ELSEVIER

Journal of Public Economics 88 (2004) 503–518

JOURNAL OF PUBLIC ECONOMICS

www.elsevier.com/locate/econbase

Direct or indirect tax instruments for redistribution: short-run versus long-run

Emmanuel Saez *

University of California at Berkeley and NBER, Department of Economics, 549 Evans Hall #3880, Berkeley, CA 94720, USA

Received 9 January 2002; received in revised form 27 June 2002; accepted 30 October 2002

Abstract

Optimal tax theory has shown that, under simple assumptions, indirect taxation such as production subsidies, tariffs, or differentiated commodity taxation, are sub-optimal and that redistribution should be achieved solely with the direct income tax. However, these important results of optimal tax theory, namely production efficiency and uniform commodity taxation under non-linear income taxation, have been shown to break down when labor taxation is based on income only and when there is imperfect substitution of labor types in the production function. These results in favor of indirect tax instruments are valid in the short-run when skills are exogenous and individuals cannot move from occupation to occupation. In the long-run, it is more realistic to assume that individuals choose their occupation based on the relative after-tax rewards. This paper shows that, in that context, production efficiency and the uniform commodity tax result are restored. Therefore, in a long-run context, direct income taxation should be preferred to indirect tax instruments to raise revenue and achieve redistribution.

JEL classification: H21; H23
Keywords: Optimal taxation; Redistribution; Occupational choices

1. Introduction

The theory of optimal taxation has derived a number of powerful properties of optimal tax structures. First and perhaps most important is the production efficiency result of Diamond and Mirrlees (1971). This result states that the economy should be on its production frontier at the optimum when the government can tax (linearly) all factors

* Tel.: +1-510-642-2239; fax: +1-510-642-6615.
E-mail address: saez@econ.berkeley.edu (E. Saez).

0047-2727/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0047-2727(02)00222-0

(inputs and outputs) at different rates. This result has two very important public policy implications. The public sector should optimize its production decisions using market prices and the government should not use tariffs, production taxes or subsidies because they create production inefficiencies. Second, Atkinson and Stiglitz (1976) showed that there is no need to use commodity taxation when the government can use a non-linear income tax and utility functions are weakly separable between goods and leisure. Atkinson and Stiglitz proved their theorem using a fixed priced model with perfect substitution between different types of labor. These two results combined imply that indirect tax instruments such as production subsidies, tariffs, or differentiated commodity taxation, are sub-optimal and that redistribution ought to be achieved solely with the direct income tax. Third, Diamond and Mirrlees (1971) showed another important result for the theoretical analysis of optimal tax structures, namely that optimal tax formulas are identical when prices of factors are fixed, as in a small open economy, and when prices are variable and derived from a general production function. This result is important because it implies that substitution between inputs in the production function can be ignored when deriving optimal tax formulas. This simplifies considerably the analysis. From now on, we call this result the Tax-Formula result.[1]

However, these three important results of optimal tax theory have been challenged by subsequent studies. Stiglitz (1982) has developed a simple two-type model (skilled and unskilled workers), where the government cannot observe workers' skills and has to base taxation on income only. In that situation, the government cannot impose freely differentiated tax rates on each type of labor as in the Diamond–Mirrlees model and the Tax-Formula result breaks down. In the model of Stiglitz (1982), there is imperfect substitution of labor types in the production function and the optimal tax formulas depend explicitly on the elasticity of substitution between skilled and unskilled labor. Stiglitz' (1982) point is important because it shows that the standard properties of the optimal non-linear income tax model of Mirrlees (1971), such as the zero top result or the positivity of the marginal tax rate, obtained under the assumption perfect substitution between labor types are not robust to the relaxation of this assumption. Recently, Naito (1999) has shown that, in the framework of the Stiglitz (1982) model, the production efficiency result of Diamond and Mirrlees (1971) and the theorem of Atkinson and Stiglitz (1976) on commodity taxation also break down. The production efficiency result breaks down because the government cannot apply differentiated rates on each type of labor and thus the taxation power of the government is restricted compared to the Diamond–Mirrlees model.[2] The Atkinson–Stiglitz Theorem breaks down because of imperfect substitution in labor types.

Therefore, relaxing two questionable assumptions of the standard model is enough to loose the three main results of optimal taxation theory. The first of these two assumptions is perfect substitution of labor inputs in the production function. The second assumption is the possibility to condition wage income tax rates on labor type. From now on, this second assumption is called the labor types observability assumption. Both the Tax-Formula result

[1] This result has received much less attention in the literature than the previous two results because it does not have such important practical policy implications.

[2] Guesnerie (1998) provides an analysis close to Naito (1999) along those lines.

and the production efficiency result of Diamond and Mirrlees (1971) are valid with imperfect substitution of labor types but no longer when the labor types observability assumption is relaxed. The Atkinson–Stiglitz theorem is valid without labor types observability but not when the perfect substitution of labor types assumption is relaxed. Naito's (1999) contribution is also important because it gives a clear sense of how the indirect tax instruments should be used to complement income taxation. When the government cares about redistribution, tariffs on low skill labor intensive goods, production subsidies for low skill intensive goods, or commodity taxes on high skill labor intensive goods, are desirable, as one would have expected.

The present paper argues that the negative results of Stiglitz (1982) and Naito (1999) hinge crucially on the way labor supply responses are modeled. In both of these papers, workers are intrinsically either skilled or unskilled and respond to incentives by varying their hours of work *within* jobs. In that case, indirect tax instruments—a production subsidy in the low skill sector for example—allow the government to target low skill wages without creating adverse incentives for the high skilled. This model might be an accurate description of labor responses in the short run once individuals have made their education decisions and chosen their occupation. However, in the long-run, when relative wages between occupations change, the adjustment does not go through changes in individual hours of work but rather through changes in relative entry levels by occupation. For example, if an industry becomes obsolete because of technological progress, then the wages in that particular industry decline and supply for this type of occupation dwindles. The adjustment of hours or work at the individual level is in this case of second order of importance in the long run. Therefore, in the long run, it seems more natural to assume that individuals choose their job depending on the (after-tax) rewards that each type of job is giving.

