May 2010

From Basic Research to Policy Recommendations: The Case for a Progressive Tax

Peter Diamond, MIT

Emmanuel Saez, UC Berkeley

Abstract:

This paper presents the case for tax progressivity based on recent results in optimal tax theory. We consider the questions of the optimal progressivity of earnings taxation and whether capital income should be taxed. For each of these questions, we critically discuss the results from academic research and whether and how such results can be used for policy recommendations. We argue that a result from basic research is relevant for policy only if (a) it is based on economic mechanisms that are empirically relevant and first order to the problem, (b) it is reasonably robust to changes in the modeling assumptions, (c) the policy prescription is implementable (i.e., is socially acceptable and is not too complex). We obtain three policy recommendations from basic research that satisfy these criteria reasonably well. First, high earners should be subject to high and rising marginal tax rates on earnings. Second, low income families should be encouraged to work with earnings subsidies, which should then be phased-out with high implicit marginal tax rates. Third, capital income should be taxed. We explain why the famous zero marginal tax rate result for the top earner in the Mirrlees model and the zero capital income tax rate results of Chamley-Judd and Atkinson-Stiglitz are not policy relevant in our view. We discuss the recent pure mechanism design approach for obtaining tax policy recommendations in dynamic models.

Peter Diamond is Professor of Economics at MIT, Cambridge Massachusetts. Emmanuel Saez is Professor of Economics, University of California, Berkeley, California. Their e-mail addresses are pdiamond@mit.edu and saez@econ.berkeley.edu, respectively. We are grateful to Henry Aaron, Alan Auerbach, James Poterba, and editor David Autor for helpful comments and discussions.

The fair distribution of the tax burden has long been a central issue in policy making. A large academic literature has developed models of optimal tax theory to cast light on the problem of optimal tax progressivity. In this paper, we explore the path from results in basic research in optimal tax theory to formulating policy recommendations.

Models in optimal tax theory typically posit that the tax system should maximize a social welfare function subject to a government budget constraint and taking into account that individuals respond to taxes and transfers. Social welfare is larger when resources are more equally distributed, but redistributive taxes and transfers can negatively affect incentives to work, save, and earn income in the first place. This creates the classical trade-off between equity and efficiency which is at the core of the optimal income tax problem. In general, optimal tax analyses maximize social welfare that is a function of individual utilities--the sum of utilities in the utilitarian case. The social marginal welfare weight for a given person measures the value of an additional dollar of consumption expressed in terms of public funds. Such welfare weights depend on the level of redistribution and are decreasing with income whenever society values more equality of income. Therefore, optimal income tax theory is first a normative theory that shows how a social welfare objective combined with constraints arising from limits on resources and behavioral responses to taxation translate into specific tax policy recommendations.[1] Second and related, optimal income tax theory can be used to evaluate current policies and suggest avenues for reform.

Moving from mathematical results, either theorems or calculated examples, to policy recommendations is a subtle process, at least when done well. That is, it is the nature of a model to be a limited picture of reality. This has two implications. First, a model may be good for one question and bad for another, depending on the robustness of the answers to the inaccuracies of the model, which will naturally vary with the question. Second, tractability concerns imply that simultaneous consideration of multiple models is appropriate since different aspects of reality can be usefully highlighted in different

[1] That is, the goal of the theory is to analyze how to achieve certain economic ends, not a description of what an actual government might do. Understanding what would be good policy, if implemented, is central to making policy recommendations.

models. Hence our reliance on trying to draw inferences simultaneously from multiple models.

In our view, a theoretical result can be fruitfully translated into a policy recommendation only if three conditions are met. First, the result needs to be based on an economic mechanism that is empirically relevant and first order to the problem at hand. This implies that it is critical to understand the economic mechanism behind a given result and then use the relevant empirical literature to assess the magnitude of the effect (or develop an empirical agenda if no such estimates exist). Second, the result needs to be reasonably robust to changes in the modeling assumptions. In particular, people have very heterogeneous tastes and there are many departures from the rational model especially in the realm of inter-temporal choice. Therefore, we should view with suspicion results that depend critically on very strong homogeneity or rationality assumptions. Deriving theoretical optimal tax formulas as a function of a few empirically estimable "sufficient statistics" is a natural way to approach those first two conditions. Third, the tax policy prescription needs to be implementable, which means that the tax policy needs to be socially acceptable and not too complex relative to the modeling of tax administration and individual responses to tax law. Some policy prescriptions such as taxing height (Mankiw and Weinzierl, 2010) are obviously not socially acceptable because they violate horizontal equity concerns that are not explicitly modeled in basic models. Such constraints effectively limit the set of tools a government can legitimately use for taxes and redistribution. The complexity constraint can be an issue when optimal taxes depend in a complex way on the full history of earnings and consumption as in the recent pathbreaking literature on optimal dynamic taxation.

We obtain three policy recommendations from basic research that we believe can satisfy these three criteria reasonably well. First, high earners should be subject to high and rising marginal tax rates on earnings. In particular, we discuss in detail why the famous zero marginal tax rate at the top of the earnings distribution is not policy relevant and why the optimal tax rate formulas derived in models that use continuum measures of populations naturally apply to models with finite numbers of people as well. Second, the earnings of low income families should be subsidized and those subsidies should then be phased-out with high implicit marginal tax rates. This result arises because labor supply responses of low earners are concentrated along the participation margin (the extensive as opposed to the intensive margin). Those two results combined imply that the optimal profile of transfers and taxes is highly nonlinear and cannot be well approximated by a flat tax along with lumpsum demogrants. Third, we argue that capital income should be taxed. We think that the Atkinson-Stiglitz and the Chamley-Judd results implying no capital income taxes are not robust enough to be policy relevant. In the end, a persuasive argument for taxing capital income (instead of taxing solely labor income) is the very simple fact that it is difficult in practice to distinguish capital and labor incomes.

The remainder of the paper is organized as follows. First, we consider the taxation of high earners, second, the taxation of low earner, and third, the taxation of capital income. Then, we discuss methodology principles and policy recommendations. We present in an appendix a discussion of the discrepancies between our lessons from optimal tax theory and those of Mankiw, Weinzierl, and Yagan (2009) recently published in this journal.

Taxation of High Earners

The US top percentile income share has increased dramatically from 9% in 1970 to 23.5% in 2007, the highest level on record since 1928 and much higher than in European countries, or Japan today (Piketty and Saez, 2003, Atkinson, Piketty, and Saez, 2010). Although the average Federal individual income tax rate of top percentile tax filers was only 22.4%, the top percentile still ended up paying 40.4% of total federal individual income taxes in 2007 (IRS, 2009). Therefore, the taxation of high earners is a central aspect of the tax policy debate not only for equity or symbolic reasons but also for revenue raising considerations. For example, increasing the average tax rate on the top 1% from the current 22.4% (as of 2007) to 29.4% would be sufficient to raise revenue by 1 percentage points of GDP.[2] This amount may well be larger than any increase in revenue that will be contemplated by 2010 National Commission on Fiscal Responsibility and Reform. Indeed, increasing the tax rate to 43.5%, which would be sufficient to raise

[2] In 2007, top 1% income earners paid $450bn in Federal individual taxes (IRS, 2009), or 3.2% of the $14,078bn GDP for 2007. Hence, increasing the average tax rate on the top 1% from 22.4% to 29.4% would raise $141bn or 1% of GDP in 2007.

revenue by 3 percentage points of GDP, would still leave the top 1% after-tax income share more than twice as high as in 1970.[3]

However, for a given measure of taxable income, higher taxes on high earners can only be achieved by increasing marginal tax rates which can discourage economic activity through behavioral responses, and hence potentially reduce tax collections, creating the standard equity-efficiency trade-off discussed in introduction. Changing the measure of taxable income would be a complementary approach to increasing tax rates.

