieux sociaux parfaitement comparables risque en effet de créer une corrélation artificielle entre la date de naissance et les caractéristiques individuelles qui sont influencées par celles des parents. Pour se faire une idée de l'ampleur des biais engendrés par la composante sociale de la date de naissance, une approche naturelle consiste à mesurer l'ampleur des variations du revenu parental moyen d'un mois de naissance à un autre. Pour y parvenir, nous avons calculé séparément le salaire moyen du père et le salaire moyen de la mère parmi les ménages dont les enfants sont nés entre 1960 et 1998, en combinant les données du recensement et des enquêtes Emploi. On utilise le recensement de 1982 pour identifier les parents dont les enfants sont nés entre le 1^{er} janvier 1960 et le 31 décembre 1979 et le recensement de 1999 pour identifier les parents dont les enfants sont nés entre le 1^{er} janvier 1980 et le 31 décembre 1998, dans la mesure où la grande majorité des enfants âgés de moins de 22 ans vivent encore chez leurs parents. Les salaires des parents sont imputés indirectement en utilisant le salaire moyen de leur profession (en 456 postes) tel qu'il peut être calculé à partir des fichiers de l'enquête Emploi (1990-2002), séparément pour les hommes et les femmes. La principale limite de cette méthodologie est qu'elle restreint le champ d'étude aux enfants appartenant à des familles dont le père ou la mère exerce une activité salariée (ce qui exclut notamment les indépendants, les professions libérales, les inactifs et les chômeurs). Par ailleurs, l'imputation salariale est nécessairement approximative et ne tient pas compte de l'évolution de la hiérarchie salariale des professions intervenue depuis le début des années 1980. Il reste que cette procédure permet de visualiser de manière relativement précise la corrélation entre revenu parental et mois de naissance depuis le début des années 1960.

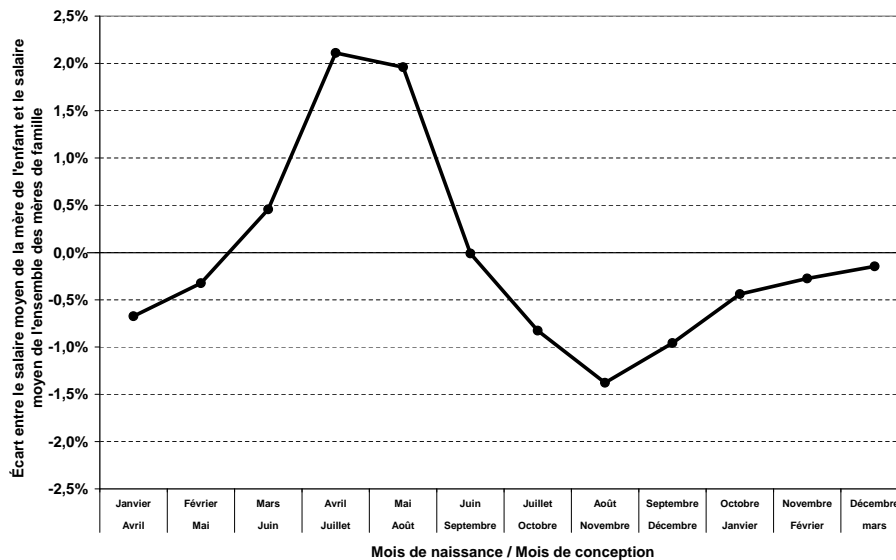
La figure 4.4 indique le salaire moyen (imputé) du père et de la mère des enfants nés au cours des décennies 1960 et 1970. La figure 4.5 correspond quant à elle aux naissances intervenues au cours des décennies 1980 et 1990. Ces différents graphiques indiquent clairement que le mois de naissance n'est pas totalement exogène par rapport au revenu des parents. On constate en effet que quelle que soit la période considérée et le sexe des parents, la courbe du salaire moyen du père et de la mère en fonction du mois de naissance des enfants a un profil sinusoïdal, avec un pic pour les naissances du mois d'avril-mai (conceptions en juillet-août) et un creux pour les

Figure 4.4 – Salaire moyen (imputé) du père et de la mère des enfants nés entre le 1^{er} janvier 1960 et le 31 décembre 1979 en fonction du mois de naissance de l'enfant, rapporté au salaire moyen de l'ensemble des pères et mères de famille. Champ : enfants dont le père ou la mère est salarié. Sources : Recensement de la Population 1982 (au 1/4) et Enquête Emploi (1990-2002).

(a) Salaire du père (imputé) / salaire moyen



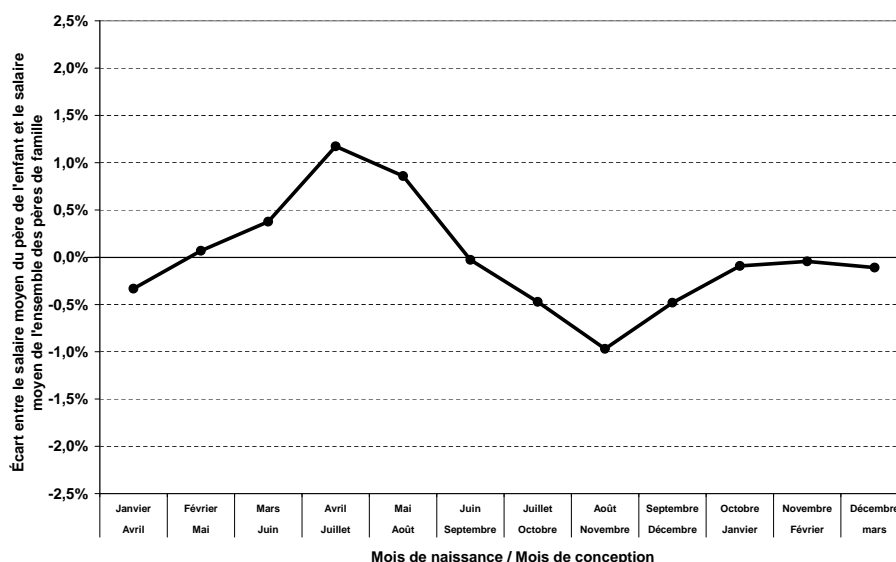
(b) Salaire de la mère (imputé) / salaire moyen



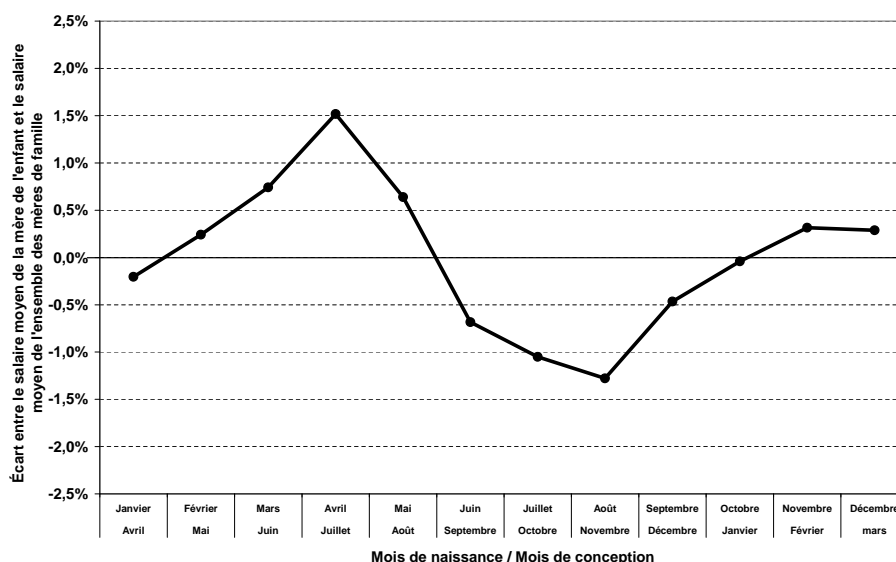
Notes: Chaque graphique est réalisé à partir des données du recensement de 1982 et de l'enquête Emploi (1990-2002). Chaque point correspond à l'écart entre le salaire moyen des pères (ou des mères) des enfants nés au cours de ce mois et le salaire moyen des pères (ou des mères) de l'ensemble des enfants nés entre le 1^{er} janvier 1960 et le 31 décembre 1979. Les salaires des pères et mères de famille sont imputés à partir de leur profession (en 456 postes) à partir du salaire moyen de chaque profession dans l'enquête Emploi (calculé séparément pour les hommes et les femmes). Ces graphiques ne couvrent par conséquent que les enfants dont le père ou la mère est salarié.

Figure 4.5 – Salaire moyen (imputé) du père et de la mère des enfants nés entre le 1^{er} janvier 1980 et le 31 décembre 1998 en fonction du mois de naissance de l'enfant, rapporté au salaire moyen de l'ensemble des pères et mères de famille. Champ : enfants dont le père ou la mère est salarié. Sources : Recensement de la Population 1999 (au 1/4) et Enquête Emploi (1990-2002).

(a) Salaire du père (imputé) / salaire moyen



(b) Salaire de la mère (imputé) / salaire moyen



Notes: Chaque graphique est réalisé à partir des données du recensement de 1999 et de l'enquête Emploi (1990-2002). Chaque point correspond à l'écart entre le salaire moyen des pères (ou des mères) des enfants nés au cours de ce mois et le salaire moyen des père (ou des mères) de l'ensemble des enfants nés entre le 1^{er} janvier 1980 et le 31 décembre 1998. Les salaire des pères et mères de famille sont imputés à partir de leur profession (en 456 postes) à partir du salaire moyen de chaque profession dans l'enquête Emploi (calculé séparément pour les hommes et les femmes). Ces graphiques ne couvrent par conséquent que les enfants dont le père ou la mère est salarié.

naissances du mois d'août (conceptions en novembre). L'allure de cette courbe peut être assez aisément reliée à la différenciation sociologique de la saisonnalité des naissances identifiée précédemment : il est probable par exemple que le salaire moyen plus élevé des parents des enfants conçus pendant l'été provienne du fait que les cadres et professions intermédiaires font davantage d'enfants pendant l'été et moins d'enfants pendant l'automne que les employés ou les ouvriers.

4.3.3 Peut-on limiter les biais induits par la composante sociologique du mois de naissance ?

S'il apparaît clairement que la date de naissance ne peut être considérée comme indépendante des caractéristiques socio-démographiques des individus, il nous semble néanmoins que les biais potentiellement associés à l'utilisation de cette variable présentent un certain nombre de propriétés qui permettent de les neutraliser, au moins en partie.

Un premier élément à relever concerne la forme de l'association entre le mois de naissance et le revenu des parents. L'analyse développée plus haut à propos de l'évolution du salaire moyen des parents en fonction du mois de naissance des enfants semble indiquer que cette relation est fondamentalement non linéaire et qu'elle a une structure plus ou moins symétrique par rapport à une date qui correspondrait au milieu de l'année civile (fin juin-début juillet). Ce type de profil ne paraît pas susceptible d'introduire des biais importants si les effets que l'on cherche à mesurer évoluent linéairement avec le mois de naissance. Dans ce cas en effet, la composante sociologique du mois de naissance se manifestera simplement par une déviation symétrique des valeurs moyennes de la variable d'intérêt (note moyenne aux évaluations, taux de redoublement, etc.) par rapport à la tendance linéaire que l'on cherche à mesurer.

La seconde observation que l'on peut formuler à la lumière des graphiques des figures 4.4 et 4.5 est que l'association entre mois de naissance et revenu des parents semble relativement faible pour les mois extrêmes (janvier et décembre), dans la mesure où le salaire moyen du père ou de la mère des enfants nés au cours de ces périodes de l'année ne diffère pas sensiblement du salaire moyen. Ce phénomène suggère que la composante sociologique du mois de naissance peut être en grande

partie neutralisée si on se limite aux individus nés en janvier ou en décembre de la même année. Lorsque les données le permettront, nous utiliserons cette approche pour vérifier la robustesse de nos résultats.

Enfin, on notera que s'il existe incontestablement une différenciation sociologique de la date de naissance, son ampleur reste relativement limitée : les graphiques des figures 4.4 et 4.5 indiquent en effet que l'amplitude des déviations du salaire du père ou de la mère des enfants nés au cours d'un mois donné par rapport au salaire moyen ne dépasse pas 3,5 %. Un tel ordre de grandeur ne paraît susceptible d'entraver l'estimation des effets du mois de naissance que si ces derniers sont modestes. Or les analyses développées dans la suite de cette étude indiquent que tel n'est pas le cas, en particulier lorsqu'on s'intéresse à l'impact sur les performances scolaires.

4.4 L'effet du mois de naissance sur les performances scolaires

Dans cette section, nous cherchons à mesurer l'impact du mois de naissance des élèves sur leurs performances tout au long de leur scolarité primaire et secondaire.

Dans le contexte éducatif français, les effets que nous mesurons pourraient *a priori* résulter de la combinaison de plusieurs facteurs : les durées inégales de scolarisation pré-élémentaire, les différences de maturité liées aux différences d'âge au moment du passage des évaluations (effet d'âge absolu), les effets d'âge relatif ou encore les effets du redoublement.

Bien que ces différents facteurs ne puissent pas être aisément dissociés, un certain nombre d'éléments empiriques nous conduisent à considérer l'âge absolu au moment des évaluations comme le facteur déterminant des écarts de performances scolaires constatés selon le mois de naissance des élèves.

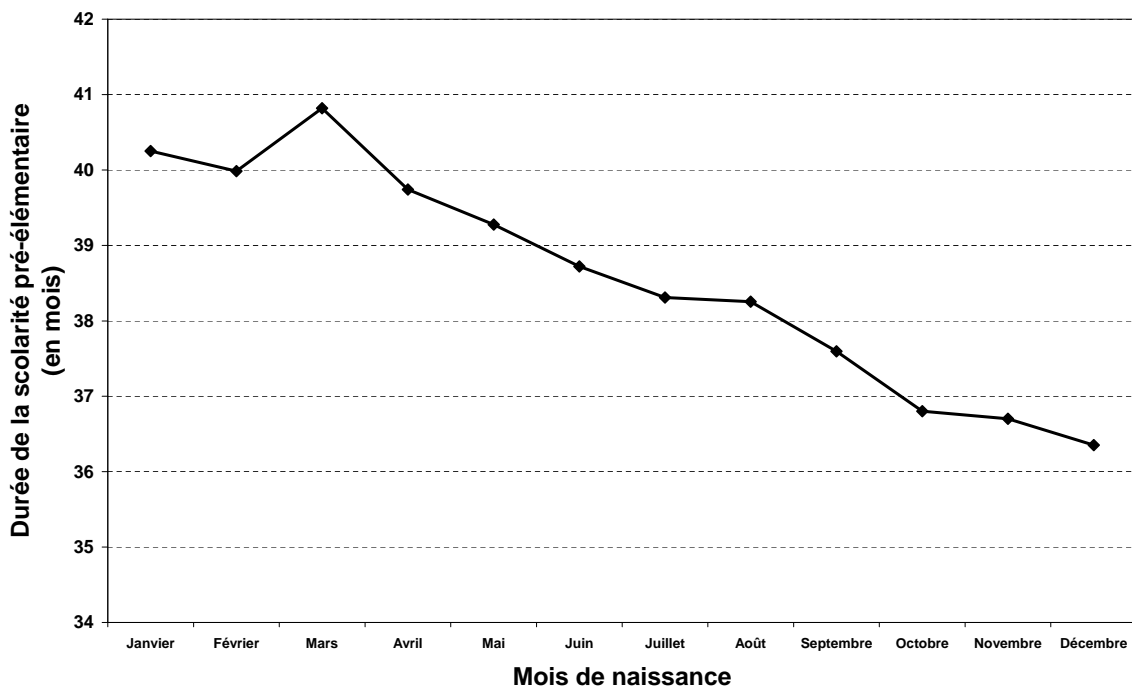
4.4.1 Mois de naissance et durée de la scolarité pré-élémentaire

Comme expliqué dans l'introduction, les règles qui régissent en France l'âge d'entrée à l'école maternelle ont tendance à privilégier les enfants les plus âgés au sein de leur cohorte au détriment des plus jeunes, si bien que ceux qui sont nés en fin

d'année ont en moyenne une scolarité pré-élémentaire plus courte que ceux qui sont nés en début d'année. Ce phénomène tient pour l'essentiel au fait qu'en raison des contraintes de capacité, les directeurs et directrices d'écoles maternelles donnent généralement, parmi les postulants à une scolarisation dès l'âge de 2 ans, la priorité aux enfants nés en début d'année par rapport aux enfants nés en fin d'année.

Les données du panel primaire de l'Éducation nationale comportent un certain nombre d'informations sur la scolarisation pré-élémentaire des élèves qui sont entrés en CP en 1997. Le graphique 4.6 indique le nombre moyen de mois que les élèves de l'échantillon ont passé à l'école maternelle en fonction de leur mois de naissance. À l'exception d'un pic local pour les élèves nés en mars (qui s'explique peut-être en partie par la surreprésentation des enfants d'enseignants pour ce mois de l'année), la tendance est nettement décroissante : les élèves de CP nés en décembre ont passé en moyenne 4 mois de moins à l'école maternelle que les élèves nés en janvier.

Figure 4.6 – *Durée moyenne (en mois) de la scolarité pré-élémentaire des élèves entrés en CP en 1997, en fonction de leur mois de naissance. Source : Panel primaire de l'Éducation nationale (1997).*



Notes: Ce graphique est construit à partir des informations sur la scolarité pré-élémentaire des élèves du panel primaire de l'Éducation nationale qui sont entrés en CP en 1997. Chaque point représente la durée moyenne (en mois) que les élèves ont passée à l'école maternelle en fonction de leur mois de naissance.

Au vu de ce graphique, il pourrait sembler naturel de penser que la plus faible durée de scolarité pré-élémentaire pénalise les performances des élèves nés à la fin de l'année indépendamment des effets de maturité liés à leur plus jeune âge au moment des évaluations. Les estimations réalisées dans l'article précédemment cité de Goux et Maurin (2008) semblent indiquer que ce n'est pas le cas : en utilisant le fait que le taux d'admission des élèves nés pendant l'été varie fortement d'une région à l'autre en raison de l'hétérogénéité des capacités d'accueil des écoles maternelles, ces auteurs parviennent à estimer séparément l'effet lié à l'âge des élèves et l'effet lié à la durée de la scolarité pré-élémentaire et trouvent que le second n'est pas significativement différent de zéro.

Dans la suite de cette étude, nous nous appuyerons sur les résultats de Goux et Maurin pour considérer que les écarts de performances scolaires constatés entre les élèves en fonction de leur mois de naissance ne proviennent pas de leur durée inégale de scolarisation pré-élémentaire.

4.4.2 Comment estimer l'impact du mois de naissance sur les performances scolaires ?

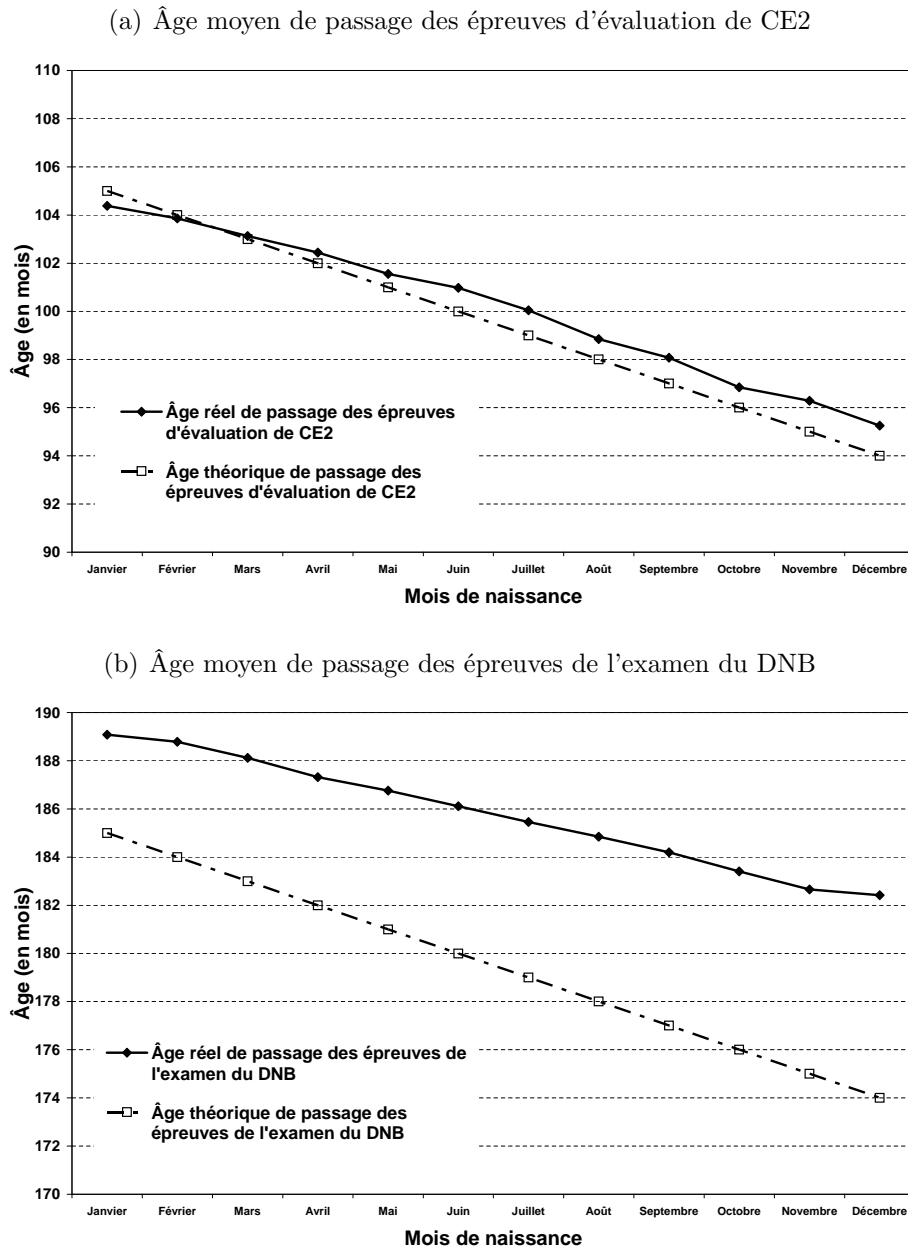
L'estimation des effets de la date de naissance sur les performances scolaires des élèves français soulève un certain nombre de difficultés méthodologiques qui tiennent aux biais de sélection et d'endogénéité que nous avons évoqués dans l'introduction.

4.4.2.1 Différences d'âge théorique et différences d'âge réelles

Dans la mesure où, en France, les cohortes scolaires regroupent tous les élèves nés au cours d'une même année civile, une première approche pour estimer l'effet de la date de naissance sur les performances scolaires pourrait consister à régresser les scores des élèves aux épreuves d'évaluation sur leurs mois de naissance. Les résultats obtenus en appliquant cette méthode sont néanmoins difficiles à interpréter, car les coefficients estimés risquent de sous-évaluer les effets de l'âge sur les résultats scolaires.

La courbe représentée en trait plein dans le graphique 4.7(a) de la figure 4.7 indique l'âge absolu (en mois) auquel les élèves du panel primaire de l'Éducation nationale (1997) ont passé les épreuves d'évaluation de CE2, qui ont lieu au mois

Figure 4.7 – Âge théorique et âge réel (en mois) de passage des épreuves d'évaluation de CE2 et de l'examen du Diplôme national du Brevet. Sources : Panel primaire de l'Éducation nationale (1997) et Base DNB (2004).



Notes: L'âge théorique de passage des épreuves d'évaluation de CE2 correspond à l'âge qu'auraient eu les élèves en octobre de l'année 1992 s'il étaient entrés en CP à l'âge normal (donc nés en 1990) et n'avaient pas redoublé ou sauté de classe par la suite. L'âge réel de passage des épreuves est calculé à partir de la date de passage des évaluations (octobre 1999 ou 2000) en fonction du mois et de l'année de naissance des élèves. L'âge théorique de passage des épreuves d'évaluation du DNB correspond à l'âge qu'auraient eu les élèves en juin de l'année 2004 s'il étaient entrés en CP à l'âge normal (donc nés en 1989) et n'avaient pas redoublé ou sauté de classe par la suite. L'âge réel de passage des épreuves est calculé à partir de la date de passage des épreuves (juin 2004) en fonction du mois et de l'année de naissance des élèves.

d'octobre. La droite représentée en trait discontinu indique quant à elle l'âge théorique auquel ces élèves auraient dû passer l'épreuve s'ils étaient entrés en CP à l'âge normal et n'avaient pas redoublé depuis. On constate que le tracé de ces deux courbes diffère sensiblement et que l'écart a tendance à croître avec le mois de naissance. Deux séries de facteurs peuvent être invoqués pour expliquer ce phénomène.

La raison pour laquelle on observe que l'âge moyen auquel les élèves nés en début d'année passent les épreuves d'évaluation de CE2 est inférieur à l'âge théorique (le phénomène inverse peut être observé pour les élève nés en fin d'année) tient au fait qu'un petit nombre d'élèves entrent à l'école primaire à un âge qui ne correspond pas à l'âge « normal ». Les données du panel primaire de l'Éducation nationale indiquent par exemple que si 97 % des élèves de l'échantillon entrés en classe de CP sont nés en 1991, 1,5 % d'entre eux sont nés en 1989 ou en 1990 et les 1,5 % restants sont nés en 1992. Or les élèves en retard ou en avance par rapport à l'âge théorique ne sont pas répartis uniformément dans l'année : les élèves scolarisés avec un an d'avance sont généralement nés en début d'année, alors que ceux qui sont scolarisés avec un ou deux ans de retard sont plutôt nés en fin d'année. Les causes d'un tel phénomène sont à rechercher du côté de la volonté de certains parents, soit d'éviter à leur enfant né en fin d'année de se retrouver parmi les plus jeunes de sa classe (cas des enfants scolarisés avec un an ou deux ans de retard par rapport à l'âge théorique), soit au contraire, pour les enfants nés en début d'année, de leur permettre d'entrer plus rapidement à l'école primaire sans pour autant être trop jeunes par rapport à leurs camarades de classe (cas des enfants scolarisés avec un an d'avance).

Si l'entrée précoce ou tardive des élèves en CP explique une petite partie de l'écart observé entre âge réel et âge théorique de passage des épreuves de CE2, la principale cause de cette différence tient au fait que les élèves n'ont pas la même probabilité de redoubler ou de sauter une classe en fonction de leur mois de naissance. L'impact du mois de naissance sur la probabilité de redoublement sera analysée en détail dans la section 4.5. Constatons simplement que ce phénomène a pour effet de réduire l'écart d'âge relatif entre les élèves nés en fin d'année et leurs camarades nés en début d'année. Le phénomène inverse se produit pour les élèves qui sautent une classe : les données du panel primaire indiquent que sur les 1,5 % d'élèves de l'échantillon qui

ont sauté une ou deux classes au cours de leur scolarité primaire, la plupart sont nés en début d'année.