This paper shows that, in the context of a job choice model, the three results of optimal taxation, namely the Tax-Formula result, Production Efficiency, and the Atkinson–Stiglitz Theorem, remain valid when both the perfect substitution of labor types and wage type observability assumptions are relaxed. Intuitively, in the long run or occupational choice model and in contrast to the Stiglitz (1982) model, indirect tax instruments increasing low skill wages induce high skill workers to switch to the low skill sector. Thus, indirect tax instruments are not more efficient than the non-linear income tax and do not allow the government to improve welfare. As a result, redistribution should be achieved using solely direct taxation of income. The result of the present paper has important policy implications because it shows that, although tariffs or production subsidies might be socially desirable in the short-run, they cannot be optimal in a long-run context. Therefore, governments with a sufficiently low discount rate should not support these policies. Though intuitively reasonable, the result is not obvious and depends in a precise way on how the behavioral responses to taxation are modeled which, as we argue below, is in principle testable empirically.

The paper proceeds as follows. Section 2 presents a simple example to contrast the desirability of tariffs in the short-run and in the long-run, and displays the central economic intuitions. Section 3 develops a general (long-run) job choice model and shows that this model can be seen as a direct extension of the Diamond and Mirrlees (1971) economy, and shows why the three main results of optimal taxation are valid in that context. Finally, Section 4 offers some concluding remarks.

2. A simple example

This section shows in a very simplified context why tariffs are desirable in the short run but no longer in the long run. I first present the structure of the economy common to both situations.

There are two types of occupations in the economy. A low skill occupation produces a low technology good (for example textile) and a high skill occupation produces a high technology good (for example computers). In each sector, one unit of high (low) skill labor produces one unit of high (low) technology good. Subscript 1 denotes the low technology good or sector and subscript 2 the high technology good or sector. We consider the case of a small open economy which takes as given the international prices of each good $p = (p_1, p_2)$. The small country can impose a tariff t per unit on imports of good 1. Therefore the domestic prices of goods are $q = (q_1, q_2) = (p_1 + t, p_2)$. We assume that the production sectors are competitive and therefore wages rates $w = (w_1, w_2)$ in each sector are equal to domestic prices $q = (q_1, q_2)$. We assume that utility is separable between consumption of goods 1 and 2, and labor choices. All individuals derive the same utility $U(c_1, c_2)$ for consuming goods 1 and 2 in quantity c_1 and c_2. The indirect utility is $v(q,x) = \max U(c_1, c_2)$ subject to $q_1c_1 + q_2c_2 \leq x$, where x denotes after tax income. The government sets an optimal (non-linear) income tax that can be based only on total labor earnings.

As the goal of this section is to contrast the desirability of tariffs in the short-run versus the long-run, we consider two models for labor choices. The first model is a short-run or choice of hours model and the second model is a long-run or occupation choice model. In the short-run, individuals are stuck into an occupation (high skill or low skill) but can vary their labor supply (hours of work) on the job.[3] In the long run however, individuals choose their occupation according to the relative rewards in each occupation. As we think that the hours choice is of second order in the long run, we assume labor supply is fixed and equal to one once a type of job is chosen in the occupational choice model.[4]

2.1. The short-run or choice of hours model

This model is a simplified version of the model of Naito (1996). Therefore, the model is presented quickly and only the intuitions for the results are given. Individuals are either unskilled (type 1) or skilled (type 2). I denote by f the immutable proportion of unskilled workers. Individuals choose their hours of work l, earn w_il and pay taxes T_i according to their type i. Total utility is equal to $V_i = v(q, w_il - T_i) - C(l)$ where $C(l)$ is an increasing and convex function of labor cost. Because the government cannot observe types directly, the income tax (T_1, T_2) must be incentive compatible: skilled workers must be better off working l_2 and earning $w_2l_2 - T_2$ after taxes rather than imitating the unskilled by working w_1l_1/w_2 and earning $w_1l_1 - T_1$. As is standard in the literature, we assume that we are in the normal redistributive case where only this incentive compatibility constraint is binding.

[3] This is the classic discrete type model of optimal taxation developed by Stiglitz (1982).
[4] This occupational model was developed by Piketty (1997) to study optimal income tax issues.

For a given level of tariffs t, the government chooses (l_1, l_2, T_1, T_2) so as to maximize a weighted sum of utilities, $W = \pi_1 f V_1 + \pi_2(1-f)V_2$ (where π_i are positive weights), subject to the incentive compatibility constraint

$$v(q, w_2 l_2 - T_2) - C(l_2) \geq v(q, w_1 l_1 - T_1) - C(w_1 l_1 / w_2), \tag{1}$$

and a budget constraint stating that total taxes collected are at least equal to zero.[5] I denote by C_1 total consumption of good 1 in the economy. As fl_1 is total production of good 1 in the economy, net imports are equal to $C_1 - fl_1$. Therefore, net taxes collected by the tariff t are equal to $t(C_1 - fl_1)$ and the budget constraint of the government is

$$fT_1 + (1-f)T_2 + t(C_1 - fl_1) \geq 0. \tag{2}$$

At the optimum, the incentive compatibility condition (1) is binding. As usual, labor supply of the high skilled is efficient ($C'(l_2) = w_2$) but labor supply of the unskilled is below the efficient level ($C'(l_1) < w_1$). Naito (1996) showed that starting from a situation with no tariffs $t = 0$, imposing a small tariff $\mathrm{d}t > 0$ increases welfare W. An intuitive explanation for this result can be presented as follows.[6]