The optimal top marginal tax rate

As shown in Saez (2001), the optimal top marginal tax rate is actually straightforward to derive. Denote the tax rate in the top bracket by τ. As depicted on Figure 1, consider a tax reform which increases τ by $d\tau$ above the income level z*. To evaluate this change we need to consider the impacts on revenue and on utilities ignoring the behavioral response as well as the impact on revenue due to the behavioral response. Ignoring behavioral responses at first, this reform mechanically raises additional revenue $dM=d\tau\ N^* [z_m(z^*) - z^*]$, where N* is the number of people with incomes above z* and $z_m(z^*)$ is their average income. As we shall see, the top tail of the income distribution is closely approximated by a Pareto distribution characterized by a power law density of the form C/z^{1+a} where a>1 is the Pareto parameter. Such distributions have the key property that the ratio $z_m(z^*)/z^*$ is the same for all z* in the top tail and equal to a/(a-1). Empirically, a is approximately 1.5. [4]

[3] The top 1% average Federal individual tax rate was 25.7% in 1970 (Piketty and Saez, 2007) and 22.4% in 2007 (IRS, 2009). The average Federal individual tax rates was 12.5% in 1970 and 12.7% in 2007. The top 1% pre-tax income share was 9% in 1970 and 23.5% in 2007. Hence, the top 1% after-tax income share was 7.6%=9%*(1-.257)/(1-.125)) in 1970 and 20.9%=23.5%*(1-.224)/(1-.127) in 2007, and would have been 16.1%=23.5%*(1-.435)/(1-.177) with a tax rate of 43.5% on the top 1% (which would increase the average tax rate to 17.7%).

[4] In the United States in 2009, the top federal income tax rate is 35% and applies to taxable incomes above $373,000, i.e., approximately the top 1% tax filers. Therefore, we have approximately z*=$400,000, z_m=$1,200,000 so that $z_m/z^*=3$ and hence a=1.5, N*=1.5m, and τ is 42.5% for ordinary labor income when combining the top federal individual tax rate of 35%, uncapped Medicare taxes of 2.9%, and an average combined state top income tax rate of 5.86% and average sales tax rate of 2.32%.

The 5.86% average top rate for states is estimated by taking the average top state income tax rate. The average across states is computed using state weights equal to the fraction of filers with AGI above $200,000 that reside in the state as of 2007 (IRS, 2009). The 2.32% average sales tax rate is estimated as 40% of the average nominal sales tax rate across states (as the average sales tax base is about 40% of total personal consumption) As the 1.45% employer Medicare tax is deductible for both Federal and State incomes taxes, and state income taxes are deductible for Federal income taxes, we have ((1-.35)*(1-.0586)-.0145)/(1.0145*1.0232)=.575 and hence τ=42.5%.

The reform obviously reduces the utility of high income tax filers. If we denote by g^* the social marginal weight on top income earners (relative to government revenue), and using a standard envelope condition, the direct welfare cost is simply $g^* dM$.[5] Because the government values redistribution, for high enough values of z^*, the social marginal value of consumption for top-bracket tax filers is small relative to that of the average person in the economy so that g^* is small and can be ignored. For example, a utilitarian social welfare criterion with marginal utility of consumption declining to zero, the most commonly used specification in optimal tax models, has this implication.[6]

Behavioral responses can be simply captured by the elasticity of reported income with respect to the net-of-tax rate $1-\tau$. Let us denote this elasticity by $e=(1-\tau)/z_m \, dz_m/d(1-\tau)$.[7] By the definition of e, the reduction in tax revenue due to behavioral responses is equal to $dB=N^* \tau \, dz_m=-N^* d\tau \, e \, z_m \, \tau/(1-\tau)$. Therefore, in net, the tax reform raises $dM + dB = N^* d\tau \, (z_m - z^*) \, [1 - e \, a \, \tau/(1-\tau)]$. At the optimum, with little social weight on the consumption of the highest earners, $dM+dB$ must be approximately zero which directly implies that the optimal top tax rate τ^* is given by the formula

$$\tau^*=1/(1+a\,e). \qquad (1)$$

The optimal top tax rate τ^* is the tax rate that maximizes tax revenue from top bracket taxpayers. Any top tax rate above τ^* would be (second-best) Pareto inefficient as reducing taxes at the top would both increase tax revenue and the welfare of top earners. τ^* is unsurprisingly decreasing with the elasticity e and the Pareto parameter a, which measures the thinness of the top of the income distribution. The optimal tax rate can be calculated for varying social weights on the consumption of the very high earners. As noted below, with plausible weights that are small relative to the weight on an average earner, the optimal tax does not change much.[8]

---

[5] Formally, g^* is the weighted average of social marginal weights on top income earners (relative to government revenue), with weights depending on income in the top bracket.

[6] For example, if utility is log in consumption, then social marginal welfare weights are inversely proportional to consumption, In 2007 in the United States, the social marginal utility at the $1.364 million average income of the top 1% (Piketty and Saez, 2003) is only 3.9% of the social marginal utility of the median family, with income $52,700 (US Census Bureau, 2009).

[7] Formally, this elasticity is an income weighted average of the individual elasticities across the N^* top bracket tax filers. It is also a mix of income and substitution effects as the reform creates both income and substitution effects in the top bracket (Saez, 2001 provides an exact decomposition).

[8] If a positive weight $g^*>0$ is set on top earners, then the optimal rate is $\tau=(1-g^*)/(1-g^*+a\,e)<\tau^*$.

The solid line in Figure 2 depicts the empirical ratio $a=z_m/(z_m-z^*)$ with z* ranging from $0 to $1,000,000 in annual income using US tax return data for 2005-the latest year for which micro tax data are available.[9] The central finding is that a is extremely stable for z* above $300,000 (and around 1.5). The excellent Pareto fit of the top tail of the distribution has been well known for over a century since the pioneering work of Pareto (1896) and verified in many countries and many periods.[10]

Therefore, if we assume that the elasticity e is roughly constant across earners at the top of the distribution, the formula $\tau=1/(1+a\ e)$ from equation (1) shows that the optimal top tax rate is independent of z* within the top tail and is also the asymptotic optimal marginal tax rate coming out of the standard nonlinear optimal tax model of Mirrlees (1971). Intuitively, the Pareto distribution is fractal in the sense that z_m/z^* is constant with z* so that the optimal top tax rate problem is independent of z* making formula (1) quite general and useful. An increase in the marginal tax rate only at a single income level in the upper tail increases the deadweight burden (decreases revenue because of reduced earnings) at that income level but raises revenue from all those with higher earnings without altering their marginal tax rates. The optimal tax rate balances these two effects – the increased deadweight burden at the income level and the increased revenue from all higher levels. With the fractal nature of the Pareto distribution and the assumed constancy of the earnings elasticity, the optimal rate is approximately the same over the range of high incomes where the distribution is Pareto and the marginal social weight on consumption is small.