L'ensemble de ces facteurs tend à réduire l'écart d'âge relatif entre les élèves au moment où ils passent les épreuves d'évaluation. On risque par conséquent de sous-estimer les effets d'âge liés à la date de naissance en régressant les scores obtenus à ces épreuves sur le mois de naissance. Il faut noter que la sous-estimation des effets de l'âge risque d'être d'autant plus importante que ces effets sont évalués à une étape plus avancée du parcours éducatif, dans la mesure où les écarts d'âge relatifs ont tendance à se réduire à mesure que l'influence du mois de naissance sur la probabilité de redoublement fait sentir ses effets. La figure 4.7(b) révèle par exemple qu'au moment où les élèves passent les épreuves du Brevet, l'écart d'âge relatif réel entre les élèves nés en décembre et les élèves nés en janvier n'est plus en moyenne que de 7 mois environ, alors que l'écart d'âge théorique est de 11 mois.

4.4.2.2 Stratégie d'estimation

Dans ces conditions, il semblerait plus légitime de mesurer l'effet de l'âge sur les performances scolaires en régressant les scores individuels non plus sur le mois de naissance, mais sur l'âge absolu des élèves à la date des évaluations. Cette approche souffre néanmoins d'un biais d'endogénéité, dans la mesure où l'âge absolu de passage d'une épreuve ne peut être considéré comme une variable indépendante des capacités de réussite scolaire, qui ne sont pas directement observables. Les élèves les plus vieux au moment du passage d'une épreuve d'évaluation sont en effet des redoublants et sont donc *a priori* moins doués à l'école que les autres élèves. Inversement, les élèves les plus jeunes au moment du passage de l'épreuve ont souvent sauté une classe, ce qui semble indiquer qu'ils manifestent de meilleures dispositions pour les études que leurs camarades. La combinaison de ces deux phénomènes induit mécaniquement une forte corrélation négative entre l'âge absolu de passage de l'épreuve et le terme d'erreur de la régression, ce qui conduit à sous-estimer l'effet causal de l'âge sur les performances scolaires, voire à en inverser le signe.

Pour résoudre la difficulté liée à l'endogénéité de la variable d'âge absolu dans l'équation de score individuel, nous nous inspirons de la méthodologie proposée par

Bedard et Dhuey (2006), qui consiste à instrumenter l'âge absolu d'un élève au moment des évaluations par la différence d'âge qui le sépare des individus nés en décembre de son année de naissance. Cette transformation affine du mois de naissance¹⁹ vaut par exemple 3 pour les élèves nés au mois de septembre, indiquant qu'ils ont trois mois de plus que les individus les plus jeunes de leur cohorte scolaire. La validité de cet instrument repose sur deux hypothèses : d'une part, il faut qu'il soit corrélé avec l'âge absolu auquel les élèves passent les épreuves d'évaluation ; d'autre part, il faut qu'il ne détermine pas les performances scolaires autrement qu'en influençant l'âge absolu de passage des épreuves. La première condition est facilement vérifiée, dans la mesure où l'âge absolu des élèves est fortement corrélé à leur position au sein de leur cohorte scolaire (cette corrélation est de 1 pour les élèves qui passent l'épreuve à l'âge normal). La seconde condition nécessite une discussion plus approfondie. Pour que notre instrument soit valide, il faut que le mois de naissance n'ait pas d'influence propre sur les performances scolaires, une fois pris en compte l'âge absolu de l'élève.

Le principal facteur susceptible d'invalider cette condition d'exclusion a été mentionné dans la partie précédente : la composante sociologique du mois de naissance implique que la position d'un élève au sein de sa cohorte scolaire est corrélée avec un certain nombre de caractéristiques parentales qui agissent directement sur les performances scolaires. Pour les raisons mentionnés plus tôt et qui tiennent essentiellement à la forme de l'association entre date de naissance et caractéristiques socio-démographiques, il ne nous semble pas que ce phénomène soit de nature à biaiser fortement nos estimations. Nous proposerons néanmoins deux méthodes pour tester la sensibilité de nos résultats à la prise en compte de ce phénomène : d'une part, en incluant dans nos régressions une série de contrôles socio-démographiques, où figurent notamment la catégorie socio-professionnelle du père et de la mère ; d'autre part, en restreignant nos échantillons aux élèves nés aux mois de janvier et de décembre, dans la mesure où la composante sociologique de ces mois de naissance paraît relativement modeste (voir plus haut).

Il faut également noter que les coefficients estimés au moyen de cette approche peuvent être contaminés par les autres effets du mois de naissance sur les perfor-

¹⁹égale à 12 - mois de naissance.

mances scolaires : l'effet de l'âge d'entrée à l'école, l'effet d'âge relatif et l'effet du redoublement. Les deux premiers effets auront tendance à augmenter l'impact apparent de l'âge de passage des évaluations (prédit par le mois de naissance) sur les performances scolaires, dans la mesure où ils ne peuvent que renforcer le handicap scolaire subi par les élèves nés en fin d'année. Le redoublement est quant à lui susceptible de jouer dans les deux sens : s'il a un impact négatif sur les performances scolaires, il aura tendance à accroître l'effet apparent de l'âge ; si au contraire il a un impact positif sur les performances scolaires, il tendra à en atténuer l'effet.

Les travaux empiriques mentionnés dans l'introduction suggèrent que l'impact de l'âge d'entrée à l'école et l'effet d'âge relatif sont relativement faibles par rapport à l'effet de l'âge absolu de passage des évaluations (Leuven et al., 2006 ; Fredriksson et Ockert, 2006 ; Crawford et al., 2007). Il nous paraît donc raisonnable de considérer que la contamination de nos estimations par les effets d'âge relatif et d'âge d'entrée à l'école est d'une ampleur limitée. L'impact du redoublement sur les performances scolaires ne peuvent pas en revanche être écartés *a priori*, sauf en ce qui concerne les élèves entrant en CP qui, par définition, ne sont pas concernés par ces effets. Nous verrons plus loin que la comparaison de nos estimations avec celles obtenues par Crawford et al. (2007) sur données anglaises semble indiquer que la part du redoublement dans les effets de l'âge sur les performances scolaires est relativement marginale.

4.4.3 Résultats des estimations

Les données scolaires dont nous disposons nous permettent de mesurer l'impact de l'âge absolu des élèves sur leurs performances scolaires tout au long de leur scolarité primaire et secondaire, du CP jusqu'aux épreuves anticipées de français du Baccalauréat. Afin de faciliter la comparaison des coefficients estimés, nous avons normalisé l'ensemble des scores utilisés de manière à ce que la moyenne de chacun soit égale à 0 et l'écart-type à 1.

Les mesures de performances disponibles. Nos estimations pour l'école primaire sont réalisées à partir du panel primaire de l'Éducation nationale, qui comporte

deux séries d'évaluations : à la rentrée de l'année de CP et à la rentrée de l'année de CE2. Les évaluations de CE2 sont disponibles pour les élèves qui n'ont pas redoublé (session de 1999) et pour ceux qui ont redoublé une fois (session de 2000). Les performances individuelles aux épreuves d'évaluation des acquis à l'entrée en CP peuvent être mesurées au moyen d'un score global qui synthétise les résultats obtenus dans les cinq grands domaines de connaissances évalués²⁰. Les résultats aux épreuves nationales d'évaluation des acquis en CE2 sont pour leur part mesurés sous la forme d'un score global en mathématiques et d'un score global en français.

Au collège, nous disposons de mesures d'évaluation des performances en sixième et en troisième. Le fichier du panel secondaire de l'Éducation nationale permet de connaître le score obtenu aux épreuves de mathématiques et de français des élèves de l'échantillon scolarisés en sixième à la rentrée 1995. Les résultats obtenus aux épreuves du Diplôme national du Brevet (DNB) peuvent être mesurés à partir de deux sources distinctes. On connaît d'une part les notes obtenues par les élèves du panel secondaire aux épreuves du contrôle continu en mathématiques, en français et en première langue vivante. Les résultats sont disponibles pour trois sessions : 1999 (élèves qui n'ont pas redoublé), 2000 (élèves qui ont redoublé une fois au collège), 2001 (élèves qui ont redoublé deux fois au collège). Pour éviter qu'un même élève apparaisse plusieurs fois (cas des élèves qui ont redoublé leur troisième), on ne retient que les résultats obtenus lors du premier passage de l'examen du Brevet. Le fichier du Diplôme national du Brevet permet quant à lui de connaître les résultats aux épreuves du contrôle terminal (mathématiques, français et histoire-géographie) de tous les élèves qui ont passé l'examen en 2004, ainsi que la note moyenne qu'ils ont obtenue à l'ensemble des épreuves du DNB²¹. Le caractère exhaustif de ce fichier permet d'atteindre une grande précision dans l'estimation des effets de l'âge sur les performances scolaires lorsque les élèves terminent le premier cycle de l'enseignement secondaire.

²⁰Les épreuves couvraient les domaines suivants : connaissances générales ; compétences verbales et familiarité avec l'écrit ; compétences logiques et familiarité avec le nombre ; concepts liés au temps et à l'espace ; comportements socio-cognitifs.

²¹La note moyenne à l'examen du Brevet est calculée en appliquant un coefficient de 6 à l'ensemble des trois épreuves du contrôle terminal (mathématiques, français, histoire-géographie) et un coefficient de 9, 10 ou 11 (selon la série) aux épreuves du contrôle continu.

Au-delà du collège, les effets sont plus délicats à mesurer. On ne dispose en effet que des résultats obtenus aux épreuves anticipées de français du Baccalauréat (qui ont lieu à la fin de l'année de première) par les élèves du panel secondaire qui ont poursuivi leurs études dans un lycée d'enseignement général et technologique et qui n'ont pas redoublé entre la sixième et la première. Cet échantillon se caractérise donc par un biais de sélection important, à la différence de ceux utilisés pour mesurer l'impact de l'âge sur les performances scolaires en primaire et au collège. Parce que les élèves nés en fin d'année ont plus souvent tendance à redoubler au collège et à être orientés dans l'enseignement professionnel (deux phénomènes que nous analyserons plus en détail dans la section 4.5), il est probable que les estimations obtenues en utilisant les notes au Bac français auront tendance à sous-estimer légèrement l'effet de l'âge de passage des épreuves sur les performances scolaires.

Estimations. Le tableau 4.1 présente les résultats de nos estimations. Les coefficients de la colonne 1 correspondent à l'approche naïve qui consiste à régresser les scores individuels sur l'âge absolu des élèves (en mois). Le fait qu'ils soient tous de signe négatif pour les évaluations réalisées après la classe de CP indique clairement que le biais d'endogénéité de l'âge contamine fortement les estimations : les performances plus faibles des élèves nés en fin d'année s'expliquent pour l'essentiel par leur probabilité plus forte d'avoir redoublé au cours de leur scolarité. Les colonnes suivantes indiquent les résultats des estimations réalisées en instrumentant l'âge absolu de passage des épreuves par la différence d'âge séparant l'élève des individus nés en décembre de son année de naissance. Les coefficients de la régression de première étape et de la forme réduite sont reportés dans les colonnes 2 et 3. La colonne 4 présente quant à elle les coefficients de la régression de seconde étape, qui mesure l'impact causal de l'âge sur les performances scolaires. Les coefficients de la colonne 2 indiquent que l'âge auquel les élèves passent les évaluations est fortement corrélé avec leur différence d'âge par rapport aux individus nés en décembre de leur année de naissance. Cette corrélation, égale à 0,9 en CP, a logiquement tendance à décroître à mesure que l'on avance dans la scolarité et que les effets du redoublement sur l'âge de passage des épreuves se font sentir avec plus d'intensité. Le fait qu'elle soit plus élevée

Table 4.1 – *Impact de l'âge absolu (en mois) sur les notes obtenues aux évaluations de CP, de CE2, de sixième, au contrôle continu et à l'examen terminal du Diplôme national du Brevet, et à l'épreuve anticipée de français du Baccalauréat. Instrument : différence d'âge séparant l'élève des individus nés en décembre de son année de naissance. Sources : Panels primaire (1997) et secondaire (1995) de l'Éducation nationale et Base DNB (2004).*

	Source	(Variable dép. :	(Variable dép. :	(Variable dép. :	(Variable dépendante :	(Variable dépendante :	(Élèves nés en janvier	Nb d'obs. Tous les élèves / élèves nés en janvier et en décembre
		score)	âge en mois)	score)	âge absolu en mois)	âge absolu en mois)	ou en décembre)	
Instrument : différence d'âge avec les individus nés en décembre de l'année de naissance								
		MCO	Première étape	Forme réduite	IV	IV	IV	
		(1)	(2)	(3)	(4)	(5)	(6)	
<u>ÉCOLE PRIMAIRE</u>								
CP : score global	PPEN 1997	0.029*** (0.003)	0.908*** (0.006)	0.056*** (0.003)	0.062*** (0.003)	0.060*** (0.003)	0.059*** (0.005)	9 342 / 1 537
CE2 : score en mathématiques	PPEN 1997	-0.011*** (0.003)	0.851*** (0.011)	0.042*** (0.003)	0.050*** (0.004)	0.048*** (0.004)	0.046*** (0.006)	7 653 / 1 257
CE2 : score en français	PPEN 1997	-0.013*** (0.003)	0.851*** (0.011)	0.031*** (0.003)	0.036*** (0.004)	0.036*** (0.004)	0.037*** (0.006)	7 653 / 1 257
<u>COLLÈGE</u>								
Sixième : mathématiques	PSEN 1995	-0.056*** (0.001)	0.724*** (0.014)	0.019*** (0.002)	0.026*** (0.003)	0.023*** (0.003)	0.020*** (0.004)	16 790 / 3 601
Sixième : français	PSEN 1995	-0.056*** (0.001)	0.724*** (0.014)	0.022*** (0.002)	0.029*** (0.003)	0.027*** (0.003)	0.026*** (0.004)	16 790 / 3 601
Troisième : DNB mathématiques (CC)	PSEN 1995	-0.051*** (0.001)	0.724*** (0.018)	0.002 (0.003)	0.002 (0.004)	0.003 (0.003)	-0.002 (0.005)	10 894 / 2 318
Troisième : DNB français (CC)	PSEN 1995	-0.049*** (0.001)	0.724*** (0.018)	0.011*** (0.003)	0.015*** (0.004)	0.016*** (0.003)	0.012** (0.005)	10 894 / 2 318
Troisième : DNB langue vivante 1 (CC)	PSEN 1995	-0.051*** (0.001)	0.724*** (0.018)	0.007*** (0.003)	0.010*** (0.004)	0.011*** (0.003)	0.012** (0.005)	10 894 / 2 318
Troisième : DNB mathématiques (EF)	DNB 2004	-0.023*** (0.000)	0.647*** (0.004)	0.004*** (0.000)	0.006*** (0.001)	0.006*** (0.000)	0.005*** (0.001)	781 391 / 127 822
Troisième : DNB français (EF)	DNB 2004	-0.017*** (0.000)	0.647*** (0.004)	0.009*** (0.000)	0.014*** (0.001)	0.014*** (0.000)	0.015*** (0.001)	781 391 / 127 822
Troisième : DNB hist. et géo. (EF)	DNB 2004	-0.020*** (0.000)	0.647*** (0.004)	0.008*** (0.000)	0.013*** (0.001)	0.014*** (0.001)	0.014*** (0.001)	781 391 / 127 822
Troisième : DNB moyenne générale	DNB 2004	-0.026*** (0.000)	0.647*** (0.004)	0.010*** (0.000)	0.016*** (0.000)	0.015*** (0.000)	0.014*** (0.001)	781 391 / 127 822
<u>LYCÉE</u>								
Première : Bac français (écrit)	PSEN 1995	-0.019*** (0.003)	0.775*** (0.016)	0.006* (0.004)	0.008* (0.005)	0.009* (0.005)	0.011 (0.005)	5 460 / 1 117
Première : Bac français (oral)	PSEN 1995	-0.018*** (0.003)	0.775*** (0.016)	-0.005 (0.004)	-0.006 (0.005)	-0.004 (0.005)	-0.005 (0.008)	5 460 / 1 117
Contrôles socio-démographiques et scolaires		Non	Non	Non	Non	Oui	Oui	

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. PPEN 1997 : panel primaire de l'Éducation nationale (1997). PSEN 1995 : panel secondaire de l'Éducation nationale (1995). DNB 2004 : base du Diplôme national du Brevet 2004. CC : contrôle continu. EF : examen final. Chaque coefficient provient d'une régression séparée. Les coefficients de la colonne 1 sont issus de la régression naïve du score sur l'âge (en mois) de l'élève. Les coefficients des colonnes 3 à 5 correspondent aux différentes étapes (première étape, forme-réduite et seconde étape) de la régression du score de l'élève sur son âge (en mois), lorsque l'âge est instrumenté par le nombre de mois qui le séparent des individus nés en décembre de son année de naissance. Les coefficient de la colonne 5 sont issus des régressions de seconde étape lorsqu'on contrôle pour les caractéristiques socio-démographiques de l'élève et le type d'établissement scolaire fréquenté (cf. section 4.4). Les coefficients de la colonne 6 sont estimés à partir de l'échantillon des élèves nés uniquement en janvier et en décembre.

pour les élèves passant les épreuves du Bac français que pour les candidats au BEPC peut s'interpréter comme la conséquence des biais de sélection mentionnés plus haut au sujet des élèves du panel secondaire qui n'ont pas redoublé entre la sixième et la première.

Plusieurs enseignements peuvent être tirés des estimations obtenues en appliquant la méthode des variables instrumentales pour mesurer l'impact causal de l'âge sur les performances scolaires (colonne 4 du tableau). Le premier constat est que les effets en jeu sont loin d'être négligeables : au CP, chaque mois supplémentaire dans l'âge des élèves se traduit par une augmentation des performances égale à 6 % de l'écart-type de la distribution des scores, ce qui signifie que les élèves nés en décembre subissent une pénalité égale à près de 70 % d'un écart-type par rapport à leurs camarades nés en janvier²², ce qui est considérable. La seconde observation que l'on peut formuler à la lumière de ces estimations est que si les effets de l'âge tendent à diminuer à mesure que les élèves progressent dans leur scolarité, ils continuent néanmoins à se manifester dans l'enseignement secondaire : par rapport au CP, la pénalité subie par les élèves nés en décembre est divisée par 2 au CE2, puis encore par deux au moment où les élèves passent les épreuves de l'examen du Brevet. Elle reste néanmoins relativement élevée, puisqu'elle représente encore entre 10 et 20 % d'un écart-type selon les épreuves. Nos estimations semblent également indiquer que les effets de l'âge sur les performances scolaires persistent au lycée : bien qu'étant probablement sous-estimé, le coefficient demeure en effet significatif au seuil de 10 % pour les notes obtenues par les élèves aux épreuves écrites de français du Baccalauréat et la pénalité associée à une différence d'âge de 11 mois est égale à environ 10 % d'un écart-type. Le troisième enseignement de ces régressions est que les effets de l'âge ne s'exercent avec la même intensité dans toutes les matières : il semblerait en particulier que la pénalité subie par les élèves nés en décembre diminue plus rapidement en mathématiques qu'en français ou en histoire-géographie. Quel que soit l'échantillon considéré (panel secondaire ou base du DNB), on constate que l'effet de l'âge sur le score obtenu par les élèves aux épreuves de mathématiques est significativement plus faible que dans les autres matières, alors qu'il est à peine inférieur en CE2.

²²Cette pénalité est calculée en multipliant par 11 les coefficients reportés dans la colonne 4 du tableau.

Tests de robustesse. Afin de vérifier la sensibilité de nos résultats à la prise en compte de la composante sociologique du mois de naissance, nous avons réalisé deux séries de tests. La première vérification passe par l'inclusion dans les régressions d'un certain nombre de contrôles socio-démographiques, où figurent notamment la PCS des parents²³. Nous avons ajouté à cette liste de variables les informations dont on dispose sur les caractéristiques de l'établissement fréquenté par l'élève²⁴. Les coefficients obtenus, qui sont reportés dans la colonne 5 du tableau, sont très proches de ceux issus des régressions dépourvues de contrôles socio-démographiques ou scolaires. La stabilité de nos estimations semble indiquer que la différenciation sociale imputable au mois de naissance n'explique pas nos résultats.

Le second test proposé consiste à restreindre les échantillons utilisés aux élèves nés en janvier ou en décembre uniquement, dans la mesure où il semble que ces deux mois de naissance ne soient pas particulièrement affectés par le biais sociologique mis en évidence dans la section 4.3. Là encore, les coefficients sont à peine modifiés et semblent confirmer la robustesse des conclusions tirées des précédentes estimations.

Analyse par sous-groupes d'élèves. Les effets de l'âge sur les performances scolaires pénalisent-ils tous les élèves de la même manière? Pour répondre à cette question, nous avons procédé à des analyses par sous-échantillons, dont les résultats sont présentés dans le tableau 4.2.

Dans les deux premières colonnes du tableau sont reportés les coefficients de la régression de seconde étape pour les garçons et les filles séparément, en incluant un certain nombre de contrôles socio-démographiques et scolaires. Ces résultats indiquent que quel que soit l'âge considéré, les effets mesurés ne diffèrent pas sensiblement selon le sexe des élèves.

Tel ne semble pas être le cas lorsqu'on compare les effets de l'âge en fonction du

²³Pour les échantillons construits à partir du PPEN 1997 et du PSEN 1995, la liste des variables de contrôle socio-démographiques inclut le sexe de l'élève, la profession du père et celle de la mère. Pour l'échantillon issu de la base du DNB 2004, la liste des contrôles inclut le sexe de l'élève et le fait qu'il soit handicapé ou non.

²⁴Pour les échantillons construits à partir du PPEN 1997 et du PSEN 1995, ces variables incluent la tranche d'unité urbaine de la commune de l'établissement scolaire, son statut (public ou privé), son appartenance ou non à une ZEP, le nombre d'élèves dans la classe fréquentée par l'élève ainsi que l'année de passage de l'évaluation. Pour l'échantillon construit à partir de la base du DNB 2004, la liste des contrôles inclut le statut (public ou privé) de l'établissement et le nombre d'élèves inscrits aux épreuves de l'examen du DNB au sein de l'établissement.

Table 4.2 – *Analyse par sous-groupes d'élèves de l'impact de l'âge absolu (en mois) sur les notes obtenues aux évaluations de CP, de CE2, de sixième, au contrôle continu et à l'examen terminal du Diplôme national du Brevet, et à l'épreuve anticipée de français du Baccalauréat. Instrument : différence d'âge séparant un élève de ses camarades nés en décembre de son année de naissance. Sources : Panels primaire (1997) et secondaire (1995) de l'Éducation nationale et Base DNB (2004).*

	Source	(Variable dépendante : score aux évaluations)			
		Garçons	Filles	PCS favorisées	PCS défav.
		IV (1)	IV (2)	IV (3)	IV (4)
<u>ÉCOLE PRIMAIRE</u>					
CP : score global	PPEN 1997	0.060*** (0.004)	0.062*** (0.004)	0.054*** (0.004)	0.063*** (0.004)
Nombre d'observations		4 861	4 668	4 008	5 334
CE2 : score en mathématiques	PPEN 1997	0.052*** (0.006)	0.046*** (0.005)	0.037*** (0.005)	0.057*** (0.005)
CE2 : score en français	PPEN 1997	0.039*** (0.005)	0.035*** (0.005)	0.030*** (0.005)	0.040*** (0.005)
Nombre d'observations		3 862	3 791	3 367	4 286
<u>COLLÈGE</u>					
Sixième : mathématiques	PSEN 1995	0.020*** (0.004)	0.025*** (0.004)	0.021*** (0.004)	0.025*** (0.004)
Sixième : français	PSEN 1995	0.026*** (0.004)	0.027*** (0.004)	0.030*** (0.004)	0.026*** (0.004)
Nombre d'observations		8 662	8 128	6 884	9 906
Troisième : DNB mathématiques (CC)	PSEN 1995	0.003 (0.005)	0.003 (0.005)	0.006 (0.005)	-0.000 (0.005)
Troisième : DNB français (CC)	PSEN 1995	0.014*** (0.005)	0.017*** (0.004)	0.019*** (0.005)	0.013*** (0.005)
Troisième : DNB langue vivante 1 (CC)	PSEN 1995	0.009* (0.005)	0.011** (0.005)	0.012** (0.005)	0.009* (0.005)
Nombre d'observations		5 142	5 752	5 224	5 670
Troisième : DNB mathématiques (EF)	DNB 2004	0.005*** (0.001)	0.006*** (0.001)	0.008*** (0.001)	0.004*** (0.001)
Troisième : DNB français (EF)	DNB 2004	0.013*** (0.001)	0.015*** (0.001)	0.017*** (0.001)	0.012*** (0.001)
Troisième : DNB hist. et géo. (EF)	DNB 2004	0.012*** (0.001)	0.013*** (0.001)	0.014*** (0.001)	0.011*** (0.001)
Troisième : DNB moyenne générale	DNB 2004	0.015*** (0.001)	0.016*** (0.001)	0.017*** (0.001)	0.014*** (0.001)
Nombre d'observations		384 263	397 128	331 115	450 276
<u>LYCÉE</u>					
Première : Bac français (écrit)	PSEN 1995	0.012 (0.008)	0.010* (0.006)	0.014** (0.007)	0.006 (0.007)
Première : Bac français (oral)	PSEN 1995	-0.001 (0.008)	-0.006 (0.006)	0.003 (0.007)	-0.012 (0.007)
Nombre d'observations		2 186	3 274	3 212	2 248
Contrôles socio-démographiques et scolaires		Oui	Oui	Oui	Oui

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. PPEN 1997 : panel primaire de l'Éducation nationale (1997). PSEN 1995 : panel secondaire de l'Éducation nationale (1995). DNB 2004 : base du Diplôme national du Brevet 2004. CC : contrôle continu. EF : examen final. Chaque coefficient provient d'une régression séparée par sous-groupe d'élèves. Les coefficients reportés proviennent des régressions de seconde étape du score individuel sur l'âge de l'élève (en mois) au moment de l'évaluation, lorsque l'âge est instrumenté par le nombre de mois qui le séparent des individus nés en décembre de son année de naissance. Les individus sont classés en « PCS favorisée » ou en « PCS défavorisée » en fonction de la PCS du chef de famille (cf. section 4.4). Toutes les régressions incluent des contrôles socio-démographiques pour l'élève et le type d'établissement scolaire fréquenté.