Suppose that the government increases tariffs by $\mathrm{d}t$, then the government collects $(C_1 - fl_1)\mathrm{d}t$ additional taxes. The tariff can be decomposed into two effects. First, the small tariff increases the price of good 1 by $\mathrm{d}t$ as would a consumption tax $\mathrm{d}t$ on good 1. Second, the tariff increases the wages of the unskilled by $\mathrm{d}t$. Therefore, the tariff is exactly equivalent to a consumption tax $\mathrm{d}t$ on good 1 plus a wage subsidy $\mathrm{d}t$ for the unskilled.[7] The consumption tax part has no first order effect on welfare because of the separability assumption between goods and labor costs. This result is a particular case of the general result of Atkinson and Stiglitz (1976).[8] Therefore, to assess the welfare effect of the tariff, we simply have to assess the welfare effect of the wage subsidy $\mathrm{d}t$ on low skill workers. It is useful to compare the wage subsidy with an income tax cut for the low skilled $\mathrm{d}T_1 = -l_1\mathrm{d}t$. As we start from an optimal income tax, this income tax change has no first order effect on welfare. Let us show why the wage subsidy is superior to the income tax change and hence has a positive first order effect on welfare.

The wage subsidy has the same effect on the utility of the unskilled and the same mechanical effect on tax revenue (ignoring behavioral responses) as the income tax change. Let us see why the wage subsidy does better on incentives than the income tax cut. Eq. (1) shows that the high skilled person mimicking the low skilled does not benefit from the low skill wage subsidy because the high skill wage w_2 is not affected by the subsidy. Intuitively, the wage subsidy allows to target redistribution to the low skilled without affecting the incentives of the high skilled because when a high skilled reduces labor supply to imitate a low skilled person, he remains in the high skill sector and thus

---

[5] Assuming that a given exogenous amount of a tax revenue should be collected would not change the analysis.

[6] Naito derives his result from the formal analysis of the first order conditions.

[7] This decomposition has been introduced by Dixit and Norman (1980, 1986).

[8] This argument is formalized in the proof of Proposition 3 in Section 3.

does not benefit from the wage subsidy. On the other hand, Eq. (1) shows that the high skill mimicking the low skill benefits from the income tax cut $\mathrm{d}T_1$. Therefore, a modification of the income tax in favor of the low skilled is going to affect labor supply of the high skilled as well because the tax schedule is common to both types. Therefore, it is clear that, for incentive reasons, the wage subsidy is preferable to the income tax cut.[9]

2.2. The long-run or occupational choice model

In the long-run model, individuals choose their occupation according to the relative rewards in each occupation and labor supply is fixed (at unity) once a type of job is chosen. Therefore a given individual decides whether to work in an unskilled occupation or a skilled occupation depending on the after-tax incomes $w_1 - T_1$ and $w_2 - T_2$ in each occupation. Individuals differ in their tastes for work in each occupation. It may be easier for example for more educated people to handle a high skill occupation than for less educated people.[10] The total population is normalized to one and the fraction f of people in the low skill job depends on $w_1 - T_1$, $w_2 - T_2$, and the price level q. Behavioral responses are built into the function $f(w_1 - T_1, w_2 - T_2, q)$. Presumably, f is increasing in $w_1 - T_1$ because if after-tax income in the low skill occupation increases while prices and after-tax income in the high skill occupation remain constant, low skill occupations become more attractive and some high skill workers may switch to low skill occupations. Similarly, f is presumably decreasing in $w_2 - T_2$.

The government sets an income tax (T_1, T_2) and can also impose a tariff t on good 1. Production of good 1 is equal to the number f of workers in the low skilled occupation. Total consumption of good 1 is denoted as above by C_1 and thus net imports are equal to $C_1 - f$. Thus, the budget constraint of the government is $fT_1 + (1-f)T_2 + t(C_1 - f) \geq 0$. We assume that the government maximizes a social welfare function W which is a weighted sum of individual utilities subject to the budget constraint. As before, starting from a situation with no tariffs ($t = 0$), we want to know whether imposing a small tariff $\mathrm{d}t$ can improve welfare. As shown above, imposing a tariff $\mathrm{d}t$ is equivalent to imposing a commodity tax $\mathrm{d}t$ on good one and a wage subsidy $\mathrm{d}t$ on low skilled jobs. As in the hours choice model, the small commodity tax has no first order effect on welfare because of separability between consumption and labor choices.

In the present model, workers base their decision on after-tax incomes $w_i - T_i$. Thus increasing the pre-tax wage w_1 by $\mathrm{d}t$ dollars is strictly equivalent to decreasing the income tax T_1 by $\mathrm{d}t$ dollars from the workers' perspective. Obviously, the fiscal cost for the government of a wage subsidy $\mathrm{d}t$ on low skilled workers is equal to a reduction $\mathrm{d}T_1 = -\mathrm{d}t$ of the income tax on low skill workers. Therefore, the wage subsidy $\mathrm{d}t$ is exactly equivalent to a reduction in the income tax $\mathrm{d}T_1 = -\mathrm{d}t$. Consequently, the small tariff can be exactly replicated using the income tax instrument. As the income tax is optimal, a small change around the optimum cannot improve welfare. As a result, the

[9] This can be shown formally using Lagrangian analysis as in Naito (1996).

[10] The precise mathematical structure of the general version of this model is presented in Section 3.

small tariff dt does not improve welfare either, implying that there should be no tariff at the optimum.