The elasticity e of top incomes

The key remaining empirical ingredient to implement formula (1) is the elasticity e of top incomes with respect to the net-of-tax rates. With the Pareto parameter a=1.5, if e=.25, then $\tau^*=1/(1+1.5*.25)=73\%$, substantially higher than the current 42.5% top US marginal tax rate (combining all taxes). If in contrast e=1, then $\tau^*=1/(1+1.5)=40\%$, slightly lower than the current top tax rate.[11]

[9] We use Adjusted Gross Income as our income definition. Tax return data oversample top incomes and are ideally suited for this exercise--also presented by Saez (2001) for wage income reported on tax returns for years 1992 and 1993.

[10] See, for example, the recent top income share studies summarized in Atkinson, Piketty, Saez (2010).

[11] Using g* of .04, these two optimal tax rates decrease by about 1 percentage point.

The literature estimating this elasticity has focused primarily on the response of reported income, either Adjusted Gross Income or Taxable Income, to net-of-tax rates.[12] The behavioral elasticity is due to real economic responses such as labor supply, business creation, or savings decisions, but also tax avoidance or evasion responses. A number of studies have shown large and quick responses of reported incomes along the tax avoidance margin at the top of the distribution but no compelling study to date has shown substantial responses along the real economic responses margin among top earners.[13] For example, in the United States, realized capital gains surged in 1986 in anticipation of the increase in the capital gains tax rate after the Tax Reform Act of 1986 (Auerbach, 1988). Similarly stock-options exercises surged in 1992 before the 1993 top rate increase took place (Goolsbee, 2000). The Tax Reform Act of 1986 also lead to a shift from corporate to individual income as it became more advantageous to be organized as a business taxed solely at the individual level rather than as a corporation taxed first at the corporate level (Slemrod, 1996, Gordon and Slemrod, 2000). Gruber and Saez (2002) is often cited for its substantial taxable income elasticity estimate (e=0.57) at the top of the distribution. However, they also found a small elasticity (e=0.17) for income before any deductions even at the top of the distribution (Table 9, p. 24).

When a tax system offers tax avoidance or evasion opportunities, the tax base is quite sensitive to tax rates, so that the elasticity e is large, and the optimal top tax rate is correspondingly low according to formula (1). Two important qualifications must be made. First, many of the tax avoidance channels such as re-timing or income shifting produce changes in tax revenue in other periods or other tax bases—called "tax externalities" and hence increase the optimal tax rate relative to formula (1).[14] Other behavioral responses, such as increased charitable giving following a tax rate increase, can also create classic externalities that also affect the optimal formula (Saez 2004b). Second and most important, the tax avoidance or evasion component of the elasticity e is not an immutable parameter and can be reduced through base broadening and tax

[12] There is a large literature which uses tax reforms to estimate e. See Saez, Slemrod, and Giertz, 2010 for a recent survey, and Slemrod, 2000 for studies focusing on the rich.

[13] Moffitt and Wilhelm (2000) provide a good illustration of this point by showing that top incomes increased significantly following the top tax rate reductions of the Tax Reform Act of 1986 but that the hours of work of top earners did not change.

[14] Saez (2004) and Saez, Slemrod, Giertz (2009) provide formulas showing how the optimal top tax rate should be modified in such cases.

enforcement (Slemrod and Kopczuk, 2002, Kopczuk 2005), while the real economic response component cannot. This makes the distinction between real responses and tax avoidance responses critical for tax policy. As an illustration using the elasticity estimates of Gruber and Saez (2002) for high income earners mentioned above, the optimal top tax rate using the current taxable income base (and ignoring tax externalities) would be $\tau^*=1/(1+1.5*0.57)=54\%$ while the optimal top tax rate using a broader income base with no deductions would be $\tau^*=1/(1+1.5*0.17)=80\%$. Taking as fixed state and payroll tax rates, such rates would correspond to top federal income tax rates equal to 48% and 76% respectively. Considerable uncertainty remains in the estimation of the long-run behavioral responses to top tax rates (Saez, Slemrod, Giertz, 2010), but the case for higher rates at the top appears robust.

Link with the zero top rate result

Formally, z_m/z^* reaches 1 when z^* reaches the level of income of the single highest income earner, in which case $a=z_m/(z_m-z^*)$ is infinite and indeed $\tau^*=0$, which proves the famous zero top rate result first demonstrated by Sadka (1976) and Seade (1977) in the context of the Mirrlees model. However, this also shows that this famous result applies only to the very top income earner; its lack of wider applicability can be verified empirically using tax data.[15] While the income distribution has a finite top, the level of the top highest earner is not known to the government ex-ante. Therefore, a reasonable assumption that fits with the empirical findings from Figure 2 is that top earners are drawn randomly from an underlying Pareto distribution so that the realized income distribution is finite but the level of the top is unknown ex-ante. In that context, as we show in the appendix, starting from the top tax rate given by formula (1), no tax rate change above any high income level can increase expected tax revenue so that formula (1), $\tau^*=1/(1+a\ e)$, remains the natural optimum tax rate. This implies that the zero top rate result and its corollary that marginal tax rates should decline at the top have

[15] If, for example, the second highest income is only one-half of the highest earner then $z_m/z^*=2$ (and hence $a=2$) when z^* is just above the second highest earner so that convergence of z_m/z^* to one really happens only between the top and second highest earner. The IRS publishes statistics on the top 400 taxpayers (IRS, 2009b). In 2007, the latest year available, the threshold to be a top 400 taxpayer was \$138.8m and the average income was \$344.8m so that $a=1.67$ at $z^*=\$138.8m$.

no policy relevance, a view that we believe is widely shared among public finance economists.

Should Marginal Tax Rates Rise with Income?

Assuming away income effects, the optimal marginal tax rate formula at any income level takes a form that can be expressed directly as a function of the income distribution as follows (Diamond (1998), see Figure 3 for a derivation):

$$T'(z)=[1-G(z)]/[1-G(z)+\alpha(z)\, e(z)] \qquad (2)$$

where e(z) is the elasticity of incomes with respect to the net-of-tax rate at income level z, G(z) is the average social marginal welfare weight across individuals with income above z, and $\alpha(z)=(zh(z))/(1-H(z))$ with h(z) the density of taxpayers at income level z and H(z) the fraction of individuals with income below z.[16] The expression $\alpha(z)$ reflects the ratio of those affected by the marginal tax rate at z (times their income) relative to the numbers of people at higher income levels. It is constant and equal to the Pareto parameter for Pareto distributions. It is depicted in dotted line on Figure 2 for the empirical 2005 US income distribution. It is inversely U-shaped, reaching a maximum of 2.17 at z=$135,000, then decreasing and staying indeed approximately constant around 1.5 above z=$400,000. Naturally, as social welfare weights are lower for higher incomes, G(z) decreases with z. Therefore, assuming a constant elasticity e across income groups, formula (2) implies that the optimal marginal tax rates should increase with income in the upper part of the distribution. This result was theoretically established by Diamond (1998) and confirmed by all subsequent simulations which use a Pareto distribution at the top as in Saez (2001) or Mankiw et al. (2009). Quantitatively, this increase is substantial. For example, assuming again an elasticity e=.25, and that G(z)=0.5 at z=$100K corresponding to the top decile threshold where α=2.05, we would have T'=49% at this income, well below the value of 73% for the top percentile as calculated above.