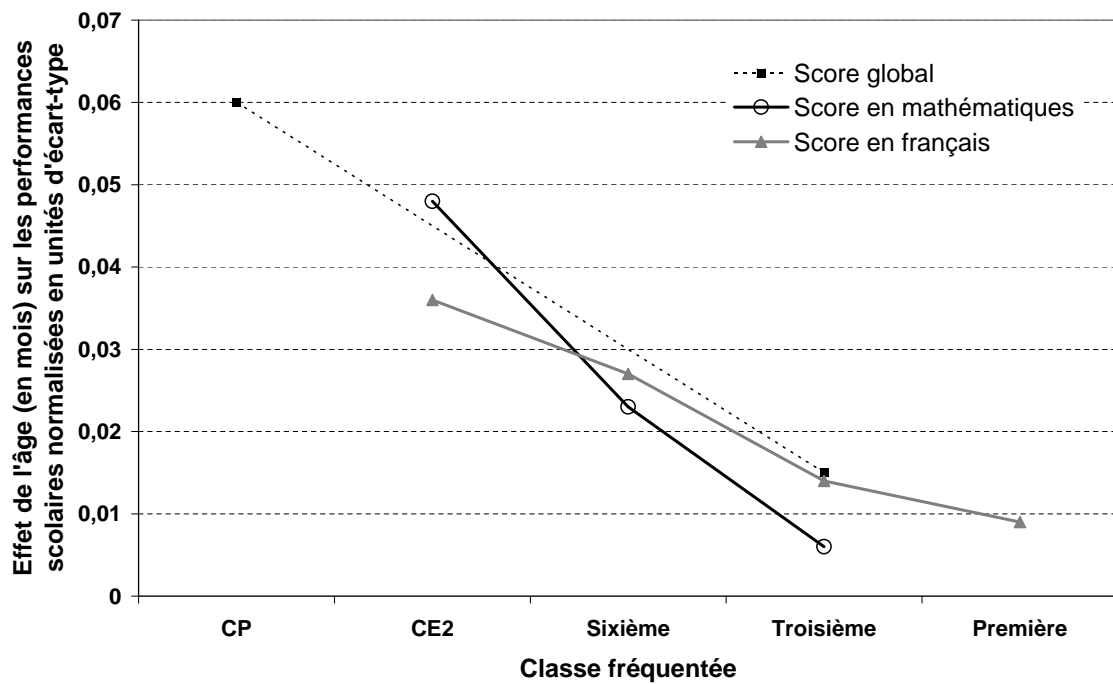
milieu social d'origine. Les coefficients reportés dans les colonnes 3 et 4 du tableau correspondent aux estimations réalisées en divisant grossièrement l'échantillon en fonction de la PCS du chef de famille : d'un côté, les élèves issus de milieux plutôt favorisés²⁵ ; de l'autre, les élèves issus de milieux plutôt défavorisés²⁶. La comparaison des coefficients indique qu'à l'école primaire, les effets d'âge sont plus prononcés pour les seconds que pour les premiers : l'impact de l'âge de passage des examens sur le score des élèves défavorisés est supérieur de 20 à 50 % à l'impact pour les élèves favorisés et l'écart est significatif au seuil de 5 % pour les scores obtenus aux épreuves de mathématiques en CE2. Plus tard dans la scolarité, l'écart tend à s'inverser, puisqu'on observe que l'impact de l'âge sur les résultats au DNB est significativement plus élevé pour les élèves défavorisés que pour les élèves favorisés. Il paraît cependant difficile d'interpréter ce phénomène comme le signe d'une disparition plus rapide des effets d'âge pour les élèves les moins favorisés socialement : ces derniers ayant davantage redoublé au cours de leur scolarité, ils sont plus âgés que les autres lorsqu'ils passent les épreuves d'évaluation, donc *a priori* moins concernés par les effets de maturité.

Synthèse. Les résultats de nos estimations sont présentés de manière synthétique dans la figure 4.8. Sur ce graphique sont reportés les coefficients de la colonne 5 du tableau 4.1, que l'on a regroupés en trois catégories : effet sur le score global, sur le score en mathématiques et sur le score en français. Malgré l'hétérogénéité des sources utilisées, les courbes indiquant l'évolution des effets tout au long du parcours éducatif sont remarquablement cohérentes : elles montrent que l'effet de l'âge sur les performances scolaires diminue à mesure que les élèves avancent dans le cursus scolaire. Elles semblent également indiquer que la décroissance est plus rapide en mathématiques qu'en français.

²⁵indépendants, professions libérales, cadres et professions intellectuelles, professions intermédiaires.

²⁶agriculteurs, employés, ouvriers, chômeurs et inactifs.

Figure 4.8 – Effet de l'âge (en mois) sur les performances scolaires (normalisées en unités d'écart-type) à différentes étapes du cursus éducatif. Sources : Panels primaire (1997) et secondaire (1995) de l'Éducation nationale et Base DNB (2004).



Notes: Chaque point de ce graphique correspond à l'estimation de l'effet marginal de l'âge (en mois) sur les performances scolaires mesurées à l'occasion des évaluations de CP, CE2, sixième, troisième et première. Les coefficients reportés ici sont ceux de la colonne 4 du tableau 4.1

4.4.4 Comparaison avec les estimations sur données britanniques

Les résultats que nous venons de commenter révèlent que l'âge des élèves influence significativement leurs performances à l'école et que la pénalité subie par les élèves nés en fin d'année persiste tout au long de leur scolarité. Ils ne permettent pas cependant de déterminer précisément quelle est la part de ces effets qui est due au redoublement. On peut néanmoins tenter de s'en faire une idée approximative en comparant nos résultats avec ceux obtenus sur données britanniques par Crawford et al. (2007). La quasi-absence de redoublement au Royaume-Uni constitue en effet une caractéristique institutionnelle particulièrement utile pour servir de point de comparaison avec la France : sous l'hypothèse que les effets de maturité ne diffèrent pas sensiblement dans les deux pays, la comparaison de la rapidité avec laquelle l'impact apparent de l'âge sur les performances scolaires s'atténue peut nous renseigner sur le rôle spécifique du redoublement en France.

Crawford et al. utilisent les scores obtenus aux tests nationaux d'évaluation des connaissances par les élèves scolarisés au Royaume-Uni pour mesurer l'impact de l'âge sur les performances scolaires à l'âge de 5 ans (*Foundation Stage*), 7 ans (*Key Stage 1*), 11 ans (*Key Stage 2*), 14 ans (*Key Stage 3*) et 16 ans (*Key Stage 4*). Les scores individuels sont globaux à 5 et 7 ans et sont disponibles séparément en mathématiques, en anglais et en sciences à l'âge de 11, 14 et 16 ans. Afin de faciliter la comparaison avec les scores obtenus par les élèves français, nous avons choisi de limiter notre comparaison à la moyenne des effets estimés en mathématiques et en anglais. Les estimations utilisées pour les élèves français sont calculées à partir des coefficients reportés dans la colonne 4 du tableau 4.1 : effet de l'âge sur le score global pour le CP, moyenne des effets sur le score en mathématiques et en français pour le CE2, la sixième et la troisième (examen final). Au Royaume-Uni, l'âge moyen des élèves au moment des évaluations est calculé en supposant qu'ils appartiennent tous à la même cohorte scolaire (ce qui est presque toujours le cas). En France, l'âge moyen de passage des évaluations est calculé à partir des informations contenues dans les panels primaire et secondaire, ainsi que dans la base du DNB 2004.

Le tableau 4.3 présente les résultats de cette comparaison et la figure 4.9 en pro-

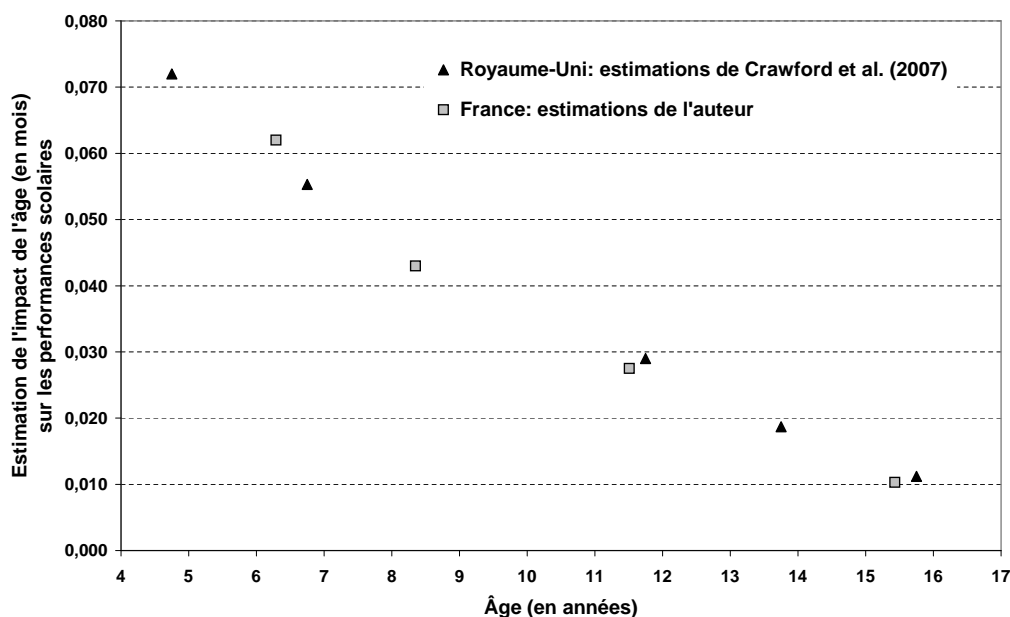
Table 4.3 – Comparaison des effets de l'âge (en mois) sur les performances scolaires en France et au Royaume-Uni, en fonction de l'âge des élèves (en années). Sources : calculs de l'auteur pour la France (cf. tableau 4.1) et estimations de Crawford et al. (2007) pour le Royaume-Uni.

Année de l'évaluation		Âge moyen de passage de l'évaluation (en années)	Effet estimé de l'âge (en mois) sur les performances scolaires	
France (1)	Royaume-Uni (2)		France (3)	Royaume-Uni (4)
	<i>Foundation Stage</i>	4.75		0.072
CP		6.29	0.062	
	<i>Key Stage 1</i>	6.75		0.055
CE2		8.35	0.043	
Sixième		11.51	0.028	
	<i>Key Stage 2</i>	11.75		0.029
	<i>Key Stage 3</i>	13.75		0.019
Troisième		15.43	0.010	
	<i>Key Stage 4</i>	15.75		0.011

Notes: Ce tableau compare les effets estimés de l'âge (en mois) sur les performances scolaires en France et au Royaume-Uni, en fonction de l'âge (en année). Les estimations pour la France sont issues des estimations réalisées dans la colonne 4 du tableau 4.1. Les estimations pour le Royaume-Uni sont tirées de Crawford et al. (2007). Les chiffres reportés dans la colonne 3 correspondent, pour le CP, à l'effet estimé de l'âge sur le score global et, pour le CE2, la sixième et la troisième, à la moyenne des effets estimés sur le score en mathématiques et en français (notes à l'examen final dans le cas du DNB). Les chiffres reportés dans la colonne 4 sont calculés à partir des tableaux 5.1 (*Foundation Stage* et *Key Stage 1*), B.1 (*Key Stage 2* et *Key Stage 3*) et 5.3 (*Key Stage 4*) de Crawford et al. (2007).

pose une visualisation graphique. Les points symbolisés par un triangle correspondent aux effets estimés de l'âge (en mois) sur les performances scolaires au Royaume-Uni, en fonction de l'âge (en années) des élèves au moment des évaluations. Les estimations obtenues sur données françaises sont quant à elles symbolisées par un carré. De manière particulièrement remarquable, le profil décroissant des effets de l'âge (en mois) sur les performances scolaires se superpose presque exactement dans les deux pays. Cette comparaison semble corroborer l'hypothèse d'un impact relativement faible du redoublement sur les performances scolaires. Si le redoublement avait un impact positif ou négatif sur les résultats des élèves, on devrait en effet observer une dissociation des deux courbes pour les effets mesurés au-delà de l'âge de 6 ans (les effets mesurés en CP n'étant pas susceptibles d'être contaminés par le redoublement). Or le graphique ne laisse pas apparaître de franche divergence, ce qui semble indiquer soit que les effets positifs ou négatifs du redoublement sont négligeables par rapport aux effets de l'âge auquel les élèves passent les examens, soit qu'ils s'annulent mutuellement.

Figure 4.9 – *Comparaison des effets de l'âge (en mois) sur les performances scolaires en France et au Royaume-Uni, en fonction de l'âge des élèves (en années). Sources : calculs de l'auteur pour la France et estimations de Crawford et al. (2007) pour le Royaume-Uni.*



Notes: Les coefficients utilisés dans ce graphique sont ceux du tableau 4.3.

4.5 L'effet du mois de naissance sur les trajectoires éducatives

Si le handicap subi par les élèves nés en fin d'année par rapport à leurs camarades nés en début d'année se limitait à un moindre développement intellectuel au moment des évaluations, il n'y aurait pas *a priori* de raison de penser que la date de naissance puisse influencer significativement leur parcours scolaire et leur vie professionnelle, dans la mesure où les effets d'âge ont presque entièrement disparu lorsque les élèves atteignent la fin de la scolarité obligatoire.

En réalité, le système éducatif français présente un certain nombre de caractéristiques institutionnelles qui pourraient contribuer à amplifier les effets de la date de naissance en les rendant responsables de trajectoires scolaires différenciées. Dans cette partie, nous nous intéressons aux effets propres de trois de ces spécificités institutionnelles : le redoublement, l'orientation et l'obligation scolaire.

4.5.1 Impact sur le redoublement

L'influence du mois de naissance sur la probabilité de redoubler a déjà été évoquée dans le cadre de l'estimation des effets de l'âge sur les performances scolaires. Nous proposons ici d'en évaluer précisément l'impact en fonction de l'âge des élèves, à partir des données du panel primaire de l'Éducation nationale et de la base Scolarité pour le secondaire.

Nous mesurons le redoublement comme la probabilité d'avoir au moins un an de retard scolaire. Pour calculer la proportion d'élèves en retard entre l'âge de 6 ans et l'âge de 10 ans, nous avons utilisé les informations contenues dans le panel primaire sur les élèves entrés en CP en 1997 : à chaque rentrée scolaire, on connaît la classe fréquentée, ce qui permet de savoir si l'élève a redoublé ou non. Le calcul de la proportion d'élèves en retard entre l'âge de 11 ans et l'âge de 15 ans est un peu plus compliqué à réaliser, dans la mesure où les éditions annuelles de la base Scolarité ne contiennent que les élèves qui sont inscrits dans les premier et second cycle de l'enseignement secondaire et ne comptabilisent pas ceux qui sont encore dans

l'enseignement primaire²⁷. Pour les besoins de notre étude, nous avons choisi de suivre le parcours secondaire des élèves nés en 1989, qu'ils soient scolarisés dans le secteur public ou dans le secteur privé. Le calcul de la proportion d'élèves en retard à l'âge de 13, 14 et 15 ans au sein de cette cohorte peut être effectué sans difficulté à partir des bases Scolarité 2002-2003, 2003-2004 et 2004-2005, dans la mesure où la quasi-totalité des élèves d'une cohorte donnée sont alors présents dans le fichier²⁸. Les bases Scolarité ne permettant pas de connaître l'année d'entrée en CP, nous supposons que tous ces élèves ont commencé l'école primaire l'année de leurs 6 ans, soit en 1983. Pour connaître la proportion d'élèves nés en 1989 qui sont en retard à l'âge de 11 et 12 ans, nous avons reconstitué les effectifs manquants des bases scolarité 2000-2001 et 2001-2002 à partir des effectifs constatés lorsque les élèves de cette cohorte ont atteint l'âge de 13 ans (base scolarité 2002-2003). Les calculs sont effectués séparément pour chaque mois de naissance. Tous les élèves qui ont été ainsi ajoutés aux bases scolarité 2000-2001 et 2001-2002 sont considérés comme ayant un an de retard ou plus.

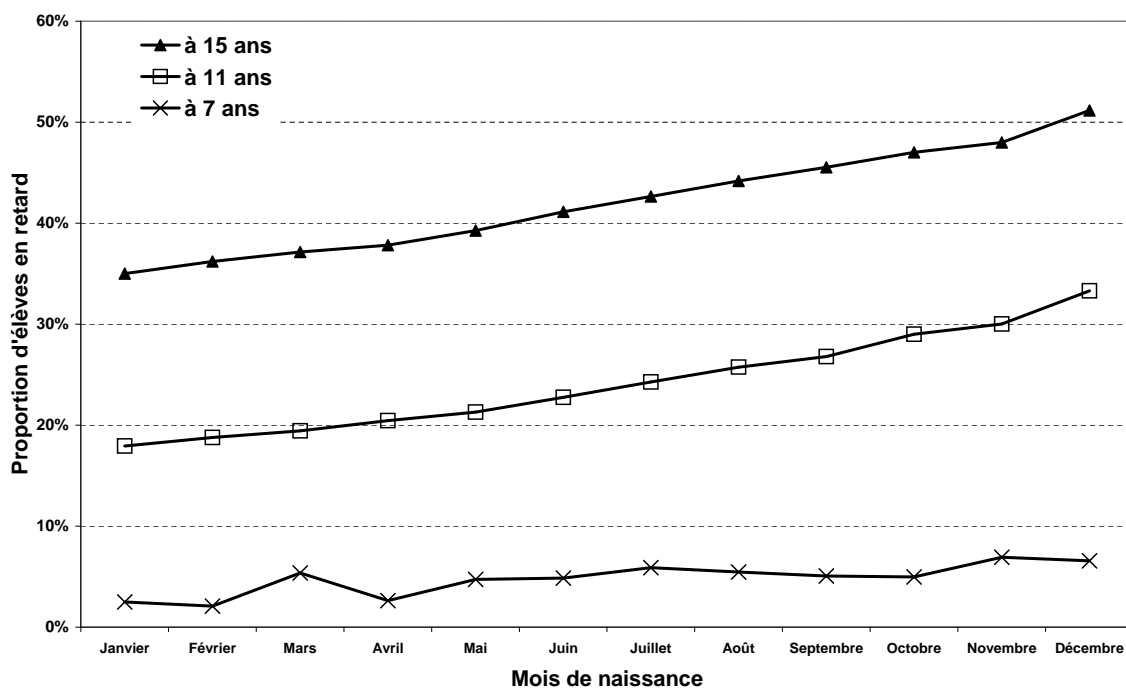
Le graphique de la figure 4.10 indique la proportion d'élèves en retard à l'âge de 7, 11 et 15 ans en fonction de leur mois de naissance. L'augmentation du taux de retard scolaire avec le mois de naissance est spectaculaire : à 15 ans, 51 % des élèves nés en décembre ont redoublé contre seulement 35 % pour ceux nés en janvier. On constate également que la pente des courbes de ce graphique n'augmente pas sensiblement entre l'âge de 11 ans et l'âge de 15 ans, ce qui semble indiquer que l'influence du mois de naissance sur le redoublement s'exerce surtout au cours de la scolarité primaire.

Pour mesurer plus précisément l'évolution de ce phénomène pendant la scolarité des élèves, nous utilisons un modèle Probit qui nous permet d'estimer l'impact marginal du mois de naissance sur la probabilité d'avoir une ou plusieurs années de retard scolaire entre l'âge de 7 ans et l'âge de 15 ans. Les résultats de nos estimations sont indiquées dans le tableau 4.4. Les effets marginaux estimés sont reportés dans la colonne 1 et la pénalité subie par les élèves nés en décembre par rapport aux élèves nés en janvier dans la colonne 2. Des estimations séparées en fonction du milieu social

²⁷Nous avons renoncé à calculer la proportion d'élèves en retard au-delà de l'âge de 15 ans dans la mesure où nombre d'entre eux sont sortis du système éducatif ou ne sont pas comptabilisés même s'ils sont encore scolarisés.

²⁸Très rares sont en effet les élèves qui ont redoublé plus de deux fois au cours de leur scolarité primaire.

Figure 4.10 – Proportion d'élèves ayant au moins un an de retard scolaire à l'âge de 7, 11 et 15 ans en fonction de leur mois de naissance. Sources : Panel primaire de l'Éducation nationale (1997) et Base Scolarité (2004-2005).



Notes: Les calculs pour la proportion d'élèves en retard à l'âge de 7 ans ont été effectués à partir des données du panel primaire de l'Éducation nationale (1997) et les calculs à l'âge de 11 et 15 ans à partir de la base Scolarité 2004-2005.

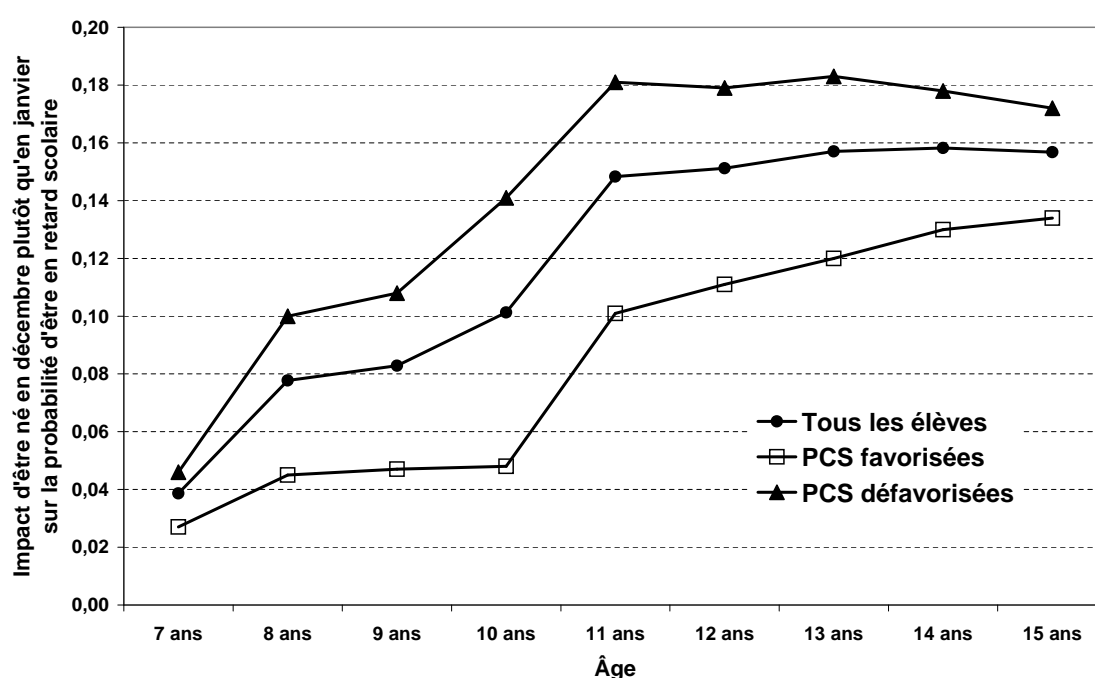
Table 4.4 – *Estimation Probit de l'impact marginal du mois de naissance sur la probabilité d'avoir au moins un an de retard scolaire en fonction de l'âge et du milieu social. Sources : Panel primaire de l'Éducation nationale (1997) et Base Scolarité (2004-2005).*

		Variable dépendante : probabilité d'avoir au moins un an de retard scolaire					
Âge	Source	Tous les élèves		PCS favorisées		PCS défavorisées	
		Effets marginaux (1)	Pénalité naissance décembre/janvier (2)	Effets marginaux (3)	Pénalité naissance décembre/janvier (4)	Effets marginaux (5)	Pénalité naissance décembre/janvier (6)
à 7 ans	PPEN 1997	0.003*** (0.001)	0.038	0.002*** (0.001)	0.027	0.004*** (0.001)	0.046
à 8 ans	PPEN 1997	0.007*** (0.001)	0.077	0.004*** (0.001)	0.045	0.009*** (0.001)	0.100
à 9 ans	PPEN 1997	0.008*** (0.001)	0.083	0.004*** (0.001)	0.047	0.010*** (0.002)	0.108
à 10 ans	PPEN 1997	0.009*** (0.001)	0.101	0.004*** (0.001)	0.048	0.013*** (0.002)	0.141
	[Nombre d'observations]	[8 663]		[3 749]		[4 914]	
à 11 ans	BASCO 2004-2005	0.013*** (0.000)	0.149	0.009*** (0.000)	0.101	0.016*** (0.000)	0.181
	[Nombre d'observations]	[799 309]		[327 140]		[472 169]	
à 12 ans	BASCO 2004-2005	0.014*** (0.000)	0.152	0.010*** (0.000)	0.111	0.016*** (0.000)	0.179
	[Nombre d'observations]	[799 309]		[327 140]		[472 169]	
à 13 ans	BASCO 2004-2005	0.014*** (0.000)	0.158	0.011*** (0.000)	0.120	0.017*** (0.000)	0.183
	[Nombre d'observations]	[799 309]		[327 140]		[472 169]	
à 14 ans	BASCO 2004-2005	0.014*** (0.000)	0.159	0.012*** (0.000)	0.130	0.016*** (0.000)	0.178
	[Nombre d'observations]	[792 577]		[326 434]		[466 143]	
à 15 ans	BASCO 2004-2005	0.014*** (0.000)	0.158	0.012*** (0.000)	0.134	0.016*** (0.000)	0.172
	Nombre d'observations	[752 877]		[317 149]		[435 728]	

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. PPEN 1997 : panel primaire de l'Éducation nationale (1997). Basco 2004-2005 : Base Scolarité (secteur public et secteur privé) de l'année scolaire 2004-2005. Chaque coefficient provient d'une régression séparée. On utilise un modèle Probit pour mesurer l'impact du mois de naissance sur la probabilité pour un élève d'avoir au moins un an de retard à un à un âge donné. La pénalité subie par les élèves nés en décembre par rapport aux élèves nés en janvier est calculée en multipliant par 11 l'effet marginal du mois de naissance sur la probabilité d'être en retard scolaire. Les élèves sont classés en « favorisés » et « défavorisés » en fonction de la PCS du chef de famille (cf. section 4.5)

d'origine sont présentées dans les colonnes 3-4 (élèves plutôt favorisés socialement) et 5-6 (élèves plutôt défavorisés socialement). La figure 4.11 propose une présentation graphique de la pénalité subie par les élèves nés en décembre par rapport aux élèves nés en janvier.

Figure 4.11 – *Estimation de l'impact d'être né décembre plutôt qu'en janvier sur la probabilité d'avoir au moins un an de retard scolaire en fonction de l'âge et du milieu social. Sources : Panel primaire de l'Éducation nationale (1997) et Base Scolarité (2004-2005).*



Notes: Ce graphique représente, en fonction de l'âge et du milieu social, l'impact du fait d'être né en décembre plutôt qu'en janvier sur la probabilité d'avoir au moins un an de retard à l'école. Chaque point du graphique correspond aux coefficients calculés dans la colonne 3 du tableau 4.4. Ce graphique montre par exemple qu'à l'âge de 13 ans, la probabilité d'être en retard scolaire est supérieure de 15 points pour les élèves nés en décembre par rapport à leurs camarades nés en janvier. Les estimations entre 7 et 10 ans ont été réalisées à partir des données du panel primaire de l'Éducation nationale (1997) et les estimations entre 11 et 15 ans à partir de la Base Scolarité 2004-2005.

Lorsque tous les élèves sont inclus dans l'échantillon, les régressions indiquent que le mois de naissance influence significativement la probabilité de redoubler tout au long de scolarité mais que l'essentiel des effets intervient au cours de la scolarité primaire : à l'âge de 11 ans, les élèves nés en décembre ont une probabilité d'avoir redoublé à l'école primaire supérieure de 14 points à ceux qui sont nés en janvier, ce qui représente un taux deux fois supérieur. On remarque également que les effets

du mois de naissance sur le redoublement se caractérisent par des augmentations particulièrement marquées à l'âge de 7 ans, 8 ans et 11 ans. Ces âges correspondent aux classes que l'on fait redoubler le plus souvent : le CP, le CE1 et le CM2. Au collège, les effets du mois de naissance n'ont plus qu'un effet très faible sur le taux de redoublement.