2.3. Interpretation

The desirability of tariffs hinges crucially on whether tariffs constitute a new tax instrument that cannot be replicated with the domestic income or commodity taxes. In the simple model we have considered, imposing a tariff on low skilled intensive goods amounts to providing a wage subsidy to low skill occupations which narrows the wage gap between the two types of jobs. In the short-run model, the wage subsidy does not affect the incentives to work of the high skilled. Thus the wage subsidy allows the government to relax the incentive compatibility constraint of the income tax and is therefore desirable. In the long run, however, the low skill wage subsidy will induce high skilled workers to move to low skill occupations. Thus, the wage subsidy does not relax the implicit incentive compatibility constraint and is therefore not desirable.

The short-run model predicts that a low skill wage subsidy would have no effect on labor supply of the high skilled whereas the long-run model predicts that such a wage subsidy would have exactly the same effect as a cut in the income tax for low incomes. Therefore, in order to assess which of the two models is the closest to the real situation, the critical empirical question is whether a wage subsidy to the low skilled would indeed have a smaller effect on incentives of the high skilled than an equivalent cut in the income tax at the low end. Unfortunately, the empirical literature on the labor supply responses to taxation does not offer a direct answer to this question but some elements should be noted.

First, labor supply studies find little cross-sectional relation between hours of work and the wage rate, suggesting that narrowly defined hours of work are not very sensitive to the wage rate (see the surveys by Pencavel (1986) and Blundell and MaCurdy (1999)). However, one should not interpret the Stiglitz (1982) model too narrowly. When the income tax increases, the high skilled might respond by reducing effort on the job producing a significant decrease in earnings but with little change in hours of work. It is important therefore to look at overall earnings and not only hours of work.

Second, studies focusing on overall income or earnings tend indeed to find larger elasticities than hours of work studies (see e.g., Feldstein (1995) for a seminal study of the response of taxable income to tax rates and Gruber and Saez (2002) for a recent survey of this literature). By itself, this piece of evidence is not conclusive for our problem because this type of response could be compatible both with the short-run model and the long-run model. It fits with the short-run model if, as mentioned just above, individuals vary their intensity of work on the job in response to taxation. It fits with the long-run model if individuals vary their labor supply in order to get into different occupations, either by getting promoted more quickly or more slowly within a firm, or by moving to other sectors. The empirical literature does not give much information on this issue. Related to this point however, a strand of the labor supply literature focuses on the response along the extensive margin, namely dropping out or entering the labor force. This margin has been shown to be sensitive to the net-of-tax wage rate, especially for secondary earners (see e.g., Meyer and Rosenbaum, 2001). This suggests that the response along the occupation margin might be more important than the response along the intensity of work on the job, at least for low skilled workers.

Last, following the path-breaking modeling work of Becker (1964), there has been substantial effort devoted to the estimation of the response of education and human capital accumulation choices to the salaries and rewards in different occupations (see e.g., the survey of Freeman, 1986). The literature finds evidence of substantial elasticities of the supply of education with respect to salaries, suggesting that the long-run occupational choice responses are large. Therefore, it is reasonable to think that the response of education to changes in the degree of the progressivity of taxation is also significant and plausibly large.[11]

3. A general model of occupational choice

In this section, we present a general model of occupational choice with many commodities and a general production function. This model generalizes the long-run simple model presented in Section 2.2. The core of the argument is to note that this model is a *generalized* version of the model developed in the seminal paper of Diamond and Mirrlees (1971). As a result, we will show that this occupational model inherits the key properties of optimal tax theory, namely production efficiency, the Tax Formula result, and the Atkinson–Stiglitz theorem. Developing the general model also allows to see more clearly what are the key structural assumptions needed to obtain these optimal tax results.

3.1. The model

In the model, each individual chooses an occupation i among a set of $I+1$ possible occupations $\{0,1,\ldots,I\}$. We assume that occupation 0 is non participation in the labor force. Once an occupation is chosen, hours of work are fixed at unity. As discussed in Section 2, this captures a long-run model of labor supply or skill acquisition. We assume that different occupations do not pay the same wage: $w_i \neq w_j$ for any $i \neq j$. This assumption is almost surely satisfied as we posited a finite number of occupations. Thus, without loss of generality, we assume that $w_0 = 0 < w_1 < \cdots < w_I$. The government sets taxes as a function of income $T_i = T(w_i)$. I denote by $m_i = w_i - T_i$ after-tax income in occupation i. Because wages are different in each occupation, the non-linear income tax amounts to imposing differentiated tax rates on the supply of each occupation. I come back in detail to this important point at the end of the section.

As in the Diamond and Mirrlees (1971) model, in addition to these I labor inputs to production, we assume that there are K consumption goods. We denote by c consumption good vector for a given individual and by $\bar{p}$ and $\bar{q}$ the before and after-tax vector prices of consumption goods. As in Diamond and Mirrlees (1971), there is a general production function defining the production possibility set linking the K consumption goods and the I

[11] Unfortunately, there appears to be no convincing study of the direct effect of income taxation on the supply of education and occupations.

labor inputs. As is standard, I assume that the production function has constant returns to scale or that the government can fully tax pure profits.

We assume that there is continuum of individuals of measure one, and that each individual is indexed by n belonging to a general index set $\mathcal{N}$ possibly multi-dimensional. Individual n maximizes a regular utility function $u^n(c,i)$ which depends on the vector of consumption goods c and on the occupation i chosen subject to the budget constraint $\bar{q}\cdot c\leq m_i$. The individual characteristic n embodies both tastes for work and skills. For example, a hard working or skilled individual will find it easier to choose a more demanding or highly skilled occupation.