As discussed above, in the current tax system with many tax avoidance opportunities at the higher end, the elasticity e is likely to be higher for top earners than for middle incomes, possibly leading to decreasing marginal tax rates at the top (Gruber and Saez, 2002). However, the natural policy response should be to close tax avoidance

[16] Technically, Saez (2001) shows that h(z) is the density of incomes when the nonlinear tax system is linearized at z. Saez (2001) also shows that a similar but more complex formula can be obtained with income effects that is quantitatively close.

opportunities, in which case the assumption of constant elasticities might be a reasonable benchmark.

Taxation of Low Earners

Transfers are naturally integrated with taxes in an optimal tax problem. In practice, means-tested transfers are often administered through specific agencies, but need to be analyzed along with taxes. Such transfers often take the form of a maximum benefit for those with no income, which is phased-out at high rates as earnings increase. For example, in the United States, TANF (Temporary Aid to Needy Families) and SNAP (Supplemental Nutrition Assistance Program, formerly known as Food Stamps) programs operate in this way. A growing fraction of means-tested transfers is now administered through refundable tax credits such as the EITC (Earned Income Tax Credit) or the Child Tax Credit. Such programs are typically first phased-in and then phased-out with earnings so that benefits are concentrated on low income working families instead of those with no earnings. Empirically, many studies have found compelling evidence of substantial labor supply responses to transfers along the extensive margin (whether of not to work). For example the EITC expansions have encouraged labor force participation of US single mothers (Meyer, 2010). However, there is much less compelling evidence of behavioral responses along the intensive margin (hours of work on the job) for lower income earners. For example, EITC recipient wage earners do not pile up around the plateau of the EITC where benefits are maximum (Saez, 2010). As we shall see, this plays a critical role in the optimal profile of transfers.

Intensive Elasticities

In the Mirrlees (1971) model, behavioral responses take place only through the intensive margin. In that context, it is optimal to phase-out transfers at the bottom at a high rate. We provide a derivation in the appendix. The intuition is the following, a high phasing out rate allows the government to target transfers to the most disadvantaged families. A high phasing-out rate reduces earnings for low income families through the intensive elasticity. However, because earnings of those in the phase-out are small to start with, this elasticity applies to a low income base. Therefore, increasing the maximum benefit (to those with no earnings) and increasing the phasing-out rate is desirable for redistribution and the behavioral responses create modest fiscal costs relative to the redistributive gains as long as the phasing-out rate is not too high.

Extensive Elasticities

Consider now a model where behavioral responses of bottom earners take place through the extensive elasticity only, i.e., whether or not to work, and that earnings when working do not respond to marginal tax rates. As discussed above, such a model is empirically the most relevant. Suppose the government starts from a transfer scheme with a positive phasing-out rate and introduces an additional small in-work benefit that increases net transfers to low income workers earning z_1 as depicted in Figure 4. Ignoring behavioral responses, such a reform is desirable if the government values redistribution to low income earners, a reasonable hypothesis. If behavioral responses are solely along the extensive margin, this reform induces some low skilled non-workers to start working to take advantage of the in-work benefit (thick curly arrow on the figure). However, because we start from a situation with a positive phasing-out rate, this behavioral response increases tax revenue as low income workers get a smaller transfer than non-workers. Hence, the reform is desirable showing that a positive phasing-out rate is not optimal. Diamond (1980) originally solved the optimal tax problem in a participation model with a continuum of earnings and Saez (2002) solved the discrete version (the derivation is outlined in Figure 4).

In practice, both extensive and intensive elasticities may be present. As depicted on Figure 4 with a light dotted curly arrow, intensive margin response would induce slightly higher earners to reduce labor supply to take advantage of the in-work benefit, reducing tax revenue. Therefore, the government has to trade-off the two effects. If, as empirical studies show, the extensive elasticity is large relative to the intensive elasticity, initially low (or even negative) phasing-out rates combined with higher phasing-out rates further up the distribution would be the optimal profile (Saez, 2002 provides simulations in such a mixed model).

Taxation of Capital Income

With the wide acceptance of the standard model for static labor supply decisions, the simplicity of a one-period model and the extensive empirical literature on labor supply elasticities, it is possible to provide useful quantitative analysis of optimal marginal tax rates. In contrast, the literature on saving behavior sees a wide variety of basic behaviors, more widely varying elasticity estimates, and a complexity that comes from the importance of the future for capital income taxation. Thus we limit our discussion to a single qualitative recommendation – capital income should be subject to significant taxation. This conclusion is not as satisfactory as the results derived for the taxation of earnings but is important in light of repeated calls for not taxing capital income. Academic arguments against capital income taxation typically draw on one or more of three theoretical analyses: (1) the aggregate efficiency theorem in Diamond and Mirrlees (1971), (2) the theorem that the optimum has no asymptotic taxation of capital income in Chamley (1986) and Judd (1985), and (3) the theorem that the optimum has no taxation of capital income in Atkinson and Stiglitz (1976). We address each of these in turn, before turning to analyses that conclude that taxing capital can improve social welfare. Key elements in our analysis are that (1) the aggregate efficiency theorem does not imply that capital income should not be taxed – that claim is a misreading of the theorem; (2) the Chamley-Judd result requires the implausible assumption that bequest behavior drives capital accumulation in the economy in a manner consistent with an infinite horizon optimization; and (3) the Atkinson-Stiglitz theorem requires uniformity in savings functions despite the documented diversity in savings behavior which supports the case for capital income taxation.[17] Three direct arguments for taxing capital income come from (1) the difficulty in distinguishing between capital and labor incomes, (2) the presence of borrowing constraints, and (3) uncertainty in future earnings opportunities. As we shall see, the optimal capital income tax literature develops many different models which lead to diverse results. Therefore, robust policy recommendations are harder to draw relative to the case of labor income taxation.

Diamond and Mirrlees

The aggregate efficiency theorem states that in the presence of a complete set of taxes the optimal allocation is on the frontier of the aggregate production set. In the usual

[17] For lengthier discussion of the Chamley-Judd and Atkinson-Stiglitz arguments, see Banks and Diamond (2009).

(nonstochastic) constant returns setting with no externalities, aggregate efficiency rules out taxes or tariffs on intermediate goods that would directly affect relative prices (for example, taxes on transactions between firms except for a uniform VAT) and rules out government production decisions that are not consistent with shadow prices that match the market prices faced by firms.

Taxes on transactions between households and firms (that do not vary with the particular firm) do not interfere with production efficiency. While taxing capital income of households will generally change the level of savings, and so investment, it does not move the economy inside the production possibility frontier – it alters the mix of aggregate first- and second-period consumptions but still leaves the economy on the frontier in first-period consumption - second-period consumption space. The logic behind the result is that (1) any interference with aggregate efficiency will necessarily change the prices of transactions between households and firms, and (2) using taxes to achieve the same changes in prices is a better strategy since it avoids production inefficiency. Thus, the aggregate efficiency theorem has <u>no</u> direct implications relative to taxing the capital income of households.