Les estimations séparées en fonction du milieu social d'origine révèlent que jusqu'à l'âge de 11 ans, l'effet marginal du mois de naissance sur la probabilité de redoubler est deux fois plus élevé pour les élèves socialement défavorisés que pour les élèves socialement favorisés. Dans la mesure où les premiers redoublent en moyenne deux fois plus que les seconds, ce résultat semble indiquer que le mois de naissance a un effet multiplicateur sur les facteurs qui conditionnent le redoublement. Le graphique de la figure 4.11 révèle également que l'impact du mois de naissance sur le redoublement persiste plus longtemps pour les élèves favorisés que pour les élèves défavorisés : l'écart entre la proportion de redoublants chez les élèves nés en décembre et les élèves nés en janvier reste stable (voire diminue un peu) après l'âge de 11 ans pour les premiers, alors qu'il continue à augmenter pour les seconds. Ce phénomène tient probablement au fait que les difficultés scolaires des élèves de milieux défavorisés sont plus prononcées, donc détectées plus tôt que celles des élèves de milieux favorisés. Dans la mesure où les redoublements multiples sont rares, les élèves du premier groupe pourraient avoir tendance à être davantage protégés contre le redoublement au collège que les élèves du second groupe. Cette analyse conduit à nuancer l'idée selon laquelle l'effet du mois de naissance sur le redoublement ne s'exercerait qu'à l'école primaire : pour les élèves issus de milieux plutôt privilégiés socialement, cet effet persiste au collège.

4.5.2 Mois de naissance et orientation scolaire

En augmentant la probabilité de redoubler, le mois de naissance ne se limite pas à retarder la progression des élèves dans leur cursus éducatif. Il est en effet probable que le signal négatif associé au retard scolaire contribue à influencer les choix d'orientation qui interviennent à la fin de la classe de troisième et qui se résument pour l'essentiel à la poursuite d'études généralistes dans un lycée d'enseignement

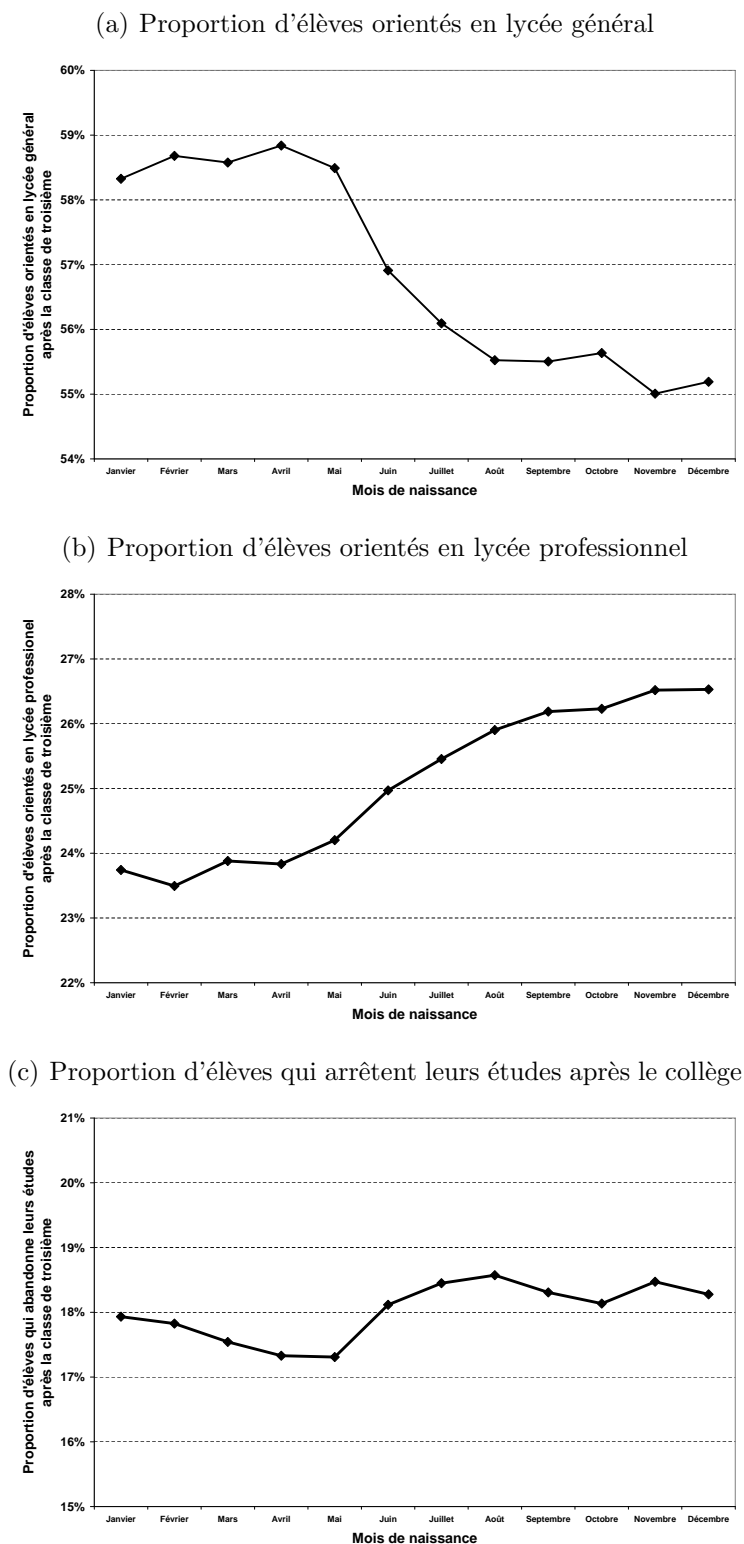
général et technologique, ou à la poursuite d'une formation professionnelle dans un lycée professionnel conduisant au bout de deux ou trois ans à l'obtention d'un CAP ou d'un BEP.

Afin de mesurer précisément l'impact du mois de naissance sur l'orientation, il aurait idéalement fallu disposer de données de panel permettant de suivre les élèves tout au long de leur scolarité. En effet, il ne suffit pas de comparer à une date donnée la répartition des élèves d'une même cohorte scolaire entre filière générale et filière professionnelle pour savoir si la date de naissance influence l'orientation, car les élèves nés en fin d'années sont plus souvent en retard par rapport à leur camarades nés en début d'année. Il est donc probable qu'on observe que les élèves qui entrent en CAP ou en BEP à l'âge de 15 ans (et qui n'ont donc pas redoublé au cours de leur scolarité) sont surreprésentés parmi les élèves nés en début d'année et qu'à l'inverse, les élèves qui entrent dans la filière professionnelle à des âges plus avancés aient plutôt tendance à être nés en fin d'année.

Bien qu'elles ne permettent pas de suivre l'ensemble de la scolarité des élèves, les bases Scolarité peuvent néanmoins être utilisées pour estimer l'impact du mois de naissance sur l'orientation scolaire, dans la mesure où elles indiquent pour chaque élève non seulement la formation suivie pendant l'année en cours, mais également la formation suivie l'année précédente. En exploitant les informations contenues dans les bases Scolarité 2000-2001 à 2004-2005, on peut ainsi connaître l'orientation en fin de collège de la quasi-totalité des élèves nés en 1986, que ceux-ci soient entrés en troisième avec un an d'avance, à l'âge normal ou avec un à trois ans de retard. Nous avons pour cela sélectionné au sein des échantillons de chacune des cinq bases les élèves nés en 1986, scolarisés dans un lycée général ou professionnel et qui étaient en troisième (générale, technologique ou SEGPA) au cours de l'année $n - 1$. Au total, cette procédure permet de reconstituer l'orientation de 82 % des élèves nés en 1986 (par rapport aux effectifs comptabilisés dans la base Scolarité 2000-2001 lorsque ces élèves avaient 14 ans). Les 18 % restants correspondent aux élèves qui ont arrêté leurs études avant d'entrer dans le second cycle de l'enseignement secondaire.

La figure 4.12 indique l'orientation suivie après la troisième par les élèves nés en 1986 en fonction de leur mois de naissance. Ces graphiques révèlent que la probabilité

Figure 4.12 – *Orientation après la classe de troisième des élèves nés en 1986.*
Sources : Bases Scolarité 2000-2001 à 2004-2005.



Notes: Ces graphiques présentent l'orientation après la classe de troisième des élèves nés en 1986, que ceux-ci aient un an d'avance en troisième, soient à l'âge normal ou aient un à trois ans de retard.

d'entrer dans la filière professionnelle plutôt que dans la filière générale est manifestement influencée par le mois de naissance. On constate en effet que 26,5 % des élèves nés en décembre sont scolarisés dans un lycée professionnel après la troisième, contre 23,7 % des élèves nés en janvier. À l'inverse, ils ne sont que 55,2 % à être scolarisés en seconde générale, contre 58,3 % pour ceux qui sont nés en janvier. Le mois de naissance ne semble pas en revanche induire un plus fort taux d'abandon des études après le collège chez les élèves nés en fin d'année par rapport à leurs camarades nés en début d'année. La courbe indiquant la proportion d'élèves qu'on ne retrouve pas dans le second cycle de l'enseignement secondaire se caractérise en effet par une forme sinusoïdale qui semble essentiellement refléter la composante sociologique du mois de naissance. Comme on a pu le voir précédemment (cf. figure 4.5), les élèves nés au second trimestre sont issus de familles légèrement plus aisées en moyenne que les élèves nés au troisième trimestre : il est donc logique de constater qu'ils sont un peu moins nombreux à abandonner leurs études après le collège.

Le tableau 4.5 présente l'estimation de l'impact du mois de naissance sur l'orientation des élèves en fin de troisième. Les effets marginaux estimés au moyen d'un modèle Probit multinomial sur l'échantillon de tous les élèves nés en 1986 sont reportés dans la colonne 1 et la colonne 2 indique la pénalité correspondante pour les élèves nés en décembre par rapport aux élèves nés en janvier. Pour réduire autant que possible l'influence de la composante sociologique du mois de naissance, on présente dans les colonnes 3 et 4 les coefficients de la même régression sur l'échantillon des élèves nés uniquement en janvier et en décembre. Ces résultats montrent que le mois de naissance augmente de manière significative la probabilité d'être orienté en lycée professionnel (la pénalité subie par les élèves nés en décembre par rapport aux élèves nés en janvier étant de l'ordre de 3 points de pourcentage) et réduit d'autant la probabilité d'être orienté en lycée général. Le mois de naissance ne semble pas avoir en revanche d'impact significatif sur la probabilité d'abandonner ses études avant le lycée.

Nous présentons le résultat des analyses par sous-groupes d'élèves dans le tableau 4.6. La comparaison des coefficients estimés séparément pour les filles et les garçons indique que l'effet du mois de naissance sur l'orientation ne varie pas en

Table 4.5 – *Impact du mois de naissance sur l'orientation des élèves en fin de troisième. Estimations réalisées à partir d'un modèle Probit multinomial. Sources : Bases Scolarité 2000-2001 à 2004-2005.*

Variable dépendante	Tous les élèves		Élèves nés en janvier ou en décembre	
	Effets marginaux (1)	Écart déc/jan (2)	Effets marginaux (7)	Écart dec/jan (8)
Orientation en lycée général	-0.004*** (0.000)	-0.044	-0.003*** (0.000)	-0.031
Orientation en lycée professionnel	0.003*** (0.000)	0.035	0.003*** (0.000)	0.028
Abandon des études après le collège	0.001*** (0.000)	0.009	0.000 (0.000)	0.003
Nombre d'observations	772 561		127 553	

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. Ce tableau présente l'estimation à l'aide d'un modèle Probit multinomial de l'impact du mois de naissance sur l'orientation en fin de troisième. L'échantillon est construit à partir des Bases scolarité 2000-2001 à 2004-2005 pour les élèves nés en 1986, que ceux-ci aient un an d'avance en troisième, soient à l'âge normal ou aient un à trois ans de retard.

Table 4.6 – *Impact du mois de naissance sur l'orientation des élèves en fin de troisième. Analyse par sous-groupes d'élèves. Estimations réalisées à partir d'un modèle Probit multinomial. Sources : Bases Scolarité 2000-2001 à 2004-2005.*

Variable dépendante	Élèves nés en janvier ou en décembre							
	Garçons		Filles		PCS favorisées		PCS défavorisées	
	Effets marginaux (1)	Écart déc/jan (2)	Effets marginaux (3)	Écart déc/jan (4)	Effets marginaux (5)	Écart déc/jan (6)	Effets marginaux (7)	Écart déc/jan (8)
Orientation en lycée général	-0.003*** (0.000)	-0.033	-0.003*** (0.000)	-0.033	-0.002*** (0.000)	-0.019	-0.004*** (0.000)	-0.041
Orientation en lycée professionnel	0.003*** (0.000)	0.028	0.003*** (0.000)	0.028	0.002*** (0.000)	0.022	0.003*** (0.000)	0.032
Abandon des études après le collège	0.000 (0.000)	0.005	0.000 (0.000)	0.002	-0.000 (0.000)	-0.003	0.001*** (0.000)	0.008
Nombre d'observations	64 475		63 078		50 426		77 127	

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. Ce tableau présente l'estimation à l'aide d'un modèle Probit multinomial de l'impact du mois de naissance sur l'orientation en fin de troisième. L'échantillon est construit à partir des Bases scolarité 2000-2001 à 2004-2005 pour les élèves nés en 1986, que ceux-ci aient un an d'avance en troisième, soient à l'âge normal ou aient un à trois ans de retard.

fonction du sexe des élèves. En revanche, les coefficients sont plus élevés pour les élèves de milieu défavorisé que pour les élèves de milieu favorisé : le mois de naissance réduit leur probabilité d'être orientés en lycée général de 4 points, augmente leur probabilité d'être orientés en lycée professionnel de 3 points et augmente d'un point leur probabilité d'abandonner les études après le collège.

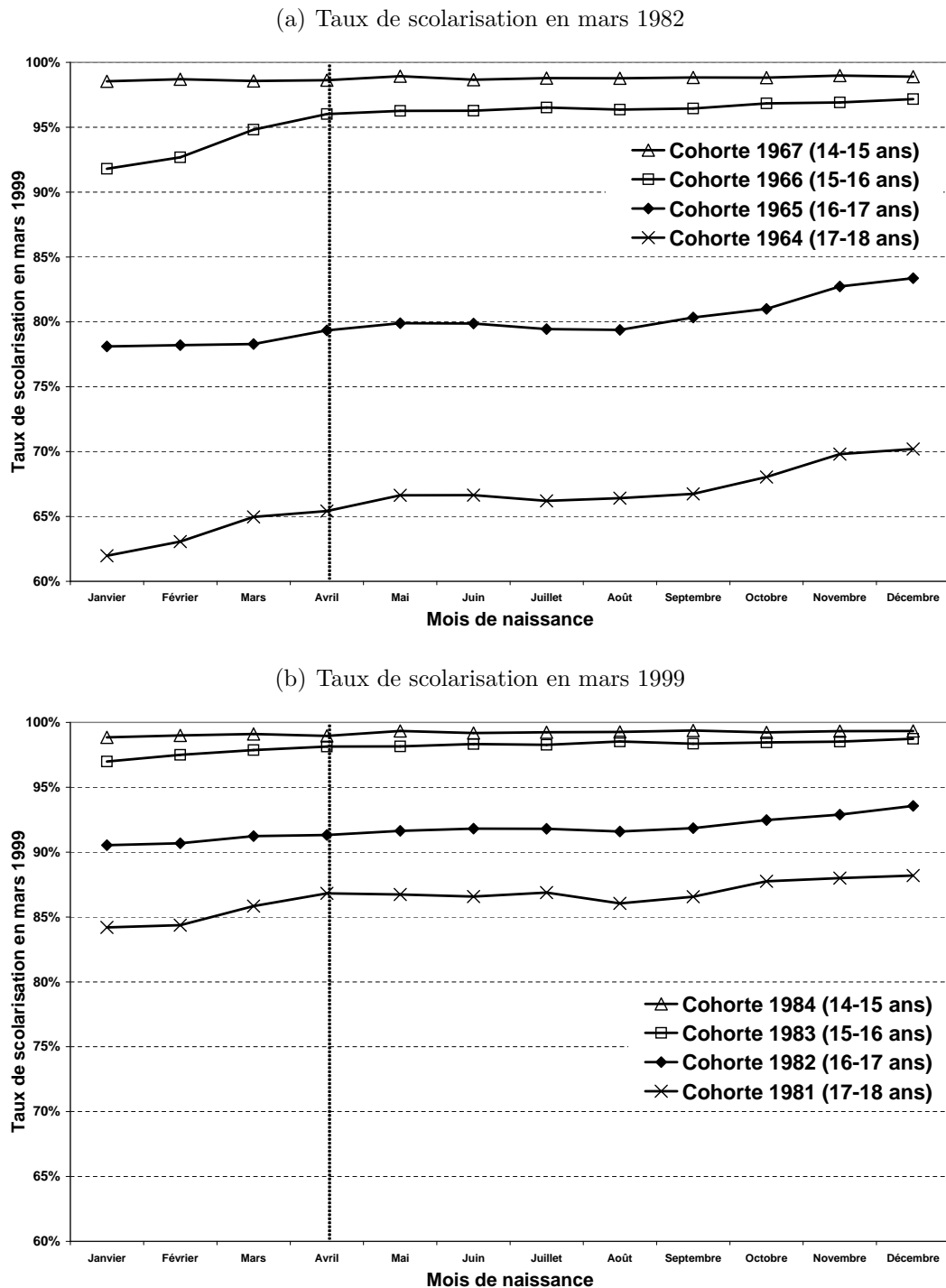
4.5.3 Mois de naissance et obligation scolaire

Le dernier effet du mois de naissance que nous analysons dans cette partie concerne le moment que les individus choisissent pour arrêter leurs études. Ainsi que nous l'avons mentionné dans l'introduction, l'obligation scolaire fonctionne en France de telle sorte que les individus nés en début d'année peuvent théoriquement quitter l'école plus tôt que ceux qui sont nés en fin d'année. On peut donc s'attendre à ce que leur durée moyenne d'études soit un peu plus courte.

Les données dont nous disposons ne nous permettent pas de mesurer ce phénomène de manière directe. En effet, aucune des sources utilisées dans la présente étude ne mentionne la date exacte à laquelle un individu a quitté l'école. On peut néanmoins analyser ce mécanisme de façon indirecte à partir des données du recensement. Ce dernier est en effet réalisé au mois de mars d'une année donnée et permet de savoir si un individu est ou non scolarisé à cette date. Nous avons donc calculé, à partir des recensements de 1982 et 1999, le taux de scolarisation en fonction de l'année et du mois de naissance pour les jeunes âgés de 14 à 18 ans.

Le RP 1982 a été effectué au début du mois de mars 1982, si bien que les individus nés en janvier ou en février de l'année n ont alors $1982 - n$ ans, alors que ceux qui sont nés plus tard dans l'année ont $1982 - n + 1$ ans. Le graphique 4.13(a) de la figure 4.13 indique le taux de scolarisation des jeunes nés entre 1964 et 1967 au mois de mars de l'année 1982, en fonction de leur mois de naissance. L'effet de l'obligation scolaire apparaît clairement sur la courbe des individus nés en 1966 : le taux de scolarisation chute de 4 points pour ceux qui atteignent l'âge de 16 ans en 1982. Le graphique révèle cependant que l'impact du mois de naissance sur le taux de scolarisation existe indépendamment de l'obligation scolaire. On remarque en effet que pour les âges situés au-delà de l'obligation scolaire, le taux de scolarisation augmente avec le mois

Figure 4.13 – Taux de scolarisation des jeunes âgés de 14 à 18 ans en mars de l'année du recensement en fonction de leur mois de naissance. Sources : Recensements de la population 1982 et 1999 (au 1/4).



Notes: Ces graphiques indiquent le taux de scolarisation des jeunes âgés de 14 à 18 ans en mars de l'année du recensement. En mars 1982, les individus nés en 1966 ont 16 ans s'ils sont nés avant le mois de mars et 15 ans s'ils sont nés après. Sur le graphique du haut, on remarque que le taux de scolarisation chute lorsque les individus atteignent l'âge de 16 ans (courbe correspondant à la cohorte 1966).

de naissance, profil qui semble indiquer que l'âge biologique joue un rôle dans la décision de quitter l'école : la probabilité de ne plus être à l'école en mars 1982 est en moyenne supérieure de 5 à 8 points pour les élèves nés en janvier par rapport aux élèves nés en décembre.

Le graphique 4.13(b) est construit selon le même principe que le précédent, mais à partir des données du recensement de 1999. Du fait de l'allongement des études, l'effet de l'obligation scolaire à 16 ans est devenu imperceptible sur la courbe des individus nés en 1984 et l'on n'observe pas de franche rupture associée au passage de 15 à 16 ans. En revanche, l'impact de l'âge sur le taux de scolarisation reste visible pour les âges situés au-delà de l'obligation scolaire : les individus nés en janvier interrompent leurs études un peu plus tôt que les individus nés en décembre.

Au total, on constate que s'il y a bien une relation entre taux de scolarisation et mois de naissance, celle-ci est très largement indépendante de l'obligation scolaire et vient en quelque sorte compenser l'effet du mois de naissance sur la durée de la scolarité pré-élémentaire (cf. section 4.4.1) : les individus nés en fin d'année arrêtent leurs études un peu plus tard que les individus nés en début d'année, ce qui fait qu'ils restent un peu plus longtemps à l'école et entrent un peu plus tard sur le marché du travail.

4.6 Les effets de la date de naissance persistent-ils à l'âge adulte ?

Les résultats des analyses menées dans la section précédente indiquent clairement que la date de naissance des individus influence non seulement leurs performances à l'école, mais également leur parcours scolaire, les élèves nés en fin d'année ayant une plus forte probabilité de redoubler ou d'être orientés en filière professionnelle que les élèves nés en début d'année. Cet effet négatif du mois de naissance pourrait être en partie compensé par le fait qu'ils quittent l'école un peu plus tard. Dans ce contexte, il paraît naturel de se demander si la différenciation des trajectoires scolaires selon le mois de naissance influence de manière significative le destin professionnel des individus. Pour répondre à cette question, nous nous appuyons principalement sur

les données de l'enquête Emploi, qui fournissent de nombreux renseignements sur le niveau de formation, la situation d'emploi et le niveau de salaire des individus à l'âge adulte.

Par construction, les informations dont on dispose concernent des cohortes antérieures aux cohortes d'élèves auxquelles nous nous sommes jusqu'à présent intéressés. Ce saut dans le temps nous paraît néanmoins légitime, dans la mesure où il n'y a guère de raison de penser que les mécanismes mis au jour dans cette étude se soient exercés très différemment pour les individus nés quelques décennies plus tôt que pour les élèves nés entre la fin des années 1980 et le début des années 1990. Dans le reste de cette partie, nous ne nous intéresserons qu'aux personnes nées entre 1945 et 1965, afin de nous assurer que nos échantillons contiennent uniquement des personnes qui ont terminé leurs études et sont entrées dans la vie active.

4.6.1 Mois de naissance et niveau de formation

En lien avec la partie précédente, nous commençons par étudier l'impact du mois de naissance sur le niveau de formation atteint au moment de l'entrée sur le marché du travail.

Âge de fin d'études. Cette question peut d'abord être analysée du point de vue de l'âge de fin d'études, exprimé en années. Il est en effet possible que le mois de naissance influence non seulement la période de l'année que les individus choisissent pour arrêter leurs études (cf. section 4.5.3), mais également la durée totale de leurs études.

On peut connaître l'âge de fin d'études des individus grâce aux données de l'enquête Emploi. Nous avons utilisé les enquêtes réalisées entre 1990 à 2002 pour calculer l'impact du mois de naissance sur cette variable, en n'utilisant que les résultats de la première vague d'enquête (afin d'éviter qu'un même individu apparaisse plusieurs fois). Le tableau 4.7 indique le coefficient mesurant l'écart entre les individus nés en décembre et les individus nés en janvier, pour différentes spécifications estimées séparément pour les hommes et pour les femmes. Les coefficients des colonnes 1 et 4 correspondent aux régressions sans variables de contrôle. On inclut des indicatrices

Table 4.7 – *Impact du mois de naissance sur l'âge de fin d'études (en années). Individus nés entre 1945 et 1965. Source : Enquête Emploi (1990-2002).*

Variable dépendante	Hommes			Femmes		
	Écart décembre/janvier			Écart décembre/janvier		
	(1)	(2)	(3)	(4)	(5)	(6)
Âge de fin d'études	-0.059 (0.055)	-0.057 (0.036)	-0.050 (0.055)	0.067** (0.023)	0.072** (0.032)	0.101** (0.048)
VARIABLES DE CONTRÔLE :						
Année d'enquête	Non	Oui	Oui	Non	Oui	Oui
Année de naissance	Non	Oui	Oui	Non	Oui	Oui
ÉCHANTILLON :						
Mois de naissance	Tous	Tous	Jan. & Déc	Tous	Tous	Jan. & Déc
Nombre d'observations	97 746	97 746	16 195	102 532	102 532	17 078

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. Ce tableau présente l'estimation de l'impact du mois de naissance sur l'âge de fin d'études. L'échantillon est construit à partir des fichiers de l'enquête Emploi (1990-2002) et comprend tous les individus nés entre 1945 et 1965. Les écarts-types sont robustes à la corrélation intra-cohorte.

pour l'année d'enquête et l'année de naissance dans les colonnes 2 et 3. Les coefficients reportés dans les colonnes 3 et 8 correspondent à la spécification complète, mais estimée uniquement pour les individus nés en janvier ou en décembre.