In order to see the link between the present model and the standard Diamond–Mirrlees economy, it is useful to treat symmetrically the consumption decision and the occupation choice. Therefore, I denote by $m=(m_0,m_1,\ldots,m_I)$ the vector of after-tax incomes, and by $p=(\bar{p},w)$ and $q=(\bar{q},m)$ the before and after-tax price vector of goods and wages, and by $\pi=(t,-T)=q-p$ the vector of tax rates. I denote by c^n the individual consumption choice vector. Similarly, the occupation choice i of individual n can be denoted as $d^n=-(0,\ldots,0,1,0,\ldots,0)$ where d^n is a vector of size $I+1$ and the unique 1 in vector d^n is the $(i+1)$th element. Therefore, I can summarize total demand of individual n by the $K+I+1$ vector $x^n=(c^n,d^n)$. Individual n picks x^n so as to maximize $u(x^n)$ subject to $q\cdot x^n\leq 0$. Let us denote by $x^n(q)$ the individual (net) demand vector, and by $V^n(q)$ the indirect utility function arising from this maximization program. Put in that form, this model looks identical to a Diamond–Mirrlees economy. The unique and key difference is that the occupation choice d^n belongs to a discrete set (as we assume that individuals cannot choose a convex combination of occupations). As a result, the individual demand $x^n(q)$ is *discontinuous* at points q where the individual is indifferent between two occupations. However, at these switching points, the individual is indifferent between these two occupations and thus gets the same utility in both occupations. As a result, the indirect utility $V^n(q)$ is continuous in q. As we will see, these discontinuities in individual demand functions are going to be smoothed out at the aggregate level under some simple conditions. Total aggregate demand is denoted by $X(q)$ and is defined as

$$X(q)=\int_{\mathcal{N}} x^n(q)\mathrm{d}v(n), \tag{3}$$

where $v(n)$ denotes the distribution of individuals over $\mathcal{N}$. We denote by $C(q)$ the vector of aggregate demand for consumption goods and $f_i(q)$ the fraction of individuals who choose occupation i when facing prices q. It is important to note that the behavioral responses to income taxation are fully embodied in the aggregate supply functions $f_i(q)$. For example, when m_i declines, individuals may move out of occupation i producing a decrease in f_i. By definition, $X(q)=(C(q),-f_0(q),\ldots,-f_I(q))$. The government sets taxes π so as to maximize a weighted sum of individual utilities

$$V(q)=\int_{\mathcal{N}} \mu(n)V^n(q)\mathrm{d}v(n), \tag{4}$$

where $\mu(n)$ is a set of non-negative weights. Exactly as in Diamond and Mirrlees (1971), the government maximizes the social welfare function $V(q)$ subject to a budget constraint and a production constraint. The budget constraint states that the government must collect enough taxes to cover an exogenous vector of public expenses E. The production constraint states that aggregate demand $X(q)$ must be technically feasible. Diamond and Mirrlees (1971) show that it is mathematically equivalent to assume that the government has full control of the production decision. Therefore, the two constraints can be collapsed into a single constraint of the form $g(X(q)+E)\leq 0$ where g is the production function. Competitive behavior implies that the before tax vector price p is equal to the vector of derivatives $Dg(X(q)+E)$.

3.2. Properties of the occupational model

3.2.1. Production efficiency

Diamond and Mirrlees (1971) show that at the optimum q^*, aggregate demand $X(q^*)$ is on the production possibility frontier. This is the Production Efficiency theorem. In addition to the key assumption that the government can tax all inputs and outputs at differentiated tax rates, Diamond and Mirrlees (1971) show that two technical assumptions are needed (Lemma 1, p. 23). First, $V(q)$ must be strictly increasing (or decreasing) for at least one of the price components q_i around q^*. This assumption is satisfied as soon as there is a consumption good that everybody likes or an input good such as labor that everybody dislikes. This is a very mild requirement and we assume that this condition holds.[12] Second, aggregate demand $X(q)$ and the indirect social welfare function $V(q)$ need to be continuous in q. In the Diamond–Mirrlees economy, continuity follows directly from convexity of preferences. In the occupational model of the present paper, continuity of aggregate demand is obtained by assuming that the number of individuals is large and preferences regularly distributed. More precisely:

Assumption 1. For each individual n, preferences are strictly convex and regular enough so that the individual demand function $x^n(q)$ is regular at any point q where individual n is not indifferent between two or more occupation choices.

For any $q\gg 0$, the set N_q of individuals n who are indifferent between two or more occupation choices is of measure zero.

By regular, we mean continuous and differentiable. As discussed above, individual demand is obviously discontinuous at price levels q where the individual switches between occupations (and hence is indifferent between two or more occupations). The first part of Assumption 1 simply states that, outside these singular points, demand functions are well behaved and regular.

The second part of Assumption 1 requires more explanation. It states that these singular points are smoothly distributed *across* individuals so that, for any price configuration q, it is never the case that an atom of individuals is indifferent between two or more occupations. If preferences for work are smoothly distributed across individuals then such

[12] In the particular case with a single consumption good, and possibly multiple labor inputs, the consumption good clearly satisfies this assumption.

an assumption would hold.[13] Assumption 1 is made in order to obtain smooth demand functions in the aggregate. Once again, this should be seen as a technical requirement in order to apply the Production Efficiency theorem and in order to be able to use calculus to characterize the optimum tax system.