Chamley and Judd

In the models analyzed in Chamley (1986) and Judd (1985), with infinitely lived agents, an asymptotically zero tax on capital income is optimal. In order to appreciate the relevance of this result for policy purposes, one needs to understand the logic of the result, and particularly its robustness to key assumptions and their degree of accuracy. As pointed out in Judd (1999), the logic for the result is straightforward. A constant capital income tax rate creates a growing tax wedge between current consumption and future consumption as the horizon grows. With interest rate r and no capital income taxes, a dollar today is worth $(1+r)^T$ after T years. If an investor is subject to an annual tax at rate τ on capital income, then the investor can convert one unit of consumption today into only $((1+(1-\tau)r)^T$ after T years. Hence, the tax wedge $1-(1+(1-\tau)r)^T/(1+r)^T$ grows with T.[18]

[18] While interest income and dividends are taxed in this compounding way, the same is not true for capital gains that are taxed on a realization basis. Nor is it true for tax-favored retirement saving (such as IRAs or 401(k)s).

For example, with r=.05 and τ=30%, the tax wedge is a modest 13.4% when T=10 but is a substantial 43.8% when T=40. If the model is extended to a dynastic model, the tax wedge becomes very large when compounding the tax across multiple generations. In order to avoid tax compounding that grows without limit in a model with stationarity, the optimal rate must go to zero. Without stationarity (as in Judd, 1999), it is the *average* rate from the near term to the asymptotic future that must be zero, reflecting both taxes on capital income and estate taxes, although no individual tax rate needs to be zero.

Therefore, the result relies critically on the assumption that individuals make consistent rational decisions about savings behavior across very long horizons as in the standard intertemporal model. When agents have long horizons, modeling their current decision-making using an infinite horizon model can be mathematically more tractable than a long finite horizon, while doing little violence to conclusions from analyses that relate to current behavior. In contrast, substituting an infinite-horizon decision maker for a sequence of finite-horizon decisionmakers can make a large difference when analyzing the asymptotic position of the economy rather than short term outcomes. In an OLG model with no dynastic linkage, the optimal capital income tax is generally not zero, even in the long-run (Diamond, 1973, Atkinson and Sandmo, 1980). Thus the strongest asymptotic zero tax result of Chamley and Judd requires that rational intertemporal decision making not only holds for entire lifetimes, but extends across dynasties as in the Barro-Becker model. Both assumptions have been heavily challenged in the empirical literature.

First, the recent behavioral economics literature has cast much doubt on the standard model of intertemporal decision making for a significant fraction of the population. In particular, there is a growing body of empirical work showing that savings decisions are heavily influenced by psychological elements (such as self-control) or minor transaction costs (default effects in employer sponsored 401(k) plans).

Second, empirical analyses of gifts and bequests, while clearly showing concerns about heirs (and so the inaccuracy of the standard OLG model with no bequests), are not supportive of the rigorous version of the dynasty model required for the Chamley-Judd result. There are many reasons why people leave bequests: unintended bequests due to lack of annuitization or love of wealth accumulation per-se, intended bequests arising out of bargaining with heirs, warm glow preferences, or altruism. The optimal tax treatment of bequests depends heavily on the mechanism behind bequests (see Cremer and Pestieau, 2004 for a recent survey). For example, unintended bequests should be taxed heavily because they do not affect donors and inheritances induce donees to work less through income effects. The dynastic model is a special form of altruism and hence likely captures only one aspect of bequest behavior. As a result, we reject the modeling approach's policy relevance.

Another straightforward conclusion coming out of the Chamley-Judd model is that it is better to tax existing wealth rather than future capital income as the former tax is lumpsum while the latter distorts intertemporal choices. While the asymptotic zero capital income tax result has drawn great attention, the initial result is largely ignored for policy purposes, although the same perspective, clearly stated in the literature, lies behind arguments for switching from income taxation to consumption taxation in OLG models as a way to transfer wealth away from older cohorts at the time of tax implementation with little in the way of distorting incentives.[19] Taxing initial wealth as much as the available tax tools allow (whether as a wealth tax or a capital income tax) strains the relevance of the assumption that the government can commit to a policy that this taxation of wealth will not be repeated. Without a genuine commitment technology, confiscatory wealth taxation would adversely affect saving behavior and have serious efficiency costs because of concerns that such taxation will return.

Atkinson and Stiglitz

In a two-period model with one period of work, the well-known Atkinson-Stiglitz theorem (1976) states that when the available tax tools include nonlinear earnings taxes, differential taxation of first- and second-period consumption is not optimal if two key conditions are satisfied: (1) all consumers have preferences that are separable between consumption and labor and (2) all consumers have the same sub-utility function of

[19] This basis for a change in taxation is very sensitive to implementation. It works for taxing consumption directly and for taxing consumption as income less savings provided initial wealth is measured, but may not work for taxing consumption as income less savings if initial wealth is not measured.

consumption. The underlying logic behind the Atkinson-Stiglitz result starts with the observation that the incentive to earn comes from the utility achievable from consumption purchases with after-tax earnings. With separable preferences and the same subutilities for everyone, differential consumption taxation can not accomplish any distinction among those with different earnings abilities beyond what is already accomplishable by the earnings tax, but would have an added efficiency cost from distorting spending choices. Thus the use of distorting taxes on consumption is a more costly way of providing the incentives for the 'optimal' earnings pattern in equilibrium.[20]

While the Atkinson-Stiglitz theorem requires an absence of a systematic pattern between earnings abilities and savings propensities., there appears to be a positive correlation between labor skill level (wage rate) and savings propensities. In the Atkinson-Stiglitz two-period certainty setting with additive preferences, this pattern of savings rates is consistent with those with higher earnings abilities discounting future consumption at a lower rate.[21] With this plausible assumption, implying that those with higher earnings abilities save more out of any given income,, then taxation of saving helps with the equity-efficiency tradeoff by being a source of indirect evidence about who has higher earnings abilities and thus contributes to more efficient redistributive taxation (Saez, 2002b).[22]

The dimensionality of worker types (relative to tax tools) matters in models of capital income taxation. This point can be brought out by contrasting the analysis of the taxation of capital income in a two-type model in Diamond (2003) with that in a four-type model in Diamond and Spinnewijn (2010). Both papers use two-period models and

---

[20] Laroque (2005) and Kaplow (2006) provide an elegant and straightforward proof of this point. They show that one can always move to a system of non-distorting consumer taxes coupled with an appropriate modification of the earned income tax and generate more government revenue while leaving every consumer with the same utility and the same labor supply.

[21] Banks and Diamond (2009) review evidence on the relationship between savings and skill levels as well as psychological evidence on discount factors. Empirical studies of savings behavior mostly find that those with higher lifetime incomes do save more, but that the full pattern of savings requires considerable complexity in the underlying model (including uncertainties about earnings and medical expenses, asset tested programs, differential availability of savings vehicles, and bequest motives) to be consistent with the different aspects of savings at different ages. Thus the higher savings rates are consistent with the preference assumption of Saez, but not, by themselves, a basis for necessarily having the discount rate pattern that Saez assumes, since these other factors are also present.