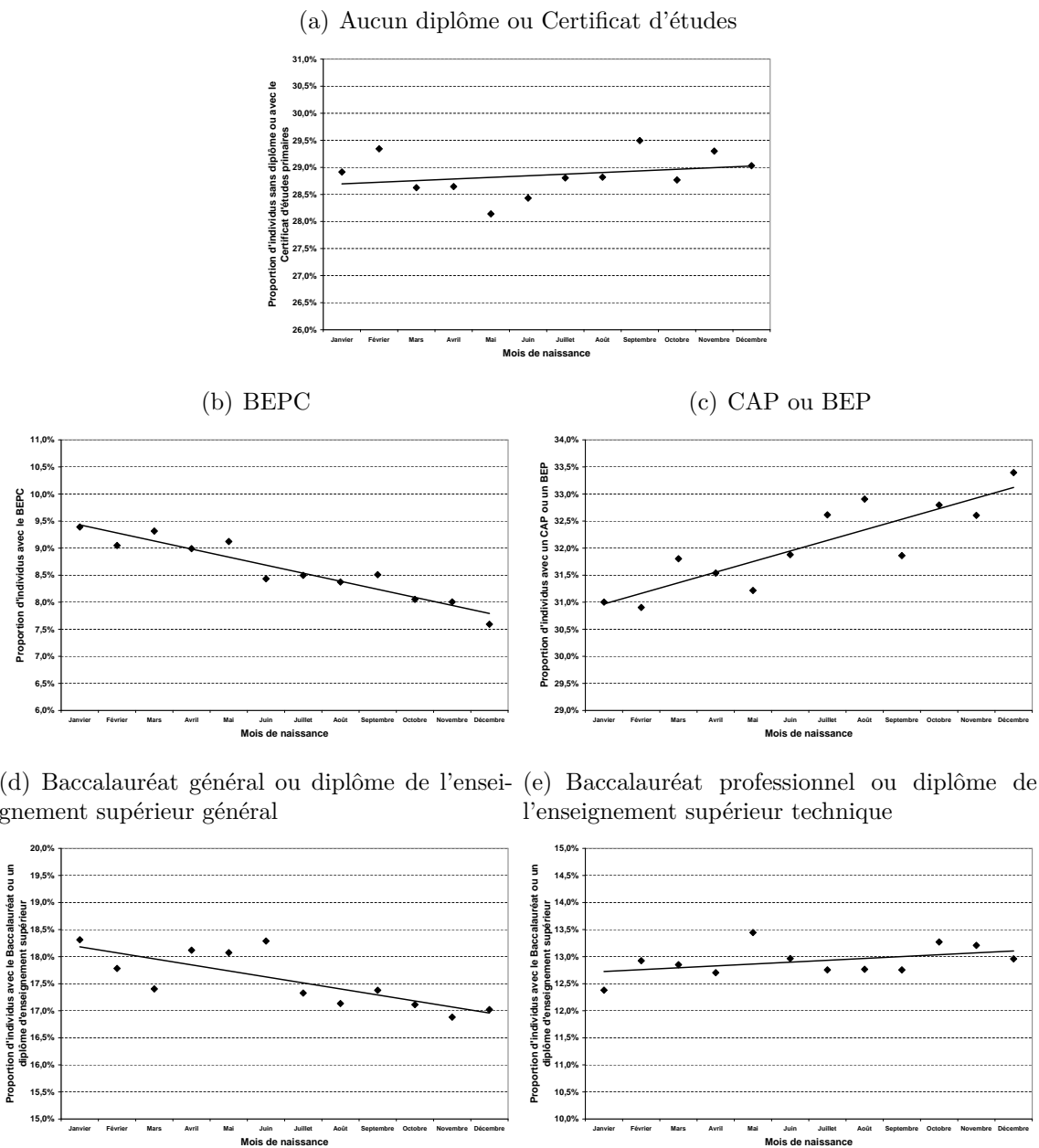
Les résultats de ces régressions indiquent que l'impact du mois de naissance sur l'âge de fin d'études est non significatif pour les hommes et très faiblement positif pour les femmes (de l'ordre d'un mois de plus pour les femmes nées en décembre par rapport aux femmes nées en janvier). À la différence de ce qui a pu être constaté, par exemple, en Suède (Fredriksson et Ockert, 2006), les moindres performances scolaires des élèves français nés en fin d'année ne semblent pas les conduire à abandonner leurs études plus tôt que les autres.

Diplômes. Si le mois de naissance ne semble pas avoir d'impact substantiel sur la durée des études, il peut néanmoins avoir un impact sur le type de diplôme que les individus détiennent au moment de leur entrée dans la vie active. La partie précédente a en effet révélé que les individus nés en fin d'année ont une plus forte probabilité d'être orientés vers des formations professionnelles. On peut donc s'attendre à ce qu'ils soient plus nombreux à détenir les diplômes auxquels conduisent ces formations.

Pour étudier cette question, nous utilisons les informations fournies par les données de l'enquête Emploi sur le dernier diplôme obtenu par les personnes interrogés, que nous avons regroupé en cinq catégories : aucun diplôme ou Certificat d'études primaires ; BEPC seul ; CAP ou BEP ; Baccalauréat général ou diplôme de l'enseignement supérieur général ; Baccalauréat professionnel ou diplôme de l'enseignement supérieur technique.

La figure 4.14 indique la répartition des diplômes en fonction du mois de naissance des individus nés entre 1945 et 1965. Le principal enseignement de ces graphiques est que le mois de naissance exerce une influence très nette sur le type de formation suivie. D'une manière générale, on constate que les individus nés en fin d'année ont une probabilité plus élevée de détenir un diplôme professionnel qu'un diplôme obtenu dans le cadre d'une formation généraliste. L'écart est particulièrement spectaculaire pour le CAP et le BEP, puisque ces diplômes sont détenus par 33,5 % des individus nés en décembre contre seulement 31 % des individus nés en janvier. De manière sy-

Figure 4.14 – *Dernier diplôme obtenu par les individus nés entre 1945 et 1965 en fonction de leur mois de naissance. Source : Enquête Emploi (1990-2002).*



Notes: Ces graphiques sont construits à partir des données de l'enquête Emploi (1990-2002) et indiquent la répartition des diplômes des individus nés entre 1945 et 1965, en fonction de leur mois de naissance.

métrique, les individus nés en décembre sont moins nombreux à arrêter leurs études avec le BEPC (8 % contre 9,5 % pour les individus nés en janvier) ou avec un diplôme de l'enseignement général de niveau supérieur ou égal au baccalauréat général (17 % contre 18,5 %). Le mois de naissance ne semble pas en revanche exercer une influence déterminante sur la probabilité d'avoir quitté l'école sans diplôme ou avec le Certificat d'études primaires²⁹, si l'on fait abstraction de la composante sociologique qui désavantage les individus nés au second trimestre par rapport aux autres. Ces observations paraissent compatibles avec ce que l'on sait du phénomène d'orientation scolaire différenciée en fonction du mois de naissance : les élèves nés en fin d'année ayant davantage tendance à être orientés en filière professionnelle, il est logique qu'on les retrouve plus nombreux à entrer sur le marché du travail avec un diplôme correspondant à ce type de formations.

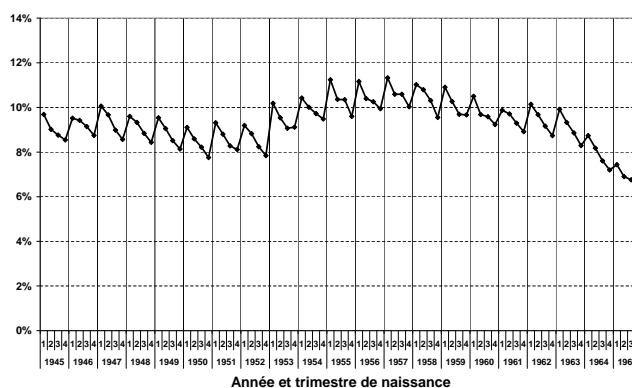
Bien que les informations contenues dans le recensement de la population ne permettent pas de distinguer aussi finement le type de diplôme détenu que l'enquête Emploi, elles peuvent néanmoins être utilisées pour confirmer l'impact du mois de naissance sur la probabilité de détenir un BEPC, un CAP-BEP ou un diplôme de l'enseignement général supérieur ou égal au baccalauréat, qui sont les trois types de diplômes dont la répartition est *a priori* la plus sensible à ce facteur. Les graphiques de la figure 4.15, construits à partir du recensement de 1999, indiquent la proportion d'individus dont le diplôme le plus élevé est soit le BEPC (graphique 4.15(a)), soit le CAP ou le BEP (graphique 4.15(b)), soit le baccalauréat général ou un diplôme universitaire général (graphique 4.15(c)) en fonction de leur année et de leur trimestre de naissance. Ces trois graphiques ont un profil en « dents de scie » particulièrement marqué, qui confirme le diagnostic établi grâce aux données de l'enquête Emploi : être né en fin d'année augmente sensiblement la probabilité de quitter l'école avec un CAP ou un BEP et diminue la probabilité de quitter l'école avec le BEPC, le Bac général ou un diplôme universitaire général (le pic du second trimestre pour ces derniers diplômes reflète la composante sociologique du mois de naissance).

Pour quantifier plus précisément l'impact du mois de naissance sur le type de diplôme détenu, nous avons procédé à un certain nombre d'estimations dont les résul-

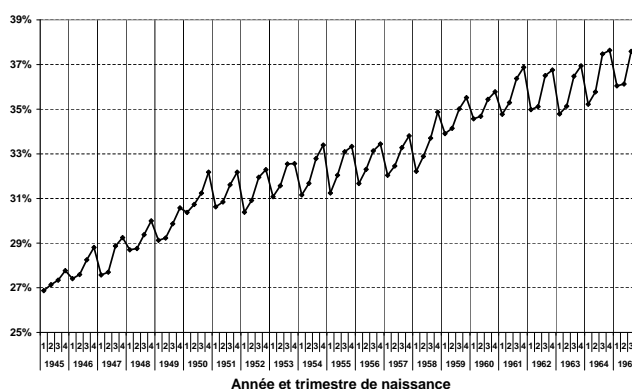
²⁹Le Certificat d'Études primaires ayant progressivement disparu à partir de la fin des années 1960, il ne compte que pour une très petite partie des diplômes détenus par les cohortes considérées.

Figure 4.15 – Proportion d'individus nés entre 1945 et 1965 dont le dernier diplôme est soit le BEPC, soit un CAP ou un BEP, soit le Baccalauréat général ou un diplôme universitaire général en fonction de leur année et de leur trimestre de naissance.
 Source : Recensement de la population 1999 (au 1/4).

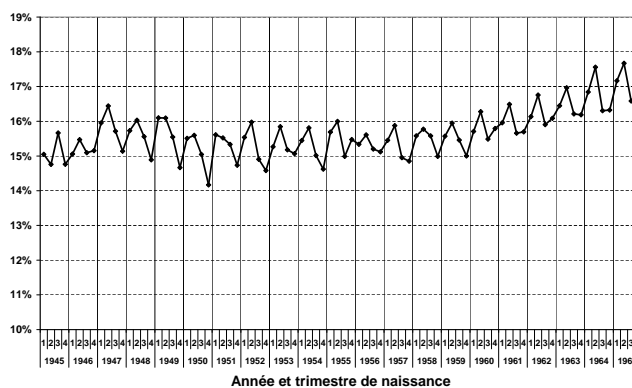
(a) BEPC



(b) CAP ou BEP



(c) Baccalauréat général ou diplôme de l'enseignement supérieur général



Notes: Ces graphiques sont construits à partir des données du recensement de 1999 et indiquent la proportion d'individus nés entre 1945 et 1965 dont le dernier diplôme est soit le BEPC, soit un CAP ou un BEP, soit le Baccalauréat général ou un diplôme universitaire général en fonction de leur année et de leur trimestre de naissance.

tats sont reportés dans le tableau 4.8. Les coefficients sont issus de régressions Probit multinomiales et les spécifications adoptées sont identiques à celles du tableau 4.7 et sont estimées séparément pour les hommes et pour les femmes. Les coefficients sont relativement proches d'une spécification à l'autre et indiquent que le fait d'être né en décembre plutôt qu'en janvier augmente significativement la probabilité de détenir un diplôme professionnel plutôt qu'un diplôme de l'enseignement général : les individus nés en fin d'année sont plus nombreux à entrer sur le marché du travail avec un CAP, un BEP, un baccalauréat professionnel ou un diplôme de l'enseignement supérieur technique qu'avec un BEPC, un baccalauréat général ou un diplôme universitaire général.

Un examen attentif des coefficients reportés dans ce tableau révèle que l'impact du mois de naissance sur la structure des diplômes pénalise davantage les hommes que les femmes. On constate en effet que pour les hommes, une naissance plus tard dans l'année réduit globalement le niveau des diplômes obtenus à l'issue de la scolarité, si l'on considère que la valeur d'un CAP ou d'un BEP ne diffère pas sensiblement de celle du BEPC³⁰ : le fait d'être né en décembre plutôt qu'en janvier augmente leur probabilité de ne pas avoir de diplôme (+0,5 point) ou d'avoir un diplôme professionnel inférieur au baccalauréat (+2 à 3 points), mais n'augmente pas leur probabilité d'avoir un diplôme professionnel de niveau équivalent ou supérieur au baccalauréat professionnel. Pour les femmes, la situation est moins défavorable : elles sont certes plus nombreuses à avoir un diplôme professionnel inférieur au Baccalauréat (+1,5 à 2 points), mais également plus nombreuses à avoir le baccalauréat professionnel ou un diplôme de l'enseignement supérieur technique (+1 point). Ce phénomène pourrait s'expliquer à la fois par la meilleure réussite scolaire des élèves de sexe féminin, mais également par la forte féminisation d'un certain nombre de formations dispensées dans le cadre de l'enseignement supérieur professionnel (notamment les diplômes des professions sociales ou de la santé).

À la lumière de ces résultats, il apparaît que si le mois de naissance influence significativement le type de diplôme obtenu à l'issue de la scolarité, son effet sur le niveau de ces diplômes n'est réellement sensible que pour les hommes. On peut donc

³⁰Les rendements salariaux apparents de ces diplômes sont relativement proches lorsqu'on les calcule à partir des données de l'enquête Emploi.

Table 4.8 – *Impact du mois de naissance sur le dernier diplôme obtenu. Estimations réalisées à partir d'un modèle Probit multinomial. Individus nés entre 1945 et 1965. Source : Enquête Emploi (1990-2002).*

Variable dépendante	Hommes			Femmes		
	Écart décembre/janvier			Écart décembre/janvier		
	(1)	(2)	(3)	(4)	(5)	(6)
DERNIER DIPLÔME OBTENU :						
Aucun diplôme ou Certificat d'études primaires	0.004 (0.005)	0.003 (0.005)	0.004 (0.007)	-0.001 (0.004)	-0.001 (0.005)	-0.004 (0.007)
BEPC	-0.012*** (0.003)	-0.012*** (0.003)	-0.016*** (0.004)	-0.021*** (0.003)	-0.021*** (0.003)	-0.025*** (0.004)
CAP ou BEP	0.024*** (0.006)	0.024*** (0.005)	0.034*** (0.007)	0.021*** (0.005)	0.021*** (0.004)	0.018*** (0.007)
Bac général ou diplôme univ. général	-0.016*** (0.005)	-0.016*** (0.004)	-0.022*** (0.006)	-0.006** (0.003)	-0.006 (0.004)	0.001 (0.006)
Bac professionnel ou diplôme univ. technique	0.001 (0.003)	0.001 (0.003)	-0.003 (0.005)	0.007** (0.004)	0.007** (0.003)	0.011** (0.005)
VARIABLES DE CONTRÔLE :						
Année d'enquête	Non	Oui	Oui	Non	Oui	Oui
Année de naissance	Non	Oui	Oui	Non	Oui	Oui
ÉCHANTILLON :						
Mois de naissance	Tous	Tous	Jan. & Déc	Tous	Tous	Jan. & Déc
Nombre d'observations	97 868	97 868	16 224	102 598	102 598	17 082

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. Ce tableau présente l'estimation de l'impact du mois de naissance sur le dernier diplôme obtenu. L'échantillon est construit à partir des fichiers de l'enquête Emploi (1990-2002) et comprend tous les individus nés entre 1945 et 1965. Les écarts-types sont robustes à la corrélation intra-cohorte.

s'attendre à ce que la situation professionnelle des hommes nés en fin d'année soit davantage dégradée que celle des femmes.

4.6.2 Mois de naissance et situation d'emploi

Nous étudions l'impact du mois de naissance sur la situation d'emploi des individus à partir des données de l'enquête Emploi, en privilégiant trois dimensions : le taux de chômage, la probabilité d'être à temps partiel et l'appartenance à la fonction publique.

L'impact du mois de naissance sur chacune de ces variable est estimé séparément pour les hommes et les femmes à l'aide d'un modèle Probit. Les résultats des régressions sont reportés dans le tableau 4.9. Les coefficients des colonnes 1 et 4 correspondent aux régressions sans variables de contrôle. On inclut des indicatrices pour l'année d'enquête et l'année de naissance dans les colonnes 2 et 3. Les coefficients reportés dans les colonnes 3 et 8 correspondent à la spécification complète, mais estimée uniquement pour les individus nés en janvier ou en décembre.

Les coefficients sont très proches d'une spécification à l'autre et semblent indiquer que le mois de naissance n'a qu'un impact très faible sur la situation d'emploi des individus. Le fait d'être né en décembre plutôt qu'en janvier n'agit pas sur la probabilité d'être à temps partiel. Les seuls effets significatifs concernent les hommes : ceux qui sont nés en décembre sont un peu plus souvent au chômage (+ 0.5 point) et plus rarement fonctionnaires (-1 point) que ceux qui sont nés en janvier.

4.6.3 Mois de naissance et salaires

L'impact salarial du mois de naissance constitue l'ultime dimension analysée dans le cadre de cette étude. Nous calculons le salaire horaire des individus en combinant les informations fournies par l'enquête Emploi sur le salaire mensuel net et le nombre d'heures habituellement travaillées par les personnes enquêtées. Nous utilisons les séries de l'indice des prix à la consommation pour convertir ces salaires en euros de 2005, avant d'en prendre le logarithme.

Pour calculer l'impact du mois de naissance sur le salaire horaire des individus, nous avons procédé à un certain nombre de régressions dont les résultats sont présen-

Table 4.9 – *Impact du mois de naissance sur la situation d’emploi des individus. Estimations réalisées à partir d’un modèle Probit. Individus nés entre le 1^{er} janvier 1945 et le 31 décembre 1965. Source : Enquête Emploi (1990-2002).*

Variable dépendante	Hommes			Femmes		
	Écart décembre/janvier			Écart décembre/janvier		
	(1)	(2)	(3)	(4)	(5)	(6)
IMPACT DU MOIS DE NAISSANCE SUR :						
La probabilité d’être au chômage	0.004** (0.001)	0.004* (0.002)	0.004* (0.002)	0.001 (0.003)	0.001 (0.003)	-0.000 (0.005)
Nombre d’observations	93 890	93 890	15 550	81 343	81 343	13 604
La probabilité d’avoir un emploi à temps partiel	0.001 (0.001)	0.001 (0.002)	0.001 (0.002)	-0.002 (0.003)	-0.002 (0.005)	-0.006 (0.009)
Nombre d’observations	88 547	88 547	14 684	73 459	73 459	12 302
La probabilité d’être fonctionnaire	-0.011*** (0.004)	-0.011*** (0.004)	-0.018*** (0.006)	-0.007 (0.005)	-0.007 (0.005)	-0.013* (0.008)
Nombre d’observations	89 097	89 097	14 768	74 129	74 129	12 425
VARIABLES DE CONTRÔLE :						
Année d’enquête	Non	Oui	Oui	Non	Oui	Oui
Année de naissance	Non	Oui	Oui	Non	Oui	Oui
ÉCHANTILLON :						
Mois de naissance	Tous	Tous	Jan. & Déc	Tous	Tous	Jan. & Déc

Notes: * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. Ce tableau présente l’estimation de l’impact du mois de naissance sur la situation des individus. L’échantillon est construit à partir des fichiers de l’enquête Emploi (1990-2002) et comprend tous les individus nés entre 1945 et 1965, dont le salaire est renseigné et qui sont interrogés pour la première fois. Les écarts-types sont robustes à la corrélation intra-cohorte.

tés séparément pour les hommes et les femmes dans le tableau 4.10. Les coefficients reportés dans les différentes colonnes correspondent à l'estimation de l'écart salarial (en log) entre les mois extrêmes d'une même cohorte. Les régressions des colonnes 1 et 6 n'incluent aucun contrôle et indiquent qu'en moyenne, les individus nés en fin d'année ont des salaires très légèrement inférieurs à ceux qui sont nés en début d'année, la pénalité étant plus forte pour les hommes (-2,3 %) que pour les femmes (-0.7 %). Les écarts mesurés sont à peine modifiés lorsqu'on contrôle pour l'année de naissance et l'année d'enquête (colonnes 2 et 7).

À la lumière des analyses développées précédemment, il semblerait que deux facteurs puissent *a priori* expliquer les écarts salariaux observés : le premier tient au fait que les individus nés en fin d'année entrent un peu plus tard sur le marché du travail que les individus nés en début d'année, comme on a pu le voir dans la section 4.5.3 ; le second est lié à l'impact du mois de naissance sur la structure des diplômes, mis en évidence dans la section 4.6.1. Pour tenter de déterminer le pouvoir explicatif de chacun de ces facteurs, nous incluons dans la régression le diplôme le plus élevé obtenu par le salarié. Cet ajout divise par deux l'écart salarial pour les hommes, mais laisse inchangé celui des femmes. Cela paraît logique, dans la mesure où l'on a vu plus haut que les hommes nés en fin d'année sont davantage pénalisés que les femmes du point de leur niveau de formation. Une fois pris en compte l'effet du mois de naissance sur les diplômes, l'écart salarial séparant les individus nés en décembre des individus nés en janvier est sensiblement le même pour les hommes et les femmes et est égal à environ 1 %. Il est probable que cet écart résiduel provienne en grande partie de l'entrée un peu plus tardive sur le marché du travail des individus nés en fin d'année.

Compte tenu de leur ampleur relativement modeste, on ne peut exclure que les écarts salariaux ainsi mesurés soient biaisés par la composante sociologique du mois de naissance. Pour tenter de neutraliser cette dernière, nous présentons le résultat des régressions effectuées sur le sous-échantillon des individus nés en janvier et en décembre. Les coefficients estimés sont reportés dans les colonnes 4 et 9 (sans l'inclusion du diplôme) et 5 et 10 (en contrôlant pour le diplôme). Les résultats de ces régressions suggèrent que les écarts salariaux sont un peu plus élevés que ceux précédemment estimés, de l'ordre de 3 % pour les hommes (dont la moitié proviendrait

Table 4.10 – *Impact du mois de naissance sur le salaire horaire (en euros de 2005). Individus nés entre le 1^{er} janvier 1945 et le 31 décembre 1965. Source : Enquête Emploi (1990-2002).*

Variable dépendante	Hommes					Femmes				
	Écart décembre/janvier					Écart décembre/janvier				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log du salaire horaire	-0.023*** (0.006)	-0.023*** (0.005)	-0.012*** (0.004)	-0.031*** (0.007)	-0.014** (0.006)	-0.007** (0.003)	-0.008* (0.005)	-0.009* (0.004)	-0.014* (0.008)	-0.014** (0.006)
VARIABLES DE CONTRÔLE :										
Année d'enquête	Non	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui
Année de naissance	Non	Oui	Oui	Oui	Oui	Non	Non	Oui	Oui	Oui
Dernier diplôme obtenu (5 cat.)	Non	Non	Oui	Non	Oui	Non	Non	Oui	Non	Oui
ÉCHANTILLON :										
Mois de naissance	Tous	Tous	Tous	Jan. & Déc	Jan. & Déc	Tous	Tous	Tous	Jan. & Déc	Jan. & Déc
Nombre d'observations	63 551	63 551	63 551	10 603	10 603	60 231	60 231	60 231	10 091	10 091

Notes : * : coefficient significatif à 10 % ; ** : coefficient significatif à 5 % ; *** : coefficient significatif à 1 %. Ce tableau présente l'estimation de l'impact du mois de naissance sur le logarithme du salaire horaire des individus. L'échantillon est construit à partir des enquêtes Emploi 1990-2004 et n'inclut que les personnes nées entre le 1^{er} janvier 1945 et le 31 décembre 1965, dont le salaire est renseigné et qui sont interrogées pour la première fois. Les cinq catégories de diplômes sont : aucun diplôme ou Certificat d'études primaires ; BEPC ; CAP ou BEP ; Baccalauréat général ou diplôme universitaire général ; Baccalauréat professionnel ou diplôme universitaire technique. Les écarts-types sont robustes à la corrélation intra-cohorte.

des effets du mois de naissance sur les diplômes) et de 1,5 % pour les femmes.

L'ensemble des analyses développées dans cette partie indique que bien que le mois de naissance exerce une influence notable sur les trajectoires scolaires des individus, il n'affecte leur situation professionnelle que de manière très marginale. Ce résultat n'est paradoxal qu'en apparence, dans la mesure où nous avons vu que la date de naissance pèse moins sur le niveau des diplômes obtenus que sur leur type (général ou professionnel). Une pénalité spécifiquement liée au niveau de formation existe néanmoins pour les hommes et se traduit par des salaires inférieurs de 1,5 % en moyenne pour ceux qui sont nés en décembre par rapport à ceux qui sont nés en janvier.

4.7 Conclusion

Cette étude avait pour objectif d'explorer systématiquement l'impact du mois de naissance sur les trajectoires scolaires et professionnelles en France, en mobilisant un grand nombre de sources statistiques permettant de suivre les individus tout au long de leur parcours éducatif ainsi qu'à l'âge adulte.

Le principal enseignement de ce travail est que les écarts de performances liés aux différences d'âge sont plus importants et persistent davantage que ne le laissent entendre les études françaises sur ce sujet. Nos estimations indiquent que les écarts de performances liés aux différences d'âge sont particulièrement forts à l'entrée en primaire (tout particulièrement pour les élèves issus de milieux sociaux défavorisés) et tendent à diminuer ensuite, mais persistent au moins jusqu'à la fin du collège : en CP, les élèves nés en fin d'année obtiennent en moyenne des scores inférieurs de 70 % d'un écart-type à ceux de leurs camarades nés en janvier ; lorsqu'ils passent les épreuves du Brevet, l'écart reste significatif et égal à 17 % d'un écart-type.

Notre étude montre que dans le contexte institutionnel français, les effets du mois de naissance sur la réussite scolaire ne sont pas transitoires, comme on le considère généralement, mais affectent durablement les trajectoires scolaires des élèves. Nos résultats montrent en effet que le mois de naissance influence non seulement le taux de

redoublement, mais également la probabilité d'être orienté dans la voie professionnelle après le collège. Le fait d'être né en décembre plutôt qu'en janvier multiplie par deux la probabilité de redoubler au cours de sa scolarité primaire et augmente de trois points la probabilité d'être orienté en lycée professionnel après la classe de troisième.

Nous montrons cependant qu'en dépit de cette différenciation des trajectoires scolaires en fonction du mois de naissance, les individus nés en fin d'années ne paraissent pas fortement pénalisés dans leur vie professionnelle par rapport aux individus nés en début d'année : leur situation d'emploi n'est pas significativement différente et l'écart de rémunération séparant les salariés nés en décembre des salariés nés en janvier est faible puisque compris entre 1,5 et 3 %. Nous interprétons ces faibles écarts salariaux, qui proviennent en grande partie de l'entrée plus tardive sur le marché du travail des personnes nées en fin d'année, comme la conséquence du fait que le mois de naissance influence moins le niveau de diplôme que le type de formation suivie (générale ou professionnelle). Notre étude révèle néanmoins qu'il existe une pénalité salariale spécifiquement liée au plus faible niveau de qualification des hommes nés en fin d'année. Celle-ci se traduit par un écart de salaire égal à 1,5 % entre ceux qui sont nés en décembre et ceux qui sont nés en janvier.

Ces résultats suggèrent que si le mois de naissance des individus n'exerce pas une influence déterminante sur leur destin professionnel, il pénalise néanmoins le parcours scolaire de ceux qui sont nés en fin d'année. Dans la mesure où l'on ne choisit pas sa date de naissance, il paraîtrait souhaitable que le système éducatif prévoie un certain nombre de mécanismes institutionnels pour corriger les inégalités scolaires liés à ce facteur. Parmi les solutions envisageables figure l'application d'un coefficient compensateur pour redresser les résultats scolaires des élèves en fonction de leur mois de naissance. Un tel dispositif, qui aurait surtout vocation à être utilisé à l'école primaire, permettrait de renforcer l'équité du système éducatif français d'une manière à la fois simple et peu coûteuse.