Lemma 1. *Under Assumption 1, aggregate demand $X(q)$ and indirect social welfare $V(q)$ are regular in q.*

The technical proof is presented in Appendix A. Using Lemma 1 and the same method as in Diamond and Mirrlees (1971), we obtain

Proposition 1. *Under Assumption 1, at the optimum, there should be production efficiency in the occupational choice model.*

Proof. Suppose by contradiction that $X(q^*)$ is in the interior of the production possibility set. Let us assume that good i is universally desired. Then by reducing the price of good i by a small amount $\mathrm{d}q_i$, by continuity of aggregate demand, $X(q^* - \mathrm{d}q_i)$ would still be in the interior of the production set. As good i is universally desired, decreasing its price would unambiguously increase welfare, $V(q^* - \mathrm{d}q_i) > V(q^*)$, showing that q^* cannot be an optimum. □

As is well known, this result has important policy implications. It implies that the government should not use taxes or subsidies specific to some sectors, or tariffs, and that the public production sector should maximize profits facing the same prices as the private sector. Naito (1999) showed that these results break down in the hours of work model of Stiglitz (1982). Proposition 1 shows that they are restored in the occupational choice model. The key reason why these indirect tax instruments are not desirable in the present model is because the government can fully control the after-tax prices q which are the only prices that matter for individual consumption and labor supply behavior. As a result, manipulating the factor prices p through indirect tax instruments does not affect the incentive constraints and cannot improve welfare.

3.2.2. Tax-formula and optimal income taxation

From the maximization program described above, max $V(q)$ subject to $g(X(q)+E)\leq 0$, and under Assumption 1, we can form the Lagrangian $L = V(q) - \lambda g(X(q)+E)$, where λ is the multiplier of the aggregate budget constraint, and take the first order conditions with respect to each q_k.

Proposition 2. Under Assumption 1, at the optimum, we have, as in Diamond and Mirrlees (1971), for each good k

$$\frac{\partial V}{\partial q_k} = \lambda \sum_j p_j \frac{\partial X_j}{\partial q_k}. \tag{5}$$

[13] For example, in the two-occupation model of Section 2.2, if we assume that there is a cost n of holding the high skill occupation, and the after-tax wage rates are m_1, m_2, then anybody with $n > m_2 - m_1$ will choose the low skill occupation. As a result, if n has a cumulated distribution $F(n)$ in the population, then fraction of unskilled workers is $1 - F(m_2 - m_1)$ which is smooth if the distribution function $F(.)$ is smooth.

Proof. The first order condition with respect to q_k is $\partial V/\partial q_k - \lambda \sum_j \partial g/\partial X_j \cdot \partial X_j/\partial q_k = 0$. As $\partial g/\partial X_j = p_j$, we obtain immediately (5). □

The important property embodied in Eq. (5) and that I called the Tax-Formula result in the Introduction is that the first order condition (5) does not depend explicitly upon the degree of substitution between factors in the production function. Put differently, in the derivation of Eq. (5), one can assume that producer prices p_j are constant.[14] Of course, in any practical application with endogenous prices, the prices p_j at the optimum depend indirectly on the demand for goods and factors and thus on the vector of taxes $\pi = q - p$. However, the Tax-Formula result simplifies considerably the *theoretical* analysis of Eq. (5).

The Tax-Formula result of Proposition 2 is important for optimal income taxation. The occupational model with one consumption good and multiple occupation choices can be seen as a model of optimal non-linear taxation. The government chooses tax rates on each occupation to maximize welfare taking into account the potentially adverse effect of taxation on incentives to work.[15] The literature on non-linear income taxation that grew out of the original contribution of Mirrlees (1971) has considered models where there is perfect substitution of labor inputs in the production function and where the space choice for individual earnings is an interval instead of a discrete set. Piketty (1997) and Saez (2002a) have shown that the discrete occupation choice model leads to formulas of the same form as in the standard continuum case. Therefore, nothing fundamental is changed by assuming a discrete set of earnings outcomes. In that context, Proposition 2 implies that, even if we relax the assumption that labor inputs are perfect substitutes, the same optimal tax formulas apply.

It is important to understand that this is not contradictory with Stiglitz (1982) who shows that relaxing the perfect substitution assumption alters optimal income tax formulas. Stiglitz' (1982) result is obtained in a model where individuals are either skilled or unskilled and vary their labor supply *within* occupations. As a result and as explained above, the non-linear income tax is not equivalent to differentiated tax rates on labor inputs, and thus the Tax-Formula result breaks down. The Mirrlees (1971) continuous model can be interpreted as an hours of work model where skills are fixed[16] in which case optimal tax formulas are not robust to relaxing the assumption of perfect substitution. But the Mirrlees (1971) model can also be interpreted as an occupation choice model where individuals choose their occupation among a continuum. In that case, the non-linear income tax is directly equivalent to differentiated tax rates on each occupation and thus the standard optimal tax formulas are still valid in the case of imperfect substitution.[17]

[14] In that case, the Lagrangian would be $V(q) + \lambda(q-p) \cdot X(q) = V(q) - \lambda p \cdot X(q)$, as $q \cdot X(q) = 0$. Thus the first order condition with respect to q_k taking p as fixed is also (5).

[15] This model was first developed by Piketty (1997) in the case of three occupations and a Rawlsian welfare criterion and extended by Saez (2002a) to any number of occupations and any social welfare function to study the problem of optimal transfers to low incomes.

[16] That was the interpretation given originally in Mirrlees (1971).

[17] As there is a continuum of choices in the Mirrlees (1971) model, one would have to extend the Diamond–Mirrlees model to the case with a continuum of factors. We conjecture that it is possible to do so rigorously and describe regularity conditions that would make Propositions 1 and 2 true in that context. However, as the mathematical degree of complication would be far greater, we think that the finite case provides an approximation good enough and thus do not pursue the continuum case any further.