[22] Golosov, Tsyvinski, and Weinzierl (2009) propose a calibration exercise and find that the quantitative uniform capital income tax is small.

assume additive preferences, with workers varying in both skill and discount factor. The two-type result is that the higher skill type is subject to no marginal taxation on either earnings or savings, while the lower skill type is marginally taxed on earnings and also on savings if the lower skill type has a lower discount factor. To have a diverse population at each earnings level, the four-types setting uses a jobs model instead of an hours model, so that all high types have the same earnings and all low skill types have the same earnings. If a high skill type were to take a job for the low-skill types, the savings decision would match that of the low skill type with the same discount factor. Considering both the introduction of a small savings tax, or linear taxation of savings (along with earnings taxation), with the tax rate on savings allowed to vary with earnings level, optimality in the paper's central case has taxation of savings of high earners and subsidization of savings of low earners. The underlying logic comes from the incentive compatibility constraints, since high discount types are more willing to work than low discount types given the same skill and savings taxes.[23]

Distinguishing between capital and labor incomes

It is often difficult to distinguish between capital and labor incomes.[24] This is clearly the case for small businesses where profits arise both from the labor of owners and returns on assets so that, to some degree, individuals can convert labor income into capital income to take advantage of any tax differential. For example, after the 1993 Finnish tax reform to a dual income tax with a lower rate on capital income, there were significant shifts of labor income to capital income among the self-employed (Pirttilä and Selin (2007)). In the United States, Gordon and MacKie-Mason (1995) and Gordon and

[23] The models considered above have variation in the population in earnings ability, and sometimes in preferences, but not in wealth at the start of the first period. With variation in initial wealth holdings and an ability to tax initial wealth, the optimum may call for full taxation of initial wealth, particularly when higher wealth is associated with higher earnings abilities. If immediate taxation of initial wealth is ruled out, the presence of capital at the start of the first period, which can earn a return when carried to the second period, can also prevent the optimality of the non-taxation of capital income if there are no fairness issues further limiting the desirability of taxation of initial wealth. As a modeling issue, one needs to ask where such wealth came from. Presumably gifts and inheritances, along with prior earnings and savings, are major sources. But since gifts and bequests might themselves be taxed and since they might be influenced by future taxation of capital income, a better treatment of this issue would be embedded in an OLG model that incorporates the different ways that people think about bequests. See Cremer and Pestieau (2004) for a recent survey.

[24] People spending time to manage their investment portfolios are converting labor time into anticipated capital income.

Slemrod (2000) have found income shifting between the corporate tax base and the individual tax base driven by tax differentials. The existence of tax differentials also puts pressure to extend the most favorable tax treatment to a wider set of incomes. For example, in the United States, compensation of private equity and hedge fund managers in the form of a share of profits generated on behalf of clients is considered realized capital gains although it is conceptually labor income

The difficulty in telling apart labor and capital income is perhaps the strongest reason why governments would be reluctant to completely exempt capital income and tax only labor income as some of the theoretical models described above recommend. Obviously, if conversion of labor income into capital income had no cost, the only option for the government would be comprehensive individual income taxation based on the sum of capital and labor income. With small costs, the tax differentials would also need to be small. Christiansen and Tuomala (2008) examine a model with costly (but legal) conversion of labor income into capital income. Despite preferences that would result in a zero tax on capital income in the absence of the ability to shift income, they find a positive tax on capital income. Similarly, the Chamley-Judd result of zero capital income taxation does not hold in a model with an inability to distinguish between entrepreneurial labor income and capital income in the same basic model (Reis, 2007).

Borrowing constraints

The models discussed above had perfect capital markets – no borrowing constraints.[25] But borrowing constraints are relevant for tax policy, providing another reason for positive capital income taxation. Since capital income taxes fall on those who are not borrowing constrained (since they have capital) raising revenue from a capital income tax allows for a lower earned income tax, including the tax on those who are so constrained – allowing for an efficiency gain when taxes are collected. For example, Aiyagari (1995) considered borrowing constrained agents in an uncertainty setting. In this model, precautionary saving is high in anticipation of future borrowing constraints.

[25] Zeldes (1989) shows that, contrary to the predictions of the consumption-smoothing model with no liquidity constraints, consumption paths track predictable changes in income for low wealth groups.

In turn, this implies that a positive capital tax is welfare improving in the standard infinitely-lived agent setup.

Uncertain future earnings

In the Atkinson-Stiglitz model, a worker knows the returns to working in every period before doing any consumption. Uncertainty about future earnings opportunities is large and pervasive (Banks and Diamond, 2009). When some consumption decisions are taken before earnings uncertainties are resolved, the Atkinson-Stiglitz result does not hold and, in a two-period model, second-period consumption should be taxed at the margin relative to first-period consumption.

The underlying logic of this result is that welfare is enhanced by providing insurance about future earnings opportunities through the tax system. The extent of insurance is limited by moral hazard concerns. When leisure is a normal good, those who save more, ceteris paribus, will tend to work less later on. Thus, discouraging savings enhances the ability to provide insurance against poor labor market possibilities. The advantage of discouraging savings is present in models with longer time horizons as well.

The literature making this point has two strands. The optimal tax strand considers optimal linear taxation of capital income along with optimal nonlinear earnings taxes. Provided a plan with less future work is accompanied by more savings, introducing such taxation raises welfare (Diamond and Mirrlees, 1982, 2000). The Atkinson-Stiglitz preference assumptions do not affect this result.

A second strand uses the mechanism design approach of social welfare optimization with the government controlling individual consumption and labor, subject to incentive compatibility constraints and aggregate resources. With additive preferences, a robust finding of this literature is the Inverse Euler Equation – that the reciprocal (inverse) of the marginal utility of consumption is equal to the expectation of the reciprocal (inverse) of the future marginal utility of consumption - $1/u'(c_1) = \mathrm{E}\{1/u'(c_2)\}$. In a certainty model, the Inverse Euler Equation and the familiar Euler Equation are the same. With uncertainty they are not and, from Jensen's inequality, the marginal utility of present consumption is less than the expected marginal utility of future consumption. Thus in the absence of restrictions, an individual would want to save more than with the socially optimal plan. To implement such an allocation one needs to have a "wedge" reflecting implicit marginal taxation of future consumption relative to earlier consumption, and so an implicit marginal tax on savings or capital income (Golosov, Kocherlakota, and Tsyvinski 2003). The Inverse Euler Condition comes from optimally balancing the incentives for today's work coming from additional compensation today and from anticipated changes in future resources as a consequence of today's additional earnings, since the inverse of marginal utility (relative to the Lagrangian on the resource constraint) is the resource cost of increasing utility. Making it less attractive for someone with higher future earnings skills to imitate someone with lower earnings skills improves the equity-efficiency tradeoff (weakens the impact of the incentive compatibility constraint).