Conclusion générale

L'ambition de cette thèse était de faire progresser notre compréhension des enjeux de la démocratisation scolaire en France en proposant une analyse économique de trois des principales politiques éducatives qui ont accompagné ce processus. Si nos études théoriques et empiriques fournissent un certain nombre d'éléments de réponses originaux, elles ne prétendent nullement dresser un bilan global de la démocratisation scolaire, mais fournissent des éclairages partiels sur la capacité du système éducatif français à réduire les inégalités scolaires. À nos yeux, la principale contribution de ce travail réside moins dans le caractère définitif des réponses apportées que dans la nouveauté du regard porté sur les politiques éducatives que nous avons choisi d'étudier.

Après avoir rappelé les principaux résultats de cette thèse, nous discutons les limites de nos travaux et esquissons les horizons qu'ils invitent à explorer.

Des résultats originaux

Allongement de la scolarité obligatoire et réduction des inégalités. La première partie de cette thèse indique que l'augmentation de l'âge de fin d'études minimum ne constitue pas une condition suffisante pour améliorer le destin professionnel des individus qui quittent l'école précocement. Nos résultats montrent que la réforme Berthoin, qui porta en France l'obligation scolaire à 16 ans révolus, n'a pas permis d'améliorer les salaires et la situation d'emploi des individus qui ont dû prolonger leurs études d'une ou deux années pour se conformer à la nouvelle loi. La comparaison effectuée avec l'allongement de la scolarité obligatoire au Royaume-Uni révèle cependant que l'échec de la réforme Berthoin n'était sans doute pas inéluctable,

et tient en grande partie au fait que cette loi n'a pas permis de réduire la proportion d'élèves quittant l'école sans aucune qualification. Par contraste, l'impact positif de l'*Education Act* de 1973 sur les salaires au Royaume-Uni semble pouvoir être attribué à l'amélioration très sensible de la structure des qualifications qui a suivi l'extension de l'obligation scolaire à 16 ans, en particulier dans la population féminine. Le succès de cette réforme dans la réduction des inégalités outre-Manche peut être interprété comme le fruit d'une mise en cohérence de l'obligation scolaire avec l'organisation du cursus éducatif : en faisant en sorte que les élèves britanniques ne puissent pas arrêter leurs études avant d'avoir atteint l'âge auquel ils terminent le collège, l'*Education Act* a conduit une grande partie des plus défavorisés d'entre eux à passer avec succès les épreuves du GCSE, améliorant ainsi leurs perspectives professionnelles. En France, au contraire, un certain nombre d'obstacles institutionnels ont empêché la bonne articulation de l'allongement de la scolarité obligatoire avec la structure du cursus éducatif secondaire : la pratique massive du redoublement, l'orientation en filière professionnelle en fin de quatrième et le statut mal défini du diplôme du BEPC expliquent que la population ciblée par la réforme Berthoin n'ait pas pu profiter de l'allongement de la durée d'études pour décrocher un diplôme avant d'avoir atteint l'âge de 16 ans.

Au-delà de sa dimension historique, cette étude fournit un certain nombre d'éléments de réflexion pour améliorer le fonctionnement actuel de l'obligation scolaire en France. Trente ans après l'entrée en vigueur de la réforme Berthoin, le système éducatif français ne parvient pas à réduire le nombre d'élèves qui quittent l'école sans aucune qualification : en 2007, près de 17 % d'entre eux (soit environ 130 000 par génération) achèvent leurs études sans diplôme du secondaire (diplôme du Brevet, CAP, BEP ou Baccalauréat). L'impuissance de l'obligation scolaire à juguler cette hémorragie tient en partie au fait que beaucoup d'élèves atteignent l'âge de 16 ans avant d'avoir pu passer le moindre diplôme, à la fois parce qu'ils ont souvent redoublé une ou deux fois au cours de leur scolarité et parce que l'enseignement professionnel court (CAP et BEP) ne débouche sur aucun diplôme avant deux ou trois années d'études après la classe de troisième, lorsque les élèves ont dépassé l'âge de fin d'études minimum.

Dans ces conditions, plusieurs pistes de réformes pourraient être envisagées pour faire en sorte que l'obligation scolaire contribue réellement à améliorer le sort des élèves les plus défavorisés : limiter la pratique du redoublement, dont les effets sur les performances scolaires et la situation d'emploi sont généralement estimés comme nuls ou négatifs (Gomes-Neto et Hanushek, 1994 ; Jacob et Lefgren, 2004 ; Cosnefroy et Rocher, 2004 ; Brodaty et al., 2008), permettrait de réduire le décalage entre l'âge théorique et l'âge réel de passage des diplômes ; étendre l'obligation scolaire à 17 ans et ne donner la possibilité d'arrêter ses études qu'à la fin de l'année scolaire (et non pas à la date d'anniversaire, comme c'est le cas aujourd'hui), se traduirait par une augmentation sensible du nombre d'élèves inscrits en deuxième année de CAP ou de BEP. Enfin, on peut envisager d'inciter financièrement les jeunes à prolonger leurs études au-delà de la limite légale et augmenter ainsi leurs chances d'acquérir une qualification. Un tel système est actuellement en cours d'expérimentation au Royaume-Uni, où l'*Education Maintenance Allowance* offre aux élèves les plus modestes des incitations financières à poursuivre une scolarité au-delà de l'obligation scolaire. D'après les évaluations de Dearden et al. (2005), cette réforme aurait permis d'augmenter de 6,5 % le taux de scolarisation dans la tranche d'âge concernée, l'impact étant d'autant plus important que les lycéens sont issus des ménages les plus pauvres.

Politiques de sectorisation et inégalités scolaires. La deuxième partie de cette thèse apporte des éléments théoriques et empiriques permettant de mieux comprendre les mécanismes par lesquels les politiques de sectorisation des établissements scolaires agissent sur les inégalités à l'école en influençant les choix résidentiels et scolaires des ménages. Notre étude explore en particulier les propriétés de modalités alternatives d'allocation des élèves aux établissements scolaires (sectorisation stricte, redécoupage de la carte scolaire ou libre choix de son école) afin d'éclairer les enjeux du débat qui s'est ouvert en France au sujet de la réforme de la sectorisation.

En nous appuyant sur la littérature théorique consacrée à l'étude de l'interaction entre choix résidentiel et scolaire, nous développons un modèle spatial qui permet d'analyser le lien entre stratification résidentielle et inégalités scolaires en présence

d'effets de pairs à l'école, tout en tenant compte les particularités du système éducatif français. Bien que les solutions de ce modèle ne soient pas calculables analytiquement, elles peuvent être étudiées au moyen de simulations numériques que nous calibrons sur les données parisiennes en utilisant les estimations empiriques réalisées dans le chapitre 3. Les résultats de ces simulations indiquent que la capacité des réformes du mode de recrutement des établissements à réduire les inégalités scolaires dépend de manière cruciale du rôle joué par le secteur privé. Ainsi, alors qu'un redécoupage de la carte scolaire peut sembler *a priori* la politique la plus efficace pour réduire les inégalités de niveau entre écoles, nous montrons qu'en accélérant la fuite des meilleurs élèves vers le privé, une telle réforme risque d'entraîner une diminution du niveau des écoles publiques et d'annuler en grande partie les bénéfices escomptés pour les élèves issus des milieux les plus modestes. À l'inverse, la suppression de la carte scolaire qui, dans notre modèle, entraîne une stratification des élèves par niveau dans les collèges publics, n'a pas forcément des effets aussi négatifs qu'attendus pour les élèves issus de milieux défavorisés. En effet, le libre choix de son collège peut améliorer sensiblement la situation des élèves les plus doués pour les études au sein de ce groupe, si elle entraîne un retour dans les collèges publics des bons élèves auparavant scolarisés dans le privé.

Cette partie comprend également une contribution empirique à l'analyse de l'interaction entre marché immobilier et inégalités scolaires, à travers une évaluation de l'impact du niveau des collèges à Paris sur le prix des logements. À partir de données immobilières et scolaires très riches, nous montrons que le niveau des collèges publics est capitalisé dans le prix de vente des appartements, à raison d'une augmentation de 1,5 à 2,5 % par écart-type supplémentaire dans les performances du collège public de secteur. Nous apportons également la première contribution empirique permettant d'évaluer la manière dont le secteur privé modifie l'interaction entre choix scolaire et choix résidentiel. Nous montrons qu'en atténuant la contrainte exercée par la carte scolaire, la présence d'établissements d'enseignement privé réduit localement la capitalisation du niveau des collèges publics dans les prix immobiliers. Ces résultats démontrent l'existence d'une composante scolaire dans les choix résidentiels lorsque l'affectation des élèves aux établissements dépend d'un critère géographique.

Ils suggèrent également que les réactions comportementales des familles sont un élément essentiel à prendre en compte lorsqu'il s'agit de déterminer la politique la plus efficace pour lutter contre les effets néfastes de la ségrégation résidentielle sur les inégalités scolaires.

Mois de naissance, redoublement et orientation. La troisième partie de cette thèse est consacrée aux effets du mois de naissance des individus sur leurs trajectoires scolaires et professionnelles. Si l'impact des différences d'âge sur les performances scolaires des élèves d'une même classe a déjà été mis en évidence par de nombreuses études réalisées en France et à l'étranger (Caille et Rosenwald, 2006 ; Bedard et Dhuey, 2006 ; Crawford et al., 2007), les mécanismes par lesquels ces effets sont susceptibles d'influencer de manière durable le destin scolaire et professionnel des individus sont encore mal connus.

La première contribution de notre étude empirique est de révéler que l'impact des différences d'âges sur les résultats scolaires des élèves est plus fort et persiste plus longtemps que ne le laissent entendre les études françaises consacrées à ce thème, dans la mesure où ces dernières ne prennent pas suffisamment en compte les biais de sélection et d'endogénéité qui affectent l'estimation des effets de l'âge sur les performances scolaires. En instrumentant l'âge auquel les élèves passent les épreuves d'évaluation par leur mois de naissance, nous montrons qu'en CP, les élèves nés en fin d'année obtiennent en moyenne des scores inférieurs de 70 % d'un écart-type à ceux de leurs camarades nés en janvier ; lorsqu'ils passent les épreuves du Brevet, l'écart reste significatif et égal à 17 % d'un écart-type. Nos analyses semblent également indiquer que les moindres performances scolaires obtenues par les élèves nés en fin d'année s'expliquent principalement par leur plus jeune âge au moment du passage des épreuves d'évaluation et que cette pénalité n'est ni atténuée, ni renforcée par le fait qu'ils redoublent plus souvent. Nos estimations sont en effet très proches de celles obtenues par Crawford et al. (2007) pour le Royaume-Uni, où le redoublement est très peu pratiqué.

La seconde contribution de ce travail empirique est de montrer que dans le contexte institutionnel français, les effets du mois de naissance sur la réussite scolaire ne sont

pas transitoires, comme on le considère généralement, mais affectent durablement les trajectoires scolaires des élèves. Nos résultats indiquent en effet que le mois de naissance influence non seulement le taux de redoublement, mais également la probabilité d'être orienté dans la voie professionnelle après le collège. Le fait d'être né en décembre plutôt qu'en janvier multiplie par deux la probabilité de redoubler au cours de sa scolarité primaire et augmente de trois points la probabilité d'être orienté en lycée professionnel après la classe de troisième.

Nous montrons cependant qu'en dépit des effets du mois de naissance sur les trajectoires scolaires, le handicap subi à l'école par les élèves nés en fin d'année ne semble pas les défavoriser de manière importante dans leur vie professionnelle. Nos estimations indiquent en effet que leur situation d'emploi n'est pas significativement affectée et que l'écart de rémunération séparant les salariés nés en décembre des salariés nés en janvier est faible puisque compris entre 1,5 et 3 %. Nous interprétons ces faibles écarts salariaux, qui proviennent en grande partie de l'entrée plus tardive sur le marché du travail des personnes nées en fin d'année, comme la conséquence du fait que le mois de naissance influence moins le niveau de diplôme que le type de formation suivie (générale ou professionnelle). Notre étude révèle néanmoins qu'il existe une pénalité salariale spécifiquement liée au plus faible niveau de qualification des hommes nés en fin d'année. Celle-ci se traduit par un écart de salaire égal à 1,5 % entre ceux qui sont nés en décembre et ceux qui sont nés en janvier.

Ces résultats suggèrent que si le mois de naissance des individus n'exerce pas une influence déterminante sur leur destin professionnel, il pénalise néanmoins le parcours scolaire de ceux qui sont nés en fin d'année. Dans la mesure où on ne choisit pas sa date de naissance, il paraîtrait souhaitable que le système éducatif prévoie un certain nombre de mécanismes institutionnels pour corriger les inégalités scolaires liés à ce facteur. Parmi les solutions envisageables figure l'application d'un coefficient compensateur pour redresser les résultats scolaires des élèves en fonction de leur mois de naissance. Un tel dispositif, qui aurait surtout vocation à être utilisé à l'école primaire, permettrait de renforcer l'équité du système éducatif français d'une manière à la fois simple et peu coûteuse pour les finances publiques.

Limites et perspectives

Bien qu'ils apportent un éclairage nouveau sur les politiques éducatives qui ont accompagné la démocratisation scolaire en France, nos travaux comportent un certain nombre de limites qu'il convient d'examiner pour mieux tracer les perspectives qu'ils tracent pour les recherches futures.

Mieux comprendre les facteurs qui conditionnent l'efficacité de l'obligation scolaire. Notre étude comparative des réformes de l'obligation scolaire en France et au Royaume-Uni indique clairement qu'il ne suffit pas d'allonger la scolarité obligatoire pour améliorer le destin professionnel des individus les moins qualifiés. Un certain nombre d'investigations complémentaires sont néanmoins nécessaires pour cerner plus précisément les facteurs qui conditionnent l'efficacité de ce type de réformes.

Malgré la richesse des données utilisées, notre estimation de l'impact salarial de la réforme Berthoin en France et de l'*Education Act* au Royaume-Uni n'est pas suffisamment précise pour déterminer si les rendements de l'obligation scolaire varient en fonction du sexe des individus. Cette question est d'importance, dans la mesure où nos analyses ont révélé que l'impact de l'allongement de la scolarité obligatoire au Royaume-Uni sur les diplômés avait été beaucoup plus fort pour les femmes que pour les hommes. Pour améliorer la précision de nos estimations, il faudrait envisager d'utiliser des sources administratives sur les salaires plutôt que les renseignements issus des enquêtes Emploi. En France, des données salariales quasi exhaustives existent dans les fichiers des Déclarations annuelles de données sociales (DADS), mais leur accès fait l'objet de sévères restrictions. Au Royaume-Uni, des informations précises sur les salaires ont été collectées dans le *New Earnings Survey Panel Dataset* (NESPD), mais ce fichier ne peut être consulté que dans des centres d'accès sécurisé payants.

Une seconde limite de notre travail concerne l'interprétation de la nature exacte des rendements que nous mesurons. En effet, si nos estimations semblent indiquer que l'impact salarial positif de l'*Education Act* au Royaume-Uni provient en partie d'une amélioration des compétences cognitives des individus, elles n'excluent pas totalement une explication fondée sur le pur effet de signal des diplômes. Or les

implications de ces deux interprétations sont très différentes du point de vue des recommandations de politique publique. En effet, dans l'hypothèse où l'allongement de la scolarité obligatoire amène un certain nombre d'élèves à acquérir, dans le cadre de la préparation des diplômes, des compétences qui sont valorisées par les employeurs, une telle politique constitue un outil privilégié pour augmenter la productivité des salariés. En revanche, si les rendements salariaux de l'obligation scolaire ne reflètent aucune amélioration des compétences, mais seulement le fait que les employeurs utilisent le diplôme comme indicateur de productivité, alors une telle politique ne peut améliorer le sort des individus avec un faible niveau d'études que si elle ne modifie qu'à la marge la structure des qualifications dans la population. Pour trancher plus nettement entre ces deux explications, il serait nécessaire d'utiliser des sources d'information complémentaires sur les compétences des individus à l'âge adulte, portant sur des échantillons plus larges que ceux utilisés dans la présente étude.

Notons également que les conclusions de notre travail ne sont pas facilement généralisables à l'ensemble des réformes de l'obligation scolaire qui ont fait l'objet d'évaluations empiriques. Elles ne permettent pas, en particulier, de comprendre les raisons de l'impact salarial positif de l'extension de l'obligation scolaire aux États-Unis ou au Canada, où la scolarité secondaire n'est pas structurée par une succession de diplômes, comme c'est le cas en France ou au Royaume-Uni. Ce paradoxe apparent invite à approfondir l'analyse des facteurs institutionnels qui conditionnent l'efficacité de l'obligation scolaire comme instrument de réduction des inégalités. Une telle perspective permettrait d'éclairer plus précisément l'influence de mécanismes aussi divers que le redoublement, l'orientation ou l'organisation des cursus éducatifs sur les bénéfices d'une scolarité obligatoire plus longue.

Enfin, notre étude ne fournit qu'une évaluation partielle des rendements de l'obligation scolaire, dans la mesure où nous avons adopté une perspective exclusivement centrée sur les salaires et l'emploi. Or, la littérature empirique consacrée aux rendements de l'éducation a révélé que les effets positifs de l'obligation scolaire ne se limitent pas nécessairement au marché du travail, mais peuvent également s'étendre à la criminalité, à la santé ou encore aux comportements civiques (Lochner et Moretti, 2004; Milligan et al., 2004; Oreopoulos, 2007). De ce point de vue, il serait

intéressant d'évaluer si l'absence de rendements salariaux de la réforme Berthoin en France ne masque pas d'autres bénéfices qui pourraient en partie compenser son échec sur le front de la lutte contre les inégalités socio-économiques.

Apporter de nouveaux éléments empiriques pour mieux appréhender les réformes de la sectorisation. Le modèle que nous avons développé pour analyser le lien entre politiques de sectorisation, stratification résidentielle et choix scolaire souffre d'un certain nombre de limites qui atténuent la portée de ses conclusions. Bien qu'il tienne compte des caractéristiques saillantes du système éducatif français, ce modèle donne sans doute une représentation trop simpliste des établissements privés, qui fonctionnent comme des clubs de parents au niveau d'un quartier. Si cette approche permet de reproduire assez fidèlement la localisation des collèges privés à Paris, elle ne rend pas compte du fait que, dans la pratique, le recrutement de ces établissements s'étend bien au-delà des quartiers dans lesquels ils sont localisés. Pour rendre nos hypothèses plus réalistes, il faudrait autoriser les écoles privées à recruter leurs élèves dans l'ensemble de la ville tout en introduisant des droits d'inscription, même modestes, et des coûts de mobilité.

Notre modélisation révèle par ailleurs que la connaissance de la disponibilité des ménages à payer pour un meilleur établissement scolaire ne suffit pas pour mesurer l'impact du mode de recrutement des établissements sur les inégalités scolaires. D'autres paramètres, tels que la corrélation entre revenu parental et réussite éducative des enfants ou encore l'importance quantitative des effets de pairs à l'école, doivent également être connus pour évaluer pleinement les effets redistributifs des réformes de la carte scolaire. En l'absence de consensus empirique sur le sujet, notre modèle tend à privilégier certaines hypothèses, ce qui limite le degré de généralité de nos résultats.

Enfin, les réformes auxquelles nous nous sommes intéressés dans le cadre de cette étude théorique correspondent à des cas volontairement limités (suppression totale de la sectorisation, redécoupage intégral des secteurs scolaires) qui, bien qu'ils se prêtent assez facilement à des simulations numériques mettant en lumière les principaux mécanismes à l'œuvre, ne rendent pas suffisamment compte de la diversité des options

envisageables pour réformer le fonctionnement de la carte scolaire. C'est pourquoi nous envisageons d'enrichir notre modèle pour étudier un certain nombre de réformes intermédiaires, telles que le redécoupage partiel des secteurs scolaires ou encore l'assouplissement de la sectorisation assorti de règles précises d'admission incluant, par exemple, la prise en compte de critères sociaux. Notre approche théorique gagnerait également en réalisme si la modélisation des interactions entre choix résidentiels et choix scolaires était envisagée dans une perspective dynamique permettant d'intégrer les évolutions démographiques et les transformations du parc immobilier.

Notre contribution empirique est quant à elle limitée par la nature des données dont nous disposons. Si notre étude permet de quantifier précisément l'impact de la performance des collèges publics sur les prix immobiliers tout en prenant en compte le rôle spécifique joué par le secteur privé, elle ne permet pas de déterminer précisément les facteurs qui sont pris en compte par les parents lorsqu'ils évaluent la qualité des établissements scolaires. En particulier, la forte corrélation qui unit la composition sociale des collèges aux résultats scolaires des élèves ne permet pas de mesurer la part qui revient à chacune de ces composantes dans la valorisation des établissements par les parents. Notre travail gagnerait en particulier à mieux quantifier les facteurs qui motivent le choix d'une scolarisation dans le secteur privé, afin de déterminer si les préférences spécifiques pour ce type d'enseignement l'emportent sur les considérations de niveau et de composition sociale des établissements. Enfin, nos estimations ne nous permettent d'évaluer la disponibilité à payer pour une meilleure école qu'au sein de la sous-population des ménages propriétaires. Idéalement, il faudrait aussi estimer la valeur de ce paramètre pour la sous-population des ménages locataires, à partir de données géolocalisées sur les loyers parisiens³¹.

Plus fondamentalement, l'analyse empirique de l'impact de la performances des établissements scolaires sur les prix immobiliers n'apporte qu'un éclairage indirect sur la question plus générale du lien entre ségrégation résidentielle, politiques de sectorisation et inégalités scolaires. Une manière plus directe d'étudier cette question consisterait à évaluer l'impact des réformes qui ont modifié le mode de recrutement des

³¹L'Observatoire des loyers de l'agglomération parisienne (OLAP) collecte depuis le début des années 1990 des informations sur le niveau des loyers et les caractéristiques d'un logement sur vingt du parc locatif privé situé dans la zone centrale de l'agglomération. Malheureusement, cet organisme n'a pas souhaité nous autoriser à exploiter ses données.

établissements scolaires. Les exemples étrangers, en particulier les expérimentations qui visent à donner plus de choix aux parents, nous apportent des éléments intéressants pour appréhender les conséquences de telles réformes (Rouse, 1998 ; Cullen et al., 2005 ; Lavy, 2006 ; Hsieh et Urquiola, 2006 ; Angrist et al., 2006), mais l'absence de consensus sur leurs effets suggère que les spécificités des différents systèmes éducatifs limitent les possibilités de comparaison. En France, un certain nombre d'expérimentations qui ont eu lieu dans le passé pourraient être évaluées pour faire progresser nos connaissances sur ces questions, qu'il s'agisse de l'assouplissement de la carte scolaire dans certains arrondissements de Paris au début des années 1990 ou de l'augmentation du nombre de dérogations à la carte scolaire accordées depuis la rentrée 2007. Bien que les enseignements à tirer de ces expérimentations soient potentiellement très riches, leur évaluation se heurte à des obstacles institutionnels et politiques très puissants, qui les empêchent aujourd'hui d'être réalisées³².

Approfondir l'analyse des enjeux éducatifs et économique de la date de naissance. Le travail empirique que nous avons consacré à l'impact économique du mois de naissance a montré qu'en influençant les performances éducatives et les trajectoires suivies à l'école, ce facteur constituait un puissant facteur d'inégalités scolaires. Si nos évaluations semblent indiquer que ce phénomène provient essentiellement des effets de l'âge sur la maturité intellectuelle des élèves, elles ne permettent pas cependant de quantifier précisément le rôle joué par les autres mécanismes potentiellement à l'œuvre.

Nous n'avons pu apprécier l'impact spécifique du redoublement sur les performances scolaires qu'indirectement, à travers la confrontation de nos estimations avec celles réalisées au Royaume-Uni par Crawford et al. (2007). Cette comparaison semble indiquer que le redoublement ne tend ni à renforcer, ni à atténuer la pénalité subie par les élèves nés en fin d'année par rapport aux élèves nés en début d'année. Une manière plus directe de tester ce résultat pourrait consister à tirer parti de la ré-

³²Malgré le soutien apporté par le Haut commissariat aux Solidarités actives contre la pauvreté au projet que nous avons déposé pour évaluer l'impact des politiques d'affectation des élèves aux établissements scolaires sur les trajectoires éducatives, le marché immobilier et la stratification résidentielle et scolaire à Paris, l'accès au fichier des élèves de l'Académie de Paris nous a toujours été refusé par le Rectorat.

forme du collège intervenue en 1995, qui eut pour conséquence de diviser par deux le taux de redoublement dans le premier cycle de l'enseignement secondaire. Même si les effets du mois de naissance sur le redoublement s'exercent surtout pendant la scolarité primaire, nos estimations indiquent qu'ils restent significatifs au collège. Il paraît donc envisageable d'exploiter la diminution exogène du taux de redoublement induite par la réforme de 1995 pour évaluer le rôle spécifique du redoublement dans les effets du mois de naissance sur les performances scolaires. Ce travail ne pourra être réalisé que lorsque les fichiers des résultats aux examens du diplôme du Brevet seront disponibles depuis le début des années 1990, et non pas seulement à partir de 2004 comme c'est le cas aujourd'hui.

Un autre aspect que nous n'avons pu analyser dans le cadre de cette étude concerne la possibilité que les élèves nés en fin d'année soient pénalisés non seulement par leur plus jeune âge au moment du passage évaluations, mais également par le fait qu'ils sont moins âgés que leurs camarades, cette position d'« infériorité » étant susceptible de nuire à leur apprentissage. Les travaux empiriques qui ont cherché à mesurer cet effet exploitent le fait que la position d'un élève dans la hiérarchie des âges ne dépend pas uniquement de sa date de naissance, mais également de celle de ses camarades. Du fait du nombre relativement réduit d'élèves par classe, la position relative d'un élève est donc susceptible de varier indépendamment de sa date de naissance. Les estimations réalisées en adoptant cette stratégie semblent indiquer que cet effet d'âge relatif sur les résultats scolaires est beaucoup plus faible que celui de l'âge absolu auquel les élèves passent les évaluations. Pour vérifier si ce résultat est valable également dans le cas français, il faudrait pouvoir apparier les résultats aux évaluations avec les données indiquant la composition exacte de la classe d'un élève, ce qui n'est possible que si l'on dispose d'un identifiant individuel commun aux différentes sources. Cet identifiant existe, mais ne figure pas pour le moment dans les fichiers accessibles aux chercheurs.

Il serait par ailleurs souhaitable de mieux préciser les raisons pour lesquelles les inégalités scolaires engendrées par la date de naissance ne se reflètent que très marginalement dans les salaires et la situation d'emploi des individus. L'hypothèse avancée dans notre étude, selon laquelle le mois de naissance influence moins le niveau de di-

plôme que le type de formation suivie, mériterait d'être davantage étayée. De ce point de vue, une analyse plus fine de l'impact du mois de naissance sur les trajectoires professionnelles pourrait être réalisée en utilisant les données issues de l'appariement des Déclarations annuelles de données sociales (DADS) et de l'Échantillon Démographique Permanent (EDP), dans la mesure où ces fichiers permettent de réaliser des analyses longitudinales et contiennent des informations précises sur le niveau de formation des salariés, leur secteur d'activité, leur niveau de rémunération ainsi que sur les caractéristiques de leur employeur.