3.2.3. Complementary commodity taxation

Atkinson and Stiglitz (1976) showed in the context of the Mirrlees (1971) model of income taxation with many consumption goods that in the presence of an optimal non-linear income tax, commodity taxation is useless when utility is weakly separable between leisure and consumption goods. Atkinson and Stiglitz proved their result in a fixed price model (i.e. with perfect substitution of labor types in the production function). As shown by Naito (1999), the Aktinson–Stiglitz theorem breaks down with imperfect substitution in the context of the hours choice model. However, we are going to show that the theorem remains valid in the occupational choice model.

More precisely, the weak separability assumption takes the following form. Individual n has a utility function of the form $U^n(v(c), i)$ where $i = 0, \ldots, I$ is the occupation choice, and $v(c)$ is the sub-utility of consumption goods.[18] We can easily prove the following proposition,

Proposition 3. *In the occupation choice model, the Atkinson–Stiglitz Theorem remains valid with imperfect substitution in labor types. Namely, weak separability implies that the tax on labor income is enough and that there is no need to tax commodities at the optimum.*

Proof. The proof goes in two steps. First, we show that, assuming fixed prices, the Atkinson–Stiglitz theorem goes through in the discrete model we are considering. This part is a purely technical adaptation to the discrete case of the original proof made in the continuum case by Christiansen (1984) and is presented in Appendix A. Second, if we now assume that prices are variables, using Proposition 2, we can apply the Tax-Formula result stating that the first order conditions for optimality with variable prices take the same form as when prices are fixed. From step one, optimal tax formulas imply that commodity tax rates are zero in the fixed price model, therefore, commodity tax rates are also zero with variable prices. □

3.2.4. Caveat

As discussed in the beginning of this section, the key assumption needed to obtain Propositions 1, 2, and 3 is that each income level corresponds to a unique occupation. This assumption is innocuous in the case of a discrete number of occupations. However, in the real world situation, there is a very large number of sectors and occupations, and individuals earning the same income can end up being in very different occupations. In that case, a general income tax cannot replicate any pattern of specific taxes for each occupation type and the formal results of Propositions 1, 2, and 3 break down.

However, it is important to note that this lack of robustness is very different from the one described in Stiglitz (1982) and Naito (1999). Indeed, the results of Stiglitz (1982) and Naito (1999) are important, not only because they show that the normal theory is not robust, but also and mostly because they give a clear sense of how policy should be tilted relative to the normal theory. Namely, the analysis of Naito (1999), as discussed in Section 2.1,

[18] As discussed in Saez (2002b), the fact that the function $v(.)$ is common to all individuals is often overlooked but is as important as the weak separability assumption to obtain the Atkinson–Stiglitz result.

provided an unambiguous justification for providing wage subsidies for industries employing low skilled workers or imposing tariffs on low skilled intensive goods.

In the occupational model, the income tax cannot discriminate between occupations generating the same earnings. Therefore, in that case, occupation specific subsidies constitute a policy instrument more powerful than the income tax. However, in contrast to Naito's (1999) situation, it is not clear whether these subsidies should be tilted toward low earnings occupations rather than higher earnings ones. Therefore, introducing this additional layer of complication does not provide any clear-cut policy recommendation as to what type of goods and industries should be subsidized. Exploring this issue is left for future research.

4. Conclusion

This paper has shown that, in a long-run context where individuals respond to tax incentives through the occupation margin, the key results of optimal tax theory, namely production efficiency, the irrelevance of substitution in production for optimal tax formulas, and the Atkinson–Stiglitz theorem on commodity taxation, are robust to the relaxation of the assumption of fixed priced and perfect observability of labor types. This stands in contrast to a short-run situation where individuals are stuck into their occupations and can only adjust labor supply on the job (Stiglitz, 1982; Naito, 1999). These results have important tax policy implications. In a short-run perspective, indirect tax instruments such as production subsidies on low skilled labor intensive sectors, tariffs or commodity taxes on high skilled labor intensive goods, are desirable to complement the redistribution achieved by progressive income taxation. However, in a long-run perspective, these indirect tax instruments are sub-optimal and redistribution should be achieved solely with the direct progressive income tax.

This set of prescriptions fits well with the real world economy. Unions support tariffs or production subsidies because union members are stuck to occupations. Using redistributing tools which lead to production inefficiencies might then be a helpful way to manipulate wage rates and improve redistribution.[19] On the other hand, in a long-run perspective, it would be unwise for the government to try to save, using large subsidies or tariffs, production sectors that can no longer compete with newer technologies or foreign production. In this context, redistribution should take place through a general income tax that does not create production inefficiencies.

The corporate income tax in the United States provides a good example of this short-run versus long-run contradiction. The corporate income tax leads to production inefficiencies because different sectors are treated differently. It is believed that the corporate income tax treats differently sectors because some sectors successfully lobby to obtain favored tax treatment (Boskin, 1996). In the short-run, a government might find it socially beneficial to provide tax breaks in some sectors in order to affect wages and enhance

[19] Diamond (1982) develops a simple model where industries decline and workers face moving costs of switching to another industry. In that situation, it might be optimal for the government to provide subsidies to moving costs or to declining industries. The present analysis focuses on the long-run and thus ignores the moving cost issue.

redistribution in a way the income tax cannot. However, in the long-run, these inefficiencies cannot be optimal and tax preferences are cleared from time to time through a general corporate income tax reform (as happened for example in the United States with the Tax Reform Act of 1986).

Acknowledgements

I thank Robin Boadway, Peter Diamond, Thomas Piketty, David Spector, and two anonymous referees for helpful discussions and comments.