The literature has examined implementation of such an optimum in simple settings using familiar tax tools. For example, Diamond and Mirrlees (1982, 1986) apply this to the adjustment of retirement benefits as a function of the age of retirement in a setting where the alternatives are a particular job or no work at all and there is uncertainty about the ability to hold the job. Implicitly taxing both work and savings allows for more redistribution to those who should retire early by discouraging savings in order to take advantage of an early retirement pension that then grows at less than an actuarially fair rate with continued work.[26] More recently, Golosov and Tsyvinski (2006) and Golosov, Troshkin, and Tsyvinski (2009) have made important progress in understanding what tax systems can decentralize the optimum. Golosov and Tsyvinski (2006) study optimal disability insurance and recognize a role for an asset test, as is widespread in programs for the poor.

However, in a many period model with a rich stochastic dynamic pattern of wage rates, full implementation of a mechanism design optimum calls for a complex, sophisticated tax structure. An example of such a structure is derived in Kocherlakota (2005). It calls for the taxes in any period to depend on the full history of earnings up to that period and has linear capital income tax rates that have a regressive relationship to

[26] Of course, there is also a concern for those who save too little for their own good – a concern that lies behind mandatory public pension systems.

contemporaneous earnings, and which collect no revenue in aggregate. This implementation discourages savings by making the return to savings stochastic even though the rate of return on investment is determinate. And the regressivity of the tax rate is designed to discourage savings by providing a higher return when marginal utility is lower. Alternatively, implementation can be done using positive non-linear capital taxes (also dependent on past earnings) (Werning, 2009). When there are alternative ways to implement a mechanism design optimum, without further research, it is not clear which approach sheds light on how to levy taxes in more realistic settings with limited tax tools.

The bottom line is that uncertain future earnings opportunities argue against zero taxation of capital income, as do savings preference heterogeneity, limited distinctions between capital and labor incomes, and borrowing constraints.

Methodology and Recommendations

If we were in government, helping to set tax policy, we would need to reach concrete conclusions on tax bases and tax rates. In our role as part of the general discussion of taxation that may influence the tax-setting process, we look to inform thinking about taxes, without necessarily getting to a concrete recommendation. In deciding what issues to promulgate and what supportive arguments to put forth, we draw on parts of the optimal tax literature. We also recognize a role of the theoretical analyses in rebutting arguments that do not seem to be a good basis for making tax policy. This approach, drawing on multiple research sources for partial insights, seems appropriate given the complexity of issues that are relevant for good tax policy, much less the even richer set of issues that would also recognize the role of arguments in a complex political process.

As a good model for addressing the many issues that matter for good tax policy, we think of the Meade Report (Meade, 1978). Chapter 2 of the Report, 'The Characteristics of a Good Tax Structure', is divided into six sections: Incentives and economic efficiency, Distributional effects, International aspects, Simplicity and costs of administration and compliance, Flexibility and stability, and Transitional problems. To consider direct taxation in the UK, the Meade Committee examined each of these issues separately and then combined the insights into a policy recommendation. It seems to us, as it seemed to Alfred Marshall, that economic analysis needs to proceed in a similar fashion.[27]

The models available for analysis, like much of the underlying theory, remain limited and still too far from reality to proceed in any other fashion than that followed by the Meade committee. So, in this paper, we have identified basic research findings that we find relevant in thinking about practical tax setting, and basic research findings that others may find relevant, but we do not. In the latter category, we have placed high implicit marginal tax rates on low earners in models with only an intensive margin (since the extensive margin is so important for low earners), the zero optimal tax rate at a known top of the earnings distribution (since the top is not known), the low and decreasing marginal tax rate on very high earners that comes from simulations using the lognormal distribution of skills (since the Pareto distribution is well documented to be a better fit), the argument for zero taxation of capital income from the aggregate efficiency result (since the theorem does not have that implication), the argument for zero taxation of capital income asymptotically (since bequest behavior does not conform with what is needed for this description of the asymptotic position of the economy), and the argument for zero taxation of capital income from the Atkinson-Stiglitz theorem (since savings rates are not uniform in the population).

Recommendations

We turn now to reviewing the prime bases for the policy recommendations that we do draw from the literature. Then, we contrast our methodological approach with some of what has been written in analyses employing mechanism design.

[27] "it [is] necessary for man with his limited powers to go step by step; breaking up a complex question, studying one bit at a time, and at last combining his partial solutions into a more or less complete solution of the whole riddle. ... The more the issue is thus narrowed, the more exactly can it be handled: but also the less closely does it correspond to real life. Each exact and firm handling of a narrow issue, however, helps towards treating broader issues, in which that narrow issue is contained, more exactly than would otherwise have been possible. With each step ... exact discussions can be made less abstract, realistic discussions can be made less inexact than was possible at an earlier stage." [Marshall, 1948, page 366.]

Recommendation 1. Tax (and transfer) policy toward low earners should include subsidization of earnings and should phase out the subsidization at a relatively high rate.

Traditional means-tested transfer programs typically have high phasing-out rates, often 100%.[28] However, in recent decades in most OECD countries, a concern arose that traditional welfare programs overly discouraged work and there has been a marked shift toward lowering the marginal tax rate at the bottom through a combination of (a) introduction and then expansion of in-work benefits such as the US EITC, (b) reduction of the statutory phasing-out rates in transfer programs for earned income (as under the US welfare reform), (c) reduction of payroll taxes for low income earners (as in the recent US Making Work Pay credit). Those reforms are consistent with the logic of optimal taxation we have outlined as they both encourage labor force participation and provide transfers to low income workers seen as a deserving group.

Such reforms were not the consequence of following recommendations from formal optimal tax theory as the reforms pre-date the development of the theory. However, the theory can now help in identifying the elements that should matter when setting the parameters of such a system, what Hahn (1973) has called the grammar of arguments about policy.[29] That is, one can learn how the size of the extensive elasticity should influence subsidy rates and how the size of the workforce at a point in the earnings distribution, relative to the population with higher earnings, should influence the rate at which the subsidy is phased out. This has not led basic research to precise recommended numbers, but a way to think about picking numbers, in light of the theory and of additional elements (e. g., degree of concern about income distribution and about

[28] Historically, most means-tested transfer programs started as narrow programs targeting specific groups deemed unable to earn enough such as widows with children, the elderly, or the disabled. For example, the ancestor the traditional US welfare program (Aid for Families with Dependent Children, renamed TANF after the 1996 welfare reform) were "mothers' pensions" state programs providing help primarily to widows with children and no resources (Katz, 1996). If beneficiaries cannot work but differ in terms of unearned income (for example, the presence of a private pension), then the optimal redistribution scheme is indeed a benefit combined with a 100% phasing-out rate.

[29] “Welfare economics is the grammar of arguments about policy, not the policy.” A role of analysis is “to make it much harder than it was to sustain an argument [contradicted by theory].” (page 106).

work per se). Having theoretical support for such a program may also encourage other countries to pursue such an approach.

> Recommendation 2. High earnings should be subject to rising marginal rates and higher rates than current US policy for top earners.

The argument for rising rates is primarily qualitative, coming from the nature of the first order conditions with the Pareto distribution and plausible properties of elasticities and social preferences. With an argument that the consumption of those with very high incomes (the top percentile) should be viewed as of small consequence relative to the consumption of average workers, it becomes possible to get roughly quantitative, at least as a function of beliefs about elasticities – we say beliefs since there are multiple empirical studies that shed some light on an estimate, but no definitive empirical study. Above we noted that a not implausible elasticity for a broad based income tax could be 0.17, translating into a level for the top tax wedge of 80%, implying a top federal income tax rate of 76%, given the current average state top income tax rate and sales tax rate.