Enfin, ce travail de recherche invite à réfléchir aux dispositifs institutionnels qui pourraient être mis en place pour corriger les inégalités scolaires qui résultent du mois de naissance. Si l'application d'un barème correcteur pour redresser les résultats scolaires des élèves scolarisés dans l'enseignement primaire constitue une piste prometteuse, ce type de mécanisme ne peut prétendre corriger les difficultés d'apprentissage liées à l'âge relatif des élèves d'une même classe. Pour cela, d'autres instruments doivent être mobilisés, mais l'incertitude quant à l'efficacité des différentes options envisageables (soutien spécifique pour les élèves les plus jeunes, rentrées décalées, classes constituées en fonction du mois de naissance des élèves, etc.) rend nécessaire leur évaluation à l'issue d'expérimentations ponctuelles.

D'une manière générale, le développement des expérimentations dans le domaine des politiques scolaires constitue une voie de recherche prometteuse pour enrichir notre connaissance des enjeux économiques de l'éducation. En plein essor dans les pays anglo-saxons, cette approche fait timidement son apparition en France, mais reste essentiellement cantonnée aux politiques d'emploi. On ne peut que souhaiter qu'elle se généralise aux politiques éducatives et vienne nourrir la réflexion collective sur l'avenir de l'école.

Bibliographie

- Acemoglu, D. and Angrist, J. (2000). “How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws”. In Bernanke, B. and Rogoff, K., editors, *NBER Macroeconomics Annual 2000*, volume 15, pages 9–59. MIT Press, Cambridge, MA.
- Angrist, J., Bettinger, E., and Kremer, M. (2006). “Long-Term Educational Consequences of Secondary School Vouchers : Evidence from Administrative Records in Colombia”. *American Economic Review*, 96(3), p. 847–862.
- Angrist, J. and Krueger, A. (1991). “Does Compulsory School Attendance Affect Schooling and Earnings”. *Quarterly Journal of Economics*, 106(4), p. 979–1014.
- Angrist, J. and Krueger, A. (1992). “The Effect of Age at School Entry on Educational Attainment : An Application of Instrumental Variables With Moments From Two Samples”. *Journal of the American Statistical Association*, 87(418), p. 328–336.
- Angrist, J., Lang, D., and Oreopoulos, P. (2007). “The Effect of High School Matriculation Awards : Evidence from Randomized Trials”. IZA Discussion Paper No. 3134.
- Angrist, J. and Lang, K. (2004). “Does School Integration Generate Peer Effects? Evidence from Boston’s Metco Program”. *American Economic Review*, 94(5), p. 1613–1634.
- Angrist, J. and Lavy, V. (1999). “Using Maimonides’ Rule To Estimate The Effect Of Class Size On Scholastic Achievement”. *Quarterly Journal of Economics*, 114(2), p. 533–575.

- Angrist, J. and Lavy, V. (2002). “The Effect of High School Matriculation Awards : Evidence from Randomized Trials”. NBER Working Paper No. 9389.
- Arrow, K. (1973). “Higher Education as a Filter”. *Journal of Public Economics*, 2, p. 193–216.
- Bayer, P., Ferreira, F., and McMillan, R. (2003). “An Unified Framework for Measuring Preferences for School and Neighborhoods”. Yale University Economic Growth Center Discussion Paper No. 872.
- Bayer, P., Ferreira, F., and McMillan, R. (2005). “Tiebout Sorting, Social Multipliers and the Demand for School Quality”. NBER Working Paper No. 10871.
- Beaud, S. (2003). *80 % au bac... et après ?* La Découverte, Paris.
- Beddard, K. and Dhuey, E. (2006). “The Persistence of Early Childhood Maturity : International Evidence of Long-Run Age Effects”. *Quarterly Journal of Economics*, 121(4), p. 1437–1472.
- Bells, J. and Daniels, S. (1990). “Are Summer Born Children Disadvantaged ? the Birthday Effect in Education”. *Oxford Review of Education*, 16, p. 67–80.
- Bickel, D., Zigmond, M., and Strayhorn, J. (1991). “Chronological age and entrance to first grade : Effects on elementary school success”. *Early Childhood Research Quarterly*, 6(2), p. 10–117.
- Black, S. (1999). “Do Better Schools Matter ? Parental Valuation of Elementary Education”. *Quarterly Journal of Economics*, 114(2), p. 577–599.
- Black, S., Devereux, P., and Salvanes, K. (2008). “Staying in the Classroom and Out of the Maternity Ward ? The Effect of Compulsory Schooling Laws on Teenage Births”. *The Economic Journal*, 118(530), p. 1025–1054.
- Bénabou, R. (1993). “Workings of a City : Location, Education and Production”. *Quarterly Journal of Economics*, 108(3), p. 619–652.
- Bénabou, R. (1996a). “Equity and Efficiency in Human Capital Investment : The Local Connection”. *The Review of Economic Studies*, 63(2), p. 237–264.

- Bénabou, R. (1996b). “Heterogeneity, Stratification and Growth : Macroeconomic Implications of Community Structure and School Finance”. *American Economic Review*, 86(3), p. 584–609.
- Bénabou, R., Kramarz, F., and Prost, C. (2005). “The French *Zones d’Education Prioritaire* : Much Ado About Nothing?”. *Economics of Education Review*, forthcoming.
- Bobak, M. and Gjonca, A. (2001). “The Seasonality of Live Birth is Strongly Influenced by Socio-demographic Factors”. *Human Reproduction*, 16(7), p. 1512–1517.
- Bogart, W. and Cromwell, B. (2000). “How Much is a Neighborhood School Worth?”. *Journal of Urban Economics*, 47(2), p. 280–305.
- Bound, J. and Jaeger, D. (2000). “Do Compulsory School Attendance Laws Explain the Association between Quarter of Births and Earnings?”. In Bernanke, B. and Rogoff, K., editors, *Research in Labor Economics : Worker Well-Being*, pages 83–108. JAI, New York.
- Brasington, D. and Haurin, D. (2006). “Educational Outcomes and House Values : a Test of the Value Added Approach”. *Journal of Regional Science*, 56(2), p. 245–268.
- Bressoux, P. and Kramarz, F. and Prost, C. (2008). “Teachers’ Training, Class Size, and Students’ Outcomes : Learning from Administrative Forecasting Mistakes”. Unpublished manuscript.
- Brodaty, T. O., Gary-Bobo, R., and Prieto, A. (2008). “Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages”. Unpublished manuscript.
- Brunello, G. and Miniaci, R. (1999). “The Economic Return to Schooling for Italian Men. An Evaluation Based on Instrumental Variables”. *Labour Economics*, 6(4), p. 509–519.
- Bumpass, L., Rindfuss, R., and Janosik, R. (1978). “Age and Pace of Subsequent Fertility”. *Demography*, 15, p. 75–86.

- Cahan, S. and Davis, D. (1987). “A Between-Grade-Levels Approach to the Investigation of the Absolute Effects of Schooling on Achievement”. *American Educational Research Journal*, 24, p. 1–12.
- Caille, J.-P. et Rosenwald, F. (2006). “Les inégalités de réussite à l’école : construction et évolution”. In *France, portrait social – Édition 2006*, pages 115–137. Insee, Paris.
- Callan, T. and Harmon, C. (1999). “The Economic Return to Schooling in Ireland”. *Labour Economics*, 6(4), p. 543–550.
- Card, D. (1999). “The Causal Effect of Education on Earnings”. In Ashenfelter, O. and Card, D., editors, *Handbook of Labour Economics*, volume 3a. North Holland, Amsterdam.
- Chay, K., McEwan, P., and Urquiola, M. (2007). “The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools”. *American Economic Review*, 95(4), p. 1237–1258.
- Cheng, M.-Y., Fan, J., and Marron, J. (1997). “On Automatic Boundary Corrections”. *Annals of Statistics*, 25(4), p. 1691–1708.
- Cheshire, P. and Sheppard, S. (2004). “Capitalising the Value of Free Schools : the Impact of Supply Characteristics and Uncertainty”. *Economic Journal*, 114(499), p. F397–F424.
- Chiswick, B. (1973). “Schooling, Screening, and Income”. In Solmon, L. and Taubman, P., editors, *Does College Matter ?* Academic Press, New York.
- Clapp, J., Nanda, A., and Ross, S. (2008). “Which School Attributes Matter ? The Influence of School District Performance and Demographic Composition on Property Values”. *Journal of Urban Economics*, 63(2), p. 451–466.
- Cosnefroy, O. et Rocher, T. (2004). “Le redoublement au cours de la scolarité obligatoire : nouvelles analyses, mêmes constats”. *Éducation et Formation*, 70, p. 73–82.
- Crawford, C., Dearden, L., and Meghir, C. (2007). “When You Are Born Matters : The Impact of Date of Birth on Child Cognitive Outcomes in England”. Institute for Fiscal Studies.

- Crosser, S. (1991). "Summer Birth Date Children : Kindergarten Entrance Age and Academic Achievement". *Journal of Educational Research*, 84, p. 140–146.
- Cullen, J., Jacob, B., and Levitt, S. (2005). "The Impact of School Choice on Student Outcomes : an Analysis of the Chicago Public Schools". *Journal of Public Economics*, 89(5-6), p. 729–760.
- Datar, A. (2006). "Does Delaying Kindergarten Entrance Give Children a Head Start?". *Economics of Education Review*, 25, p. 43–62.
- de Bartolome, C. (1990). "Equilibrium and Inefficiency in a Community Model With Peer Groups Effects". *Journal of Political Economy*, 98(1), p. 110–133.
- de Monredon, S. (2008). "Les élèves du premier degré à la rentrée 2007 dans les écoles publiques et privées". Note d'information du Ministère de l'Éducation nationale.
- de Saboulin, M. (1978). "Un nouveau calendrier des première naissances". *Économie et Statistique*, 100.
- Dearden, L., Emmerson, C., Frayne, C., and Meghir, C. (2005). "Conditional Cash Transfers and School Dropout Rates". forthcoming *Journal of Human Resources*.
- Dearden, L., Machin, S., and Reed, H. (1997). "Intergenerational Mobility in Britain". *Economic Journal*, 107(440), p. 47–66.
- Devereux, P. and Hart, R. (2008). "Forced to be Rich? Returns to Compulsory Schooling in Britain". Stirling Economics Discussion Paper 2008-02.
- Doblhammer, G. and Vaupel, J. (2001). "Life Span Depends on Month of Birth". *Proceedings of the National Academy of Sciences*, 98(5), p. 2934–2939.
- Downes, T. and Zabel, J. (2002). "The Impact of School Characteristics on House Prices : Chicago 1987-1991". *Journal of Urban Economics*, 52(1), p. 1–25.
- Dubet, F. et Duru-Bellat, M. (2000). *L'hypocrisie scolaire. Pour un collège enfin démocratique*. Seuil, collection « Épreuve des faits », Paris.

- Dupaquier, M. (1976). “Le mouvement saisonnier des naissances en France de 1853 à 1973”. Thèse de démographie.
- Duru-Bellat, M. (2006). *L'inflation scolaire*. La République des Idées/Seuil, Paris.
- Eicker, F. (1963). “Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regressions”. *Annals of Mathematical Statistics*, 34, p. 447–456.
- Eide, E. and Showalter, M. (2001). “The effect of Grade Retention on Educational and Labor Market Outcomes”. *Economics of Education Review*, 20(6), p. 563–576.
- Epple, D. and Romano, R. (1998). “Competition between Private and Public Schools, Vouchers, and Peer-Group Effects”. *American Economic Review*, 88(1), p. 33–62.
- Epple, D. and Romano, R. (2003). “Neighborhood Schools, Choice, and the distribution of Education Benefits”. In Hoxby, C. M., editor, *The Economics of School Choice*, pages 227–286. University of Chicago Press, Chicago, IL.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, London.
- Fernández, R. (2003). “Sorting, Education and Inequality”. In Dewatripont, Hansen, and Turnovsky, editor, *Advances in Economics and Econometrics : Theory and Applications, Eighth World Congress, Volume II*, pages 1–40. Cambridge University Press, Cambridge, MA.
- Fernández, R. and Rogerson, R. (1996). “Income Distribution, Communities and the Quality of Public Education”. *Quarterly Journal of Economics*, 111(1), p. 135–164.
- Fernández, R. and Rogerson, R. (1998). “Income Distribution and Public Education : A Dynamic Quantitative Analysis of School Finance Reform”. *American Economic Review*, 88(4), p. 813–833.
- Fernández, R. and Rogerson, R. (2003). “Equity and Resources : An Analysis of Education Finance Systems”. *Journal of Political Economy*, 111(4), p. 858–897.

- Ferreira, M. M. (2007). "Estimating the Effects of Private School Vouchers in Multi-district Economies". *American Economic Review*, 97(3), p. 789–817.
- Ferrier, J. (2003). "L'avance et le retard scolaires à l'école élémentaire et au collège". *Les Cahiers de l'Éducation*, 23, p. 9–18.
- Fertig, M. and Kluge, J. (2005). "The Effect of Age at School Entry on Educational Attainment in Germany". IZA Discussion Paper No. 1507.
- Figlio, D. N. and Lucas, M. (2004). "What's in a Grade? School Report Cards and House Prices". *American Economic Review*, 93(3), p. 591–604.
- Flores-Lagunes, A. and Light, A. (2007). "Interpreting Sheepskin Effects in the Returns to Education". Unpublished manuscript.
- Florin, A., Cosnefroy, O. et Guimard, P. (2004). "Trimestre de naissance et parcours scolaire". *Revue Européenne de Psychologie Appliquée*, 54(4), p. 237–246.
- François, C. et Poupeau, F. (2004). "L'évitement scolaire et les classes moyennes à Paris". *Éducation et Sociétés*, 14, p. 51–66.
- Fredriksson, P. and Öckert, B. (2006). "Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance". IFAU Working Paper 2006 :12.
- Gary-Bobo, R. et Mahjoub, M.-B. (2006). "Estimation of Class-Size Effects, using 'Maimonides' Rule" : The Case of French Junior High Schools". Unpublished manuscript.
- Gary-Bobo, R. et Trannoy, A. (2008). "Efficient Tuition Fees and Examinations". *Journal of the European Economic Association* forthcoming.
- Gibbons, S. and Machin, S. (2003). "Valuing English Primary Schools". *Journal of Urban Economics*, 53(2), p. 197–219.
- Gibbons, S. and Machin, S. (2006). "Paying for Primary Schools : Supply Constraints, School Popularity or Congestion". *The Economic Journal*, 116(510), p. C77–C92.

- Gledhill, J., Ford, T., and Goodman, R. (2002). “Public Schooling for Young Children and Maternal Labor Supply”. *American Economic Review*, 92(1), p. 307–322.
- Goldin, C. and Katz, L. (2002). “The Power of the Pill : Oral Contraceptives and Women’s Career and Marriage”. *Journal of Political Economy*, 110(4), p. 730–770.
- Gomes-Neto, J. and Hanushek, E. (1994). “Causes and Consequences of Grade Repetition : Evidence from Brazil”. *Economic Development and Cultural Change*, 43(1), p. 117–148.
- Goux, D. and Maurin, É. (2008). “Preschool Enrolment, Mothers’ Participation in the Labour Market, and Children’s Subsequent Outcomes”. Unpublished manuscript.
- Goux, D. et Nouveau, C. (2007). “Expansion scolaire et insertion professionnelle : une évaluation de l’ouverture du lycée depuis le début des années 1980”. In Cohen, D., editor, *Une jeunesse difficile : portrait économique et social de la jeunesse française*, pages 61–81. Cerepremap, Paris.
- Graue, M. and DiPerna, J. (1990). “The Effects of Chronological Age, Gender, and Delay of Entry on Academic Achievement and Retention : Implications for Academic Redshirting.”. *Psychology in the Schools*, 27, p. 260–263.
- Graue, M. and DiPerna, J. (2002). “Redshirting and Early Retention : Who Gets the ‘Gift of Time’ and what are its Outcomes?”. *American Educational Research Journal*, 37(2), p. 509–534.
- Gravel, N. and Trannoy, A. (2006). “Measuring the Social Value of Local Public Goods : an Empirical Analysis Within Paris Metropolitan Area ”. *Applied Economics*, 38(16), p. 1945–1961.
- Grenet, J. (2008). “PISA : une enquête bancale?”. *La Vie des Idées*.
- Gurgand, M. (2005). *Économie de l’éducation*. La Découverte, collection « Repères », Paris.
- Gurgand, M. et Maurin, É. (2007). “Démocratisation du secondaire et inégalités salariales : l’expansion éducative d’après-guerre”. In Cohen, D., editor, *Une jeu-*

- nesse difficile : portrait économique et social de la jeunesse française*, pages 82–105. Cepremap, Paris.
- Hahn, J., Todd, P., and van der Klauw, W. (2001). “Identification and Estimation of Treatment Effects with a Regression Discontinuity Design”. *Econometrica*, 69(1), p. 201–209.
- Harmon, C. and Walker, I. (1995). “Estimates of the Economic Return to Schooling for the United Kingdom”. *American Economic Review*, 85(5), p. 278–1286.
- Hastings, J., Van Weelden, R., and Weinstein, J. (2007). “Preferences, Information and Parental Choice Behavior in Public School Choice”. NBER Working Paper No. 12995.
- Hoxby, C. (2000). “Does Competition Among Public Schools Benefit Students and Taxpayers?”. *American Economic Review*, 90(5), p. 789–817.
- Hoxby, C. (2001). “Peer Effects in the Classroom : Learning from Gender and Race Variation”. NBER Working Paper No. 7867.
- Hoxby, C. (2003). “School Choice and School Productivity : Could School Choice be a Tide that Lifts All Boats?”. In Hoxby, C., editor, *The Economics of School Choice*, pages 287–352. University of Chicago Press, Chicago, IL.
- Hoxby, C. and Weingarth, G. (2005). “School Reassignment and the structure of peer effects”. Working Paper.
- Hsieh, C.-T. and Urquiola, M. (2006). “The Effects of Generalized School Choice on Achievement and Stratification : Evidence from Chile’s Voucher Program”. *Journal of Public Economics*, 90(8-9), p. 1477–1503.
- Huber, P. (1967). “The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions”. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, p. 221–233.
- Imbens, G. and Angrist, J. (1994). “Identification and Estimation of Local Average Treatment Effects”. *Econometrica*, 62(2), p. 467–475.

- Imbens, G. and Lemieux, T. (2007). “Regression Discontinuity Designs : A Guide to Practice”. *Journal of Econometrics*, 127(2), p. 615–635.
- J., B., D., J., and R., B. (1995). “Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak”. *Journal of the American Statistical Association*, 40(430), p. 443–450.
- Jacob, B. and Lefgren, L. (2004). “Remedial Education and Student Achievement : A Regression Discontinuity Analysis”. *The Review of Economics and Statistics*, 86(1), p. 226–244.
- Jaspar, M.-L. (2008). “Les élèves du second degré à la rentrée 2007 dans les établissements publics et privés”. Note d’information du Ministère de l’Education nationale, Paris.
- Jones, M. and Mandeville, K. (1990). “The Effect of Age at School Entry on Reading Achievement Scores among South Carolina Students”. *Remedial and Special Education*, 11, p. 56–62.
- Jürges, H. and Schneider, K. (2004). “What Can Go Wrong Will Go Wrong : Birthday Effects and Early Tracking in the German School System”. CESifo working paper No. 2055.
- Kane, T., Riegg, S., and Staiger, D. (2006). “School Quality, Neighborhoods and Housing Prices”. *American Law and Economic Review*, 8(3), p. 183–212.
- Kane, T. and Staiger, D. (2002). “The Promise and Pitfalls of Using Imprecise School Accountability Measures”. *Journal of Economic Perspectives*, 16(4), p. 91–114.
- Kane, T., Staiger, D., and Samms, G. (2003). “School Accountability Ratings and Housing Values”. *Brookings-Wharton Papers on Urban Affairs*, pages 83–137.
- Kramarz, F. and Machin, S. and Ouazad, A. (2007). “What Makes a Test Score ? The Respective Contributions of Pupils, Schools, and Peers in Achievement in English Primary Education”. Unpublished manuscript.

- LaFerrère, A. (2005). “Hedonic Housing Price Indexes : the French Experience”. In *The Economics of School Choice*, volume 21, pages 271–287. Bank for International Settlements.
- Lang, K. and Kropp, D. (1986). “Human Capital vs Sorting : the Effect of Compulsory Attendance Laws”. *Quarterly Journal of Economics*, 101(3), p. 609–624.
- Lange, F. and Topel, R. (2006). “The Social Returns to Education and Human Capital”. In Hanushek, E. and Welch, F., editors, *Handbook of the Economics of Education*, volume 1. North-Holland, Amsterdam.
- Langer, P., Kalk, J., and Searls, D. (1984). “Age of Admission and Trends in Achievement : A Comparison of Blacks and Caucasians”. *American Educational research Journal*, 21(1), p. 61–78.
- Langouët, G. and Léger, A. (1997). *Le choix des familles : école publique ou école privée*. Paris, Éditions Fabert.
- Lavy, V. (2006). “From Forced Busing to Free Choice in Public Schools : Individual and General Equilibrium Effects”. NBER working paper No. 11969.
- Le Bastard-Landrier, S. (2005). “L’expérience subjective des élèves de seconde : Influence sur les résultats scolaires et les voeux d’orientation”. *L’orientation scolaire et professionnelle*, 34(2), p. 143–164.
- Lee, D. (2008). “Randomized Experiments from Non-random Selection in the U.S. House Elections”. *Journal of Econometrics*, 142(2), p. 675–697.
- Lee, D. and Card, D. (2008). “Regression discontinuity inference with specification error”. *Journal of Econometrics*, 142(2), p. 655–674.
- Lee, D., Moretti, E., and Butler, M. (2004). “Do Voters Affect or Elect Policies? Evidence from the U.S. House”. *Quarterly Journal of Economics*, 119(3), p. 807–859.
- Lefranc, A. and Trannoy, A. (2004). “Intergenerational Earnings Mobility in France : is France more mobile than the US?”. IDEP Working Paper No. 0401.

- León, A. (2004). “The Effect of Education on Fertility : Evidence From Compulsory Schooling Laws”. Unpublished manuscript.
- Leridon, H. (1988). “Les conceptions du premier janvier ou les « étrennes de septembre »”. *Population*, 3.
- Leuven, E., Lindhal, M., Oosterbeek, H., and Webbink, D. (2006). “Expanding schooling opportunities for 4-year-olds”. IZA Discussion Paper No. 2434.
- Levine, J. and Plug, E. (1999). “Instrumenting Education and the Returns to Schooling in the Netherlands”. *Labour Economics*, 6(4), p. 521–534.
- Lleras-Muney, A. (2005). “The Relationship Between Education and Adult Mortality in the United States”. *Review of Economic Studies*, 72(1), p. 189–221.
- Lochner, L. and Moretti, E. (2004). “The Effect of Education on Crime : Evidence from Prison Inmates, Arrests, and Self-Reports”. *American Economic Review*, 94(1), p. 155–189.
- Ludwig, J. and Miller, D. (2005). “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design”. NBER Working Paper No. 11702.
- Ludwig, J. and Miller, D. (2007). “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design”. *Quarterly Journal of Economics*, 122(1), p. 159–208.
- Lutinier, B. (1987). “La saison des mariages”. *Economie et Statistique*, 204, p. 21–28.
- MacKinnon, J. and White, H. (1985). “Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties”. *Journal of Econometrics*, 29(3), p. 305–325.
- Marchand, O. et Thélot, C. (1997). *Le travail en France : 1800-2000*. Nathan, Paris.
- Marini, M. and Hodson, P. (1981). “Effect of the Timing of Marriage and First Birth on the Spacing of Subsequent Births”. *Demography*, 18(4), p. 529–548.

- Martin, R., Foels, P., Clanton, G., and Moon, K. (2004). "Season of Birth is Related to Child Retention Rates, Achievement, and Rate of Diagnosis of Specific LD". *Journal of Learning Disabilities*, 37(4), p. 307–317.
- Martínez Mora, F. (2006). "The Existence of Non-Elite Private schools". *Journal of Public Economics*, 90(8-9), p. 1505–1518.
- Maurin, É. (2007). *La nouvelle question scolaire. Les bénéfices de la démocratisation*. Le Seuil, Paris.
- May, D., Kundert, D., and Brent, D. (1995). "Does Delayed School Entry Reduce Later Grade Retentions and Use of Special Education Services?". *Remedial and Special Education*, 16, p. 288–294.
- Mayer, S. and Knutson, D. (1998). "Does Age at Enrollment in First Grade Affect Children's Cognitive Test Scores?". *Compare*, 31(3), p. 381–391.
- Mazumder, B. (2003). "Revised Estimates of Intergenerational Income Mobility in the United States". Federal Reserve Bank of Chicago, WP 2003-16.
- McClelland, M., Morrison, F., and Holmes, D. (2000). "Children at Risk for Early Academic Problems : The Role of Learning Related Social Skills". *Early Childhood Research Quarterly*, 7, p. 155–174.
- Meghir, C. and Pålme, M. (2005). "Educational reform, Ability and Parental Background". *American Economic Review*, 95(1), p. 414–424.
- MEN-DEP (2007). *L'État de l'école*, No. 17. Ministère de l'Éducation nationale, Paris.
- Milligan, K., Moretti, E., and Oreopoulos, P. (2004). "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom". *Journal of Public Economics*, 88(9-10), p. 1667–1695.
- Ministère de l'Éducation Nationale (1969). *Tableaux de l'Éducation Nationale. Statistiques rétrospectives 1958-1968*. Service Central des Statistiques et Sondages, Paris.

- Ministère de l'Éducation Nationale (1978). *Tableaux des Enseignements et de la Formation. Statistiques rétrospectives 1967-1977*. Service Central des Statistiques et Sondages, Paris.
- Mizala, A., Romaguera, P., and Urquiola, M. (2007). "Socioeconomic Status or Noise? Tradeoffs in the Generation of School Quality Information". *Journal of Development Economics*, 84(1), p. 61–75.
- Mizala, A. and Urquiola, M. (2008). "School markets : The impact of Information Approximating Schools' Effectiveness". NBER Working Paper No. 13676.
- Morrison, F., Griffith, E., and Alberts, D. (1997). "Nature-Nurture in the Classroom : Entrance Age, School Readiness, and Learning in Children". *Developmental Psychology*, 11, p. 56–62.
- Nechyba, T. (1997). Existence of equilibrium and stratification in local and hierarchical tiebout economies with property taxes and voting. *Economic Theory*, 10.
- Nechyba, T. (1999). "School Finance Induced Migration and Stratification Patterns : The Impact of Private School Vouchers". *Journal of Public Economic Theory*, 1(1), p. 5–50.
- Nechyba, T. (2000). "Mobility, Targeting and Private School Vouchers". *American Economic Review*, 90(1), p. 130–146.
- Nechyba, T. (2003). "Introducing School Choice into Multi-District School Systems". In Hoxby, C., editor, *The Economics of School Choice*, pages 145–194. University of Chicago Press, Chicago, IL.
- Oberti, M. et Prêteceille, E. (2004). "Les classes moyennes et la ségrégation". *Éducation et Sociétés*, 14, p. 135–153.
- Odry, D. (dir.) (2006). *L'orientation, c'est l'affaire de tous*. Centre régional de documentation pédagogique (CRDP) de l'académie d'Amiens, Amiens.
- Oosterbeek, H. and Webbink, D. (2007). "Wage Effects of an Extra Year of Basic Vocational Education". *Economics of Education Review*, 26, p. 408–419.