Appendix A

Proof of Lemma 1

The regularity of $X(q)$ follows from Lebesgue's Dominated Convergence Theorem (see e.g. Rudin, 1966). Let $q \gg 0$ and q^j be a sequence converging to q. By the regularity assumption on utility functions, $x^n(.)$ is continuous at q when the individual n is not indifferent between two or more occupation choices when facing price q. By assumption, this is true except for a set of measure zero. Therefore, $x^n(q^j) \rightarrow x^n(q)$ when $j \rightarrow \infty$ almost surely in n. For $q \gg 0$, it is clear that the demand functions $x^n(q)$ are bounded. Thus, Lebesgue's theorem of dominated convergence implies immediately that $X(q^j) \rightarrow X(q)$, implying that $X(q)$ is continuous. The proof of the continuity of $V(q)$ is even simpler because the individual $V^n(q)$ functions are continuous. The proof of the differentiability of $X(q)$ (and $V(q)$) proceeds in the same way using Lebesgue's theorem of dominated convergence. □

Proof of the Atkinson–Stiglitz result in the discrete model

Weak separability implies that, for a given set of consumption prices $\bar{q}$ and after tax disposable labor income m (equal to $m_i = w_i - T_i$ in occupation i), the consumption choice vector is independent of type n and thus can be written as $c(\bar{q}, m)$. Let us denote by $V(\bar{q}, m) = \max_c v(c)$ subject to $\bar{q} \cdot c \leq m$ the indirect utility. Individual n then chooses occupation i to maximize $U^n(V(\bar{q}, m_i), i)$.

Starting from no commodity taxation and optimal income taxation, let us consider, as in Christiansen (1984), the introduction of a small tax $\mathrm{d}t_1$ on (say) good 1. The proof consists in showing that the effects on tax revenue and welfare of this change can be reproduced by a small income tax change such that $\mathrm{d}T_i = c_1(\bar{q}, m_i)\mathrm{d}t_1$ for each $i = 0, \ldots, I$.

First note that the change $\mathrm{d}T_i$ is well defined because the function $c_1(\bar{q}, m_i)$ is the same for all individuals. That is why the weak separability (and uniform sub-utility $v(c)$) assumption is key to the result. Second, from Roy's identity, we have $V_{\bar{q}_1} = -V_m c_1$, thus both changes have the same effect on individual utility and hence on welfare. Third, because both changes have the same effect on the sub-utility $V(\bar{q}, m)$, any individual who switches occupations because of one of the tax changes also switches occupations because of the other one (and vice-versa). Therefore, the behavioral responses to the two tax changes are identical. Thus the effect on tax revenue due to behavioral responses is the same in both cases. Last, the mechanical change in tax revenue is the same in both cases and equal to $\mathrm{d}t_1 \sum f_i c_1(\bar{q}, m_i)$. Therefore, the small commodity tax is fully equivalent to the small income tax^{i}change. As the income tax is optimal, the income tax change, and hence the commodity tax change $\mathrm{d}t_1$ do not improve welfare, implying that no commodity taxation is optimal.

References

Atkinson, A.B., Stiglitz, J.E., 1976. The design of tax structure: direct versus indirect taxation. Journal of Public Economics 6, 55–75.

Becker, G.S., 1964. Human Capital. National Bureau of Economic Research, New York.

Blundell, R., MaCurdy, T., 1999. Labor supply: a review of alternative approaches. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A. Elsevier, Amsterdam.

Boskin, M., 1996. Frontiers of Tax Reform. Hoover Institution Press, Stanford.

Christiansen, V.A., 1984. Which commodity taxes should supplement the income tax? Journal of Public Economics 24, 195–220.

Diamond, P.A., 1982. Protection, trade adjustment assistance, and income distribution. In: Bhagwati, J.N. (Ed.), Import Competition and Response. The University of Chicago Press, Chicago.

Diamond, P.A., Mirrlees, J.A., 1971. Optimal taxation and public production. American Economic Review 61, 8–27 also 261–278.

Dixit, A., Norman, V., 1980. Theory of International Trade. Cambridge University Press, Cambridge.

Dixit, A., Norman, V., 1986. Gains from trade without lump-sum compensation. Journal of International Economics 21, 111–122.

Feldstein, M., 1995. The effect of marginal tax rates on taxable income: a panel study of the 1986 tax reform act. Journal of Political Economy 103, 551–572.

Freeman, R., 1986. Demand for education. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, Vol. 1. Elsevier, Amsterdam.

Gruber, J., Saez, E., 2002. The elasticity of taxable income: evidence and implications. Journal of Public Economics 84, 1–32.

Guesnerie, R., 1998. Peut-on toujours redistribuer les gains à la spécialisation et à l'echange? Un retour en Pointillé sur Ricardo et Heckscher-Ohlin. Revue Economique 49, 555–579.

Meyer, B., Rosenbaum, D., 2001. Welfare, the earned income tax credit, and the labor supply of single mothers. Quarterly Journal of Economics 116, 1063–1114.

Mirrlees, J.A., 1971. An exploration in the theory of optimal income taxation. Review of Economic studies 38, 175–208.

Naito, H., 1996. Tariffs and production subsidies as devices to relax the incentive problem of a progressive income tax system. Research Seminar in International Economics Discussion Paper No. 391, University of Michigan.

Naito, H., 1999. Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency. Journal of Public Economics 71, 165–188.

Pencavel, J., 1986. Labor supply of men. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, Vol. 1. Elsevier, Amsterdam.

Piketty, T., 1997. La redistribution fiscale face au chômage. Revue Française d'Economie 12, 157–201.

Rudin, W., 1966. Real and Complex Analysis. McGraw-Hill, New York.

Saez, E., 2002a. Optimal income transfer programs: intensive versus extensive labor supply responses. Quarterly Journal of Economics 117, 1039–1073.

Saez, E., 2002b. The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. Journal of Public Economics 83, 217–230.

Stiglitz, J.E., 1982. Self-selection and Pareto efficient taxation. Journal of Public Economics 17, 213–240.