- Oppenheimer, V. (1988). “A Theory of Marriage Timing”. *American Journal of Sociology*, 94, p. 563–591.
- Oreopoulos, P. (2006a). “Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter”. *American Economic Review*, 96(1), p. 152–175.
- Oreopoulos, P. (2006b). “The Compelling Effects of Compulsory Schooling : Evidence from Canada”. *Canadian Journal of Economics*, 39(1), p. 22–52.
- Oreopoulos, P. (2007). “Do Dropouts Drop out too Soon ? Wealth, Health and Happiness from Compulsory Schooling”. *Journal of Public Economics*, 91(11-12), p. 2213–2229.
- Oreopoulos, P., Page, M., and Stevens, A. (2006). “The Intergenerational Effects of Compulsory Schooling”. *Journal of Labor Economics*, 24(4), p. 729–760.
- Piketty, T. et Valdenaire, M. (2006). *L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français – Estimations à partir du panel primaire 1997 et du panel secondaire 1995*. Ministère de l'Éducation nationale, Paris.
- Pischke, J.-S. and Manning, A. (2006). “Comprehensive Versus Selective Schooling in England and Wales : What Do We Know?”. NBER Working Paper No. 12176.
- Pischke, J.-S. and von Wachter, T. (2006). “Zero Returns to Compulsory Schooling in Germany : Evidence and Interpretation”. NBER Working Paper No. 11414.
- Pischke, J.-S. and von Wachter, T. (2008). “Zero Returns to Compulsory Schooling in Germany : Evidence and Interpretation”. *Review of Economics and Statistics*, 90(3), p. 592–598.
- Plug, E. (2001). “Season of Birth, Schooling and Earnings”. *Journal of Economic Psychology*, 22(5), p. 641–660.
- Pons, E. and Gonzalo, M. (2002). “Returns to Schooling in Spain : How Reliable are Instrumental Variable Estimates”. *Labour*, 16(4), p. 747–770.

- Prioux, F. (1988). “Mouvement saisonnier des naissances : influence du rang et de la légitimité dans quelques pays d’Europe occidentale”. *Population*, 3, p. 587–609.
- Prost, A. (2004). *Histoire de l’enseignement et de l’éducation, tome IV, L’école et la famille dans une société en mutation depuis 1930*. Éditions Perrin, Paris.
- Puhani, P. and Weber, A. (2005). “Does the Early Bird Catch the Worm ? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany”. IZA Discussion Paper No. 1827.
- Reback, R. (2005). “House Prices and the Provision of Local Public Services : Capitalization Under School Choice Programs”. *Journal of Urban Economics*, 57(2), p. 275–301.
- Retherford, R. and Sewell, W. (1989). “How Intelligence Affects Fertility”. *Intelligence*, 13, p. 169–185.
- Régnier-Loilier, A. (2004). “Les naissances ‘au fil des saisons’ ”. *Revue Européenne des Sciences Sociales*, 42(129).
- Rosenwald, F. (2006). Les scolarités des filles et des garçons. *Administration et éducation*, 110, p. 25–37.
- Rothstein, J. (2006). Good principals or good peers : Parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4), p. 1333–1350.
- Rouse, C. (1998). “Private School Vouchers and Student Achievement : An Evaluation of the Milwaukee Parental Choice Program”. *Quarterly Journal of Economics*, 113(2), p. 553–602.
- Sandron, F. (1988). *Les naissances de la pleine lune et autres curiosités démographiques*. Éditions L’Harmattan, Paris.
- Sharp, C. (1995). “What’s Age Got to Do With it ? A Study of Patterns of School Entry and the Impact of Season of Birth on School Attainment”. *Education Research*, 37, p. 255–265.

- Solon, G. (1992). “Intergenerational income mobility in the United States”. *American Economic Review*, 82(3), p. 393–408.
- Spence, M. (1973). “Job Market Signaling”. *Quarterly Journal of Economics*, 87, p. 355–374.
- Stiglitz, J. (1975). “The Theory of ‘Screening’, Education, and the Distribution of Income”. *American Economic Review*, 65, p. 283–300.
- Stipek, D. and Byler, P. (1995). “Considering Primary School Effectiveness : an Analysis of 1992 Key Stage 1 Results”. *The Curriculum Journal*, 6, p. 279–295.
- Stipek, D. and Byler, P. (2001). “Academic Achievement and Social Behaviors Associated with Age of Entry into Kindergarten”. *Journal of Applied Developmental Psychology*, 22, p. 175–189.
- Strøm, B. (2004). “Student Achievement and Birthday Effects”. Unpublished manuscript.
- Sweetland, J. and De Simone, P. (1987). “Age of Entry, Sex, and Academic Achievement in Elementary School Children”. *Psychology in the Schools*, 234, p. 406–412.
- Thélot, C. et Vallet, L.-A. (2000). “La réduction des inégalités sociales devant l’école depuis le début du siècle”. *Économie et Statistique*, 334, p. 3–32.
- Tiebout, C. (1956). “A Pure Theory of Local Public Expenditures”. *Journal of Political Economy*, 64(5), p. 416–424.
- Van Zanten, A. (2001). *L’école de la périphérie. Scolarité et ségrégation en banlieue*. Presses Universitaires de France, Paris.
- Vieira, J. (1999). “Returns to Education in Portugal”. *Labour Economics*, 6(4), p. 535–541.
- Weiss, A. (1983). “A Sorting-cum-Learning Model of Education”. *Journal of Political Economy*, 91, p. 420–442.
- White, H. (1980). “A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity”. *Econometrica*, 48, p. 817–838.

Liste des tableaux

1.1	Summary statistics (France). Source: <i>Enquête Emploi</i> (1990-2002).	49
1.2	Summary statistics (United Kingdom). Source: Quarterly Labour Force Survey (1993-2004).	51
1.3	Naive OLS estimates of the returns to schooling in France and Britain, calculated separately for female and male workers. Sources: <i>Enquête Emploi</i> (1990-2002) and Quarterly Labour Force Survey (1993-2004).	52
1.4	Estimated impact of the French Berthoin reform (1967) on the distribution of school leaving ages of women and men. Source: <i>Enquête Emploi</i> (1990-2002).	58
1.5	Estimated impact of the French Berthoin reform (1967) on schooling and hourly wages (in 2005 euros) of female and male workers. Source: <i>Enquête Emploi</i> (1990-2002).	61
1.6	Estimated impact of the British Education Act (1973) on the distribution of school leaving ages of women and men. Source: Quarterly Labour Force Survey (1993-2004).	65
1.7	Estimated impact of the British Education Act (1973) on schooling and hourly wages (in 2005 pounds) of female and male workers. Source: Quarterly Labour Force Survey (1993-2004).	67
1.8	Inclusion of age effects in the estimation of the impact of the French Berthoin reform (1967) and of the British Education Act (1973) on schooling and hourly wages of female and male workers. Sources: <i>Enquête Emploi</i> (1990-2002) and Quarterly Labour Force Survey (1993-2004).	73

1.9	Sensitivity to the choice of bandwidth: estimated impact of the French Berthoin reform (1967) on schooling and hourly wages (in 2005 euros) of female and male workers using local linear regression with different bandwidth values. Source: <i>Enquête Emploi</i> (1990-2002).	75
1.10	Sensitivity to the choice of bandwidth: estimated impact of the British Education Act (1973) on schooling and hourly wages (in 2005 pounds) of female and male workers using local linear regression with different bandwidth values. Source: Quarterly Labour Force Survey (1993-2004).	76
1.11	Testing for jumps in schooling and hourly wages of French female and male workers at non-discontinuity points. Source: <i>Enquête Emploi</i> (1990-2002).	77
1.12	Testing for jumps in schooling and hourly wages of British female and male workers at non-discontinuity points. Source: Quarterly Labour Force Survey (1993-2004).	78
1.13	Estimated impact of the French Berthoin reform (1967) on the probability of being in employment and on the probability of being self-employed, for women and men separately. Source: <i>Enquête Emploi</i> (1990-2002).	84
1.14	Estimated impact of the British Education Act (1973) on the probability of being in employment and on the probability of being self-employed, for women and men separately. Source: Quarterly Labour Force Survey (1993-2004).	87
1.15	Educational career of the 1952 and 1953 cohorts between the age of 6-7 and the age of 18-19. Source: <i>Tableaux de l'Éducation Nationale</i> , Ministère de l'Éducation nationale (1969, 1978).	94
1.16	Estimated impact of the French Berthoin reform (1967) on the distribution of the highest held academic credential of women and men. Source: <i>Enquête Emploi</i> (1990-2002).	100
1.17	Estimated impact of the British Education Act (1973) on the distribution of the highest held academic credential of women and men. Source: Quarterly Labour Force Survey (1993-2004).	101

1.18	Estimated impact of the British Education Act (1973) on the literacy scores of female and male participants to the 1994 International Adult Literacy Survey.	108
2.1	Simulation example: residential and educational segregation outcomes from alternative school enrollment policies.	165
2.2	Simulation example: distribution of educational gains induced by school zone reassignments and open enrollment in the presence or absence of a private sector.	166
2.3	Calibration for Paris: residential and educational segregation outcomes from alternative school enrollment policies.	180
2.4	Calibration for Paris: distribution of educational gains induced by school zone reassignments and open enrollment.	181
3.1	Summary statistics: Public and private middle schools in Paris in 2004. Sources: <i>IPES</i> dataset (2004) and <i>OCEAN</i> national examinations database (2004).	217
3.2	Correlation between the three standardized indexes of public middle school performance. Sources: <i>OCEAN</i> (2004) national examinations database and <i>SCOLARITE</i> pupil database (1997-2004).	218
3.3	Summary statistics: characteristics of flats located within 250, 300 and 350 meters from a school attendance boundary and sold during school years 1997 to 2004. Source: <i>BIEN</i> dataset.	223
3.4	Summary statistics: Neighborhood characteristics and average public middle school performance at the district level for all sales and sales located within 250, 300 and 350 meters of a school attendance boundary, sold during school years 1997 to 2004. Sources: 1999 National Census and school catchment areas in academic year 1998-1999. . . .	225
3.5	Regressions results - Naive estimates of the impact of middle school performance on housing prices. School years 1997-2004.	227

3.6	Regression of cross-boundary housing price differentials on corresponding school performance differentials. Sales are matched across public middle school attendance boundaries. School years 1997-2004.	229
3.7	Estimation of the impact of school performance on housing prices including the full set of school attendance boundaries dummies. School years 1997-2004.	233
3.8	Comparison of flat features on the “good” and “bad” side of school attendance boundaries (in terms of the school’s average <i>DNB</i> exam score in 2004). School years 1997-2004.	235
3.9	Comparison of socio-demographic characteristics of neighborhoods located on the “good” and “bad” side of school attendance boundaries (in terms of the school’s average <i>DNB</i> exam score in 2004). School years 1997-2004.	237
3.10	Comparison of the occupation of the household head (computed separately for families and households without children) in neighborhoods located on the “good” and “bad” side of school attendance boundaries where the school performance differential is larger than half a standard deviation(in terms of the school’s average <i>DNB</i> exam score in 2004). School years 1997-2004.	239
3.11	Location of the school attended by public and private school Parisian pupils with respect to their arrondissement of residence. School year 2003-2004.	242
3.12	Regression of cross-boundary housing price differentials on corresponding school performance differentials by quartile of proximity to the closest private middle school. Sales are matched across public middle school attendance boundaries. School years 1997-2004.	248
3.13	Regression of cross-boundary housing price differentials on corresponding school performance differentials by quartile of private school density. Sales are matched across public middle school attendance boundaries. School years 1997-2004.	250

3.14	Comparison of the socio-demographic characteristics of the four quartiles of private school proximity for sales located within 250 meters of a public middle school attendance boundary. School years 1997-2004.	252
3.15	Comparison of the socio-demographic characteristics of the four quartiles of private school density for sales located within 250 meters of a public middle school attendance boundary. School years 1997-2004.	253
3.16	Example of a hedonic regression for zone 1 (<i>arrondissements</i> 1 to 4), used to convert observed housing prices into the “typical” flat equivalence prices. School years 1997-2004.	259
4.1	Impact de l’âge absolu (en mois) sur les notes obtenues aux évaluations de CP, de CE2, de sixième, au contrôle continu et à l’examen terminal du Diplôme national du Brevet, et à l’épreuve anticipée de français du Baccalauréat. Instrument : différence d’âge séparant l’élève des individus nés en décembre de son année de naissance. Sources : Panels primaire (1997) et secondaire (1995) de l’Éducation nationale et Base DNB (2004).	314
4.2	Analyse par sous-groupes d’élèves de l’impact de l’âge absolu (en mois) sur les notes obtenues aux évaluations de CP, de CE2, de sixième, au contrôle continu et à l’examen terminal du Diplôme national du Brevet, et à l’épreuve anticipée de français du Baccalauréat. Instrument : différence d’âge séparant un élève de ses camarades nés en décembre de son année de naissance. Sources : Panels primaire (1997) et secondaire (1995) de l’Éducation nationale et Base DNB (2004).	317
4.3	Comparaison des effets de l’âge (en mois) sur les performances scolaires en France et au Royaume-Uni, en fonction de l’âge des élèves (en années). Sources : calculs de l’auteur pour la France (cf. tableau 4.1) et estimations de Crawford et al. (2007) pour le Royaume-Uni.	321
4.4	Estimation Probit de l’impact marginal du mois de naissance sur la probabilité d’avoir au moins un an de retard scolaire en fonction de l’âge et du milieu social. Sources : Panel primaire de l’Éducation nationale (1997) et Base Scolarité (2004-2005).	326

4.5	Impact du mois de naissance sur l'orientation des élèves en fin de troisième. Estimations réalisées à partir d'un modèle Probit multinomial. Sources : Bases Scolarité 2000-2001 à 2004-2005.	332
4.6	Impact du mois de naissance sur l'orientation des élèves en fin de troisième. Analyse par sous-groupes d'élèves. Estimations réalisées à partir d'un modèle Probit multinomial. Sources : Bases Scolarité 2000-2001 à 2004-2005.	333
4.7	Impact du mois de naissance sur l'âge de fin d'études (en années). Individus nés entre 1945 et 1965. Source : Enquête Emploi (1990-2002).	338
4.8	Impact du mois de naissance sur le dernier diplôme obtenu. Estimations réalisées à partir d'un modèle Probit multinomial. Individus nés entre 1945 et 1965. Source : Enquête Emploi (1990-2002).	344
4.9	Impact du mois de naissance sur la situation d'emploi des individus. Estimations réalisées à partir d'un modèle Probit. Individus nés entre le 1 ^{er} janvier 1945 et le 31 décembre 1965. Source : Enquête Emploi (1990-2002).	346
4.10	Impact du mois de naissance sur le salaire horaire (en euros de 2005). Individus nés entre le 1 ^{er} janvier 1945 et le 31 décembre 1965. Source : Enquête Emploi (1990-2002).	348

Table des figures

1.1	Evolution of the distribution of school leaving ages in France (women and men). Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	55
1.2	Evolution of the distribution of school leaving ages in France, calculated for women and men separately. Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	56
1.3	Impact of the French 1967 Berthoin reform on the average age left full-time education, calculated for women and men separately. Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	60
1.4	Evolution of the distribution of school leaving ages in the United Kingdom (women and men). Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).	63
1.5	Evolution of the distribution of school leaving ages in the United Kingdom, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).	64
1.6	Impact of the British 1973 Education Act on the average age left full-time education, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).	66
1.7	Impact of the French 1967 Berthoin reform on hourly wages (in 2005 euros), calculated for women and men separately. Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	69
1.8	Impact of the British 1973 Education Act on hourly wages (in 2005 pounds), calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).	71

1.9	Impact of the French 1967 Berthoin reform on the probability of being in employment, calculated for women and men separately. Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	83
1.10	Impact of the British 1973 Education Act on the probability of being in employment, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004). . . .	86
1.11	Impact of the French 1967 Berthoin reform on the probability of being self-employed, calculated for women and men separately. Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	88
1.12	Impact of the British 1973 Education Act on the probability of being self-employed, calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004). . . .	89
1.13	Total school enrollment of French school cohorts 1952 and 1953 every year from the age of 6-7 to the age of 18-19. Source: <i>Tableaux de l'Éducation Nationale</i> , Ministère de l'Éducation nationale (1969, 1978).	92
1.14	Impact of the French 1967 Berthoin reform on the distribution of academic credentials, calculated for female and male workers separately. Cohorts 1944-1962. Source: <i>Enquête Emploi</i> (1990-2002).	96
1.15	Impact of the British 1973 Education Act on the distribution of academic credentials, calculated for female and male workers separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).	97
1.16	Impact of the British 1973 Education Act on the fraction of individuals holding no academic qualification and on the fraction holding a junior secondary schooling certificate (GCE O-Level or CSE), calculated for women and men separately. Cohorts 1949-1967. Source: Quarterly Labour Force Survey (1993-2004).	99
1.17	Literacy scores of British women and men on three scales (prose, document and quantitative literacy). Individuals born between 1947 and 1969. Source: International Adult Literacy Survey (1994).	107
2.1	Single crossing condition: indifference curves in the housing quality (q) and housing price (p) space.	122

2.2	Layout of the city used in the simulations: a checkerboard of 100 neighborhoods and 16 school zones.	146
2.3	Layout of school zones, neighborhoods, blocks and cells in the city. . .	147
2.4	Simulating school zone reassignments: the new pattern of school attendance boundaries.	149
2.5	Distribution of the log of intrinsic neighborhood qualities, the log of family income and the log of pupil ability in the sample of 10,000 households used in the simulations.	160
2.6	Simulation example: spatial distribution of neighborhood intrinsic quality.	161
2.7	Simulation example: equilibrium spatial distribution of school quality and final cell desirability under strict school zoning.	163
2.8	Simulation example: relationship between income and cell desirability, intrinsic neighborhood quality and mean ability in the local public school under strict school zoning.	164
2.9	Simulating school zone reassignments: comparing initial school quality to equilibrium school quality.	168
2.10	Simulating strict school zoning with private schools: equilibrium public and private school quality.	169
2.11	Comparing the size of the private sector under strict school zoning, before and after the reassignments of school zones.	171
2.12	Comparing public and private school quality under strict school zoning immediately after the reassignments of school zones and in the final equilibrium.	172
2.13	Simulation example: spatial distribution of the estimated quintiles of intrinsic neighborhood quality in the city of Paris.	177
2.14	Calibrations for Paris. distribution of educational gains induced by school zone reassignments and open enrollment, dy deciles of income and ability. Parameter values: $\lambda=0.6$. $\rho=0.5$. $\beta=0.5$. $s_0=0.4$	183
3.1	Matching sales across boundaries: an example.	197
3.2	School catchment areas in Paris during school year 2003-2004.	208

3.3	School catchment areas in the 19 th <i>arrondissement</i> of Paris. School year 2003-2004.	209
3.4	School catchment areas and <i>arrondissements</i> boundaries in Paris in 2004.	211
3.5	School catchment areas and boundaries of high school districts in Paris in 2004.	211
3.6	Spatial distribution of the quartiles of public middle school performance as measured by the average <i>Diplôme National du Brevet</i> exam score in 2004. Source: <i>OCEAN</i> national examinations database (2004). 220	
3.7	Spatial distribution of the quartiles of public middle school performance as measured by the fraction of pupils admitted into general curriculum high schools (averaged over the period 1997-2004). Source: <i>SCOLARITE</i> database (1997-2004).	220
3.8	Spatial distribution of the quartiles of public middle school performance as measured by the fraction of pupils from privileged socio-economic backgrounds (averaged over the period 1997-2004). Source: <i>SCOLARITE</i> database (1997-2004).	221
3.9	Spatial distribution of the quartiles of average price per square meter (in 2004 euros) within each public middle school catchment area in school year 2003-2004.	221
3.10	Map of local private middle school proximity in Paris.	245
3.11	Map of local private middle school density in Paris.	245
4.1	Déviation de la distribution des mois de naissance des individus nés à l'étranger et des individus nés en France entre le 1 ^{er} janvier 1900 et le 31 décembre 1998 par rapport à une répartition uniforme des naissances au cours de l'année. Source : Recensement de la population 1999 (au 1/4).	292

4.2	Mouvement saisonnier des naissances par décennie de naissance. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Décennies 1920-1959. Sources : Recensement de la population 1982 et 1999 (au 1/4).	294
4.2	Mouvement saisonnier des naissances par décennie de naissance. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Décennies 1960-1998. Sources : Recensement de la population 1982 et 1999 (au 1/4).	295
4.3	Mouvement saisonnier des naissances en fonction de la catégorie socio-professionnelle de la mère de famille. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Naissances survenues entre le 1 ^{er} janvier 1980 et le 31 décembre 1998. Source : Recensement de la population 1999 (au 1/4).	298
4.3	Mouvement saisonnier des naissances en fonction de la catégorie socio-professionnelle de la mère de famille. Déviation du nombre de naissances constatées chaque mois par rapport à une répartition uniforme des naissances au cours de l'année. Naissances survenues entre le 1 ^{er} janvier 1980 et le 31 décembre 1998. Source : Recensement de la population 1999 (au 1/4).	299
4.4	Salaire moyen (imputé) du père et de la mère des enfants nés entre le 1 ^{er} janvier 1960 et le 31 décembre 1979 en fonction du mois de naissance de l'enfant, rapporté au salaire moyen de l'ensemble des pères et mères de famille. Champ : enfants dont le père ou la mère est salarié. Sources : Recensement de la Population 1982 (au 1/4) et Enquête Emploi (1990-2002).	301

4.5	Salaire moyen (imputé) du père et de la mère des enfants nés entre le 1 ^{er} janvier 1980 et le 31 décembre 1998 en fonction du mois de naissance de l'enfant, rapporté au salaire moyen de l'ensemble des pères et mères de famille. Champ : enfants dont le père ou la mère est salarié. Sources : Recensement de la Population 1999 (au 1/4) et Enquête Emploi (1990-2002).	302
4.6	Durée moyenne (en mois) de la scolarité pré-élémentaire des élèves entrés en CP en 1997, en fonction de leur mois de naissance. Source : Panel primaire de l'Éducation nationale (1997).	305
4.7	Âge théorique et âge réel (en mois) de passage des épreuves d'évaluation de CE2 et de l'examen du Diplôme national du Brevet. Sources : Panel primaire de l'Éducation nationale (1997) et Base DNB (2004).	307
4.8	Effet de l'âge (en mois) sur les performances scolaires (normalisées en unités d'écart-type) à différentes étapes du cursus éducatif. Sources : Panels primaire (1997) et secondaire (1995) de l'Éducation nationale et Base DNB (2004).	319
4.9	Comparaison des effets de l'âge (en mois) sur les performances scolaires en France et au Royaume-Uni, en fonction de l'âge des élèves (en années). Sources : calculs de l'auteur pour la France et estimations de Crawford et al. (2007) pour le Royaume-Uni.	322
4.10	Proportion d'élèves ayant au moins un an de retard scolaire à l'âge de 7, 11 et 15 ans en fonction de leur mois de naissance. Sources : Panel primaire de l'Éducation nationale (1997) et Base Scolarité (2004-2005).	325
4.11	Estimation de l'impact d'être né décembre plutôt qu'en janvier sur la probabilité d'avoir au moins un an de retard scolaire en fonction de l'âge et du milieu social. Sources : Panel primaire de l'Éducation nationale (1997) et Base Scolarité (2004-2005).	327
4.12	Orientation après la classe de troisième des élèves nés en 1986. Sources : Bases Scolarité 2000-2001 à 2004-2005.	330

4.13	Taux de scolarisation des jeunes âgés de 14 à 18 ans en mars de l'année du recensement en fonction de leur mois de naissance. Sources : Recensements de la population 1982 et 1999 (au 1/4).	335
4.14	Dernier diplôme obtenu par les individus nés entre 1945 et 1965 en fonction de leur mois de naissance. Source : Enquête Emploi (1990-2002).	340
4.15	Proportion d'individus nés entre 1945 et 1965 dont le dernier diplôme est soit le BEPC, soit un CAP ou un BEP, soit le Baccalauréat général ou un diplôme universitaire général en fonction de leur année et de leur trimestre de naissance. Source : Recensement de la population 1999 (au 1/4).	342

Résumé : Cette thèse évalue l'impact économique de trois des principales politiques éducatives qui ont accompagné la démocratisation de l'école en France : l'allongement de l'obligation scolaire, la sectorisation des établissements scolaires et l'orientation des élèves. L'objectif de ce travail est de mesurer sur données françaises les conséquences de ces politiques sur les inégalités éducatives, entendues au sens large : inégalités économiques, inégalités socio-spatiales et inégalités liées à la date de naissance des individus.

Nous commençons par analyser l'impact de la réforme Berthoin, qui porta en 1967 l'obligation scolaire de 14 à 16 ans révolus, en la comparant à l'*Education Act* de 1973 qui porta au Royaume-Uni l'âge de fin d'études minimum de 15 à 16 ans. En exploitant les discontinuités législatives engendrées par ces deux réformes, nous montrons que contrairement à sa version britannique, l'allongement de la scolarité obligatoire en France n'a pas amélioré le destin scolaire et professionnel des générations concernées. Nous interprétons ce phénomène comme la conséquence du fait que cette réforme n'a pas permis de réduire significativement la proportion d'individus quittant l'école sans diplôme.

Nous proposons ensuite une analyse théorique et empirique du lien unissant politiques de sectorisation, stratification résidentielle et inégalités scolaires. Nous développons d'abord un modèle théorique destiné à évaluer les propriétés de modalités alternatives d'allocation des élèves aux établissements scolaires : sectorisation stricte, redécoupage de la carte scolaire ou libre choix de son école. Nous nous attachons ensuite à évaluer l'impact de la performance des collèges publics et privés sur le prix des logements à Paris, en comparant les prix des transactions immobilières localisées de part et d'autre des frontières entre secteurs scolaires. Nos résultats indiquent qu'une augmentation d'un écart-type du niveau d'un collège entraîne une augmentation des prix immobiliers comprise entre 1,5 et 2,5 % selon l'indicateur de performance retenu, et que cet écart a tendance à diminuer quand la densité locale du secteur privé augmente.

Nous réalisons enfin une évaluation systématique de l'impact de la date de naissance des individus sur leurs trajectoires scolaires et professionnelles. Nous montrons que le handicap subi à l'école par les élèves nés en fin d'année par rapport à leurs camarades nés en début d'année, lié au fait qu'ils sont les plus jeunes au sein de leur cohorte scolaire, est important et affecte durablement leur destin scolaire en augmentant leur probabilité de redoubler ou d'être orientés dans la voie professionnelle. Nos résultats indiquent cependant qu'en dépit de cette différenciation des trajectoires scolaires en fonction du mois de naissance, les individus nés en fin d'année ne sont pas fortement pénalisés dans leur vie professionnelle par rapport aux individus nés en début d'année.

Discipline : Sciences Économiques (05)

Mots-clés : Éducation ; inégalités ; démocratisation scolaire ; obligation scolaire ; rendements de l'éducation ; sectorisation ; ségrégation résidentielle ; mois de naissance.

Intitulé et adresse du laboratoire : Paris-Jourdan Sciences Économiques (PSE), 48 boulevard Jourdan, 75014 Paris.