Being a quantitative estimate, one can ask how seriously one should take the number in light of the elements not included in its derivation. Among the prime issues raised when considering this issue are the following. In the US income tax structure, the marginal tax rate is applied to net income incorporating both labor income and capital income, with the latter entering in part in a somewhat separate way (lower rates for capital gains and qualified dividends). Does the presence of capital income mean that earnings should be taxed significantly differently? So far we have not seen a decisive argument why that might be the case. If the tax system continues the current relative treatments of labor and capital incomes, does that argue for a lower tax rate since some capital income is included? Serious consideration of this issue would need to recognize the tax advantage for capital income from deferral and the larger impact of anticipated inflation on nominal capital income than on labor income. Models that have a constraint of uniform tax rates for bases that one might want to tax differently tend to have a tax rate lying between the two separately desirable rates. We return to this below. Should the elasticities be measured for labor incomes or total incomes, given the tax base and the substitutability of one for the other for many very high earners? Perhaps most critically, does a derivation in a single period model still apply when recognizing that people earn and pay income taxes year after year. Earlier decisions do affect earnings opportunities and there is intertemporal substitution in filling a lifetime budget constraint. It is not clear that any of these makes for large differences, but we do not have much to draw on to answer this question either way.

Recommendation 3. Capital income should be taxed.

We have seen four arguments for taxation – the difficulty of distinguishing between capital and labor incomes, the positive correlation between earnings opportunities and savings propensities, the role of capital income taxes in easing the tax burden on those who are borrowing constrained, and the role of discouraging savings in encouraging later labor supply. The arguments are based on lifecycle analyses. Yet the empirical literature finds that the lifecycle approach, while helpful, is limited in its success in explaining savings behavior. The belief, important in considering Social Security, that many people do not save enough for their own retirements and the current size of Social Security together call for policies to encourage savings, particularly retirement savings. The most widely used method is forced saving through mandatory contributions to social security retirement systems. This can also be complemented with a combination of taxing capital income and having tax-favored retirement savings (including some subsidies) targeted to those liable to save too little. A cap on allowable tax-favored amounts and subsidies for savings by very low earners serve such a targeting role.

Mechanism design for tax analysis

Optimality analyses of taxation have flourished in two (mostly) separate research communities. The public economics community has been actively doing optimal tax analyses since the mid 1960's. The standard optimal tax analysis begins with a set of allowable tax structures and optimizes the tax rates and/or tax bases in the allowable structure. The macro community (New Dynamic Public Finance - NDPF) has been active since the mid 1980's, using a mechanism design approach. Mechanism design only rules out taxes that are assumed to require information that the government does not have. Beyond this constraint (and the resource/technology constraint), there are no further restrictions, allowing complex structures that might be assumed as unavailable for being too complex. That is, individuals choose from the allowable set of complete lifetime consumption and earnings levels. The macro community has made future uncertain earnings opportunities central to its analyses, an element largely lacking in the optimal tax approach. This has made for significant advances in our understanding of this realistic phenomenon. The standard mechanism design approach derives each individual's marginal rates of substitution consistent with the individual's consumption and labor allocation. The next step is to describe a mechanism that can implement the allocation using familiar government tools.

Analysts using the two approaches sometimes differ in how they approach policy implications. While the public economics community looks for lessons for diverse settings, according to Kocherlakota (2009) which provides a comprehensive treatment of the NDPF, "The goal of this book is to figure out at least some characteristics of the best possible tax system." (Page 1.) Indeed Kocherlakota (2009) criticizes the Chamley-Judd approach[30] "We see that the Ramsey [linear tax] approach is disturbingly non-robust. In particular, the set of possible tax instruments makes a big difference in the answers to various optimal tax questions. (Page 27.)" and "These are but two examples of a general problem with the Ramsey approach: the answers depend critically on the set of possible instruments. … This general lack of robustness is due to a fundamental limitation of the Ramsey approach: it takes the set of possible tax instruments as given. We need to *endogenize* the government's set of possible taxes in some fashion. (page 30.)"[31] Indeed a lack of robustness in this sense is a virtue not a vice when considering recommendations in a setting with a limited agenda for tax policy. While lessons from mechanism design have added to our understanding of taxation, this methodological narrowness rejects analyses that might be directly on point for a government considering a limited tax reform rather than a complete tax optimization. This may occur out of

[30] The NDPF community refers to this as Ramsey taxation, a term with a somewhat different meaning to the public economics community, and so avoided here.
[31] This quote is consistent with some personal interactions suggesting that some in the macro community take mechanism design to be the only right way of doing optimal tax analyses.

political feasibility or to recognize a value in historical continuity or for reasons not captured in the usual mechanism design formulation (such as limits in acceptable complexity and record-keeping requirements). Therefore, in our view, limited tax reform analysis can inform relevant policy questions and hence should not be rejected on methodological principles.[32]

Using the revelation principle, the mathematical analysis looks for the best allocation derived from learning each person's true type. In thinking about an entire economy with a large number of people and a complex uncertain future, one can no more identify all types and infer what type each person is than we can have a complete set of Arrow-Debreu markets. Both are impossible, starting with listing all the states of nature or fully describing all the types. This does not mean that studying these allocations is a waste of time, but rather that one needs to be careful in choosing what lessons one takes away from such a study, including lessons from examination of allocations in less general, but more realistic settings.

Payoff design in optimal contracting theory calls for making use of any variable correlated with the key unobservable variable (Holmstrom, 1979). For actual tax systems, this is not feasible. We think it important to recognize that a model (for example a game-theoretic equilibrium) that may be perfectly sensible with a small number of sophisticated agents may not be helpful for a large population with limitations in attention to long-term consequences, information about tax structures and payoff possibilities. In particular, complexity is a key issue that we do not know how to model. Hence we substitute judgment for formal modeling in letting complexity concerns affect the lessons we take from formal modeling rather than assuming that the degree of complexity is irrelevant. More concretely, legislators, tax administrators and taxpayers have limited abilities to design, enforce and comply with complex tax structures. Hence the design of tax policies should not rely on a structure that is so complex that the supporting model does not describe behavior well. Since we do not have a good way to incorporate complexity limitations into economic equilibrium analyses, there is little

[32] Another difference is that the public economics community draws on multiple models, and so often seeks insights from most analyses, not precise answers. But Kocherlakota (2009) says: "The ultimate goal of the NDPF is to provide relatively precise recommendations as to what taxes should be." (Page 5.)

choice but to assume restrictions rather than deriving them. Of course, one could pretend to derive them by simply labeling different tax structures as too complex and not too complex. That would not be very different from some treatments of "observability" – where different measures, all with some cost and all with some inaccuracy are labeled either costlessly and perfectly available or simply unavailable.[33] Just as recognizing complexity should limit allowable tools, there is a similar role for public perceptions of tax fairness.

For us, policy relevance is a large part of the appeal of working on the intellectually interesting optimal tax problems. That relevance relies on methodologically suitable use of the findings. We have tried to spell out the methodology we have used in reaching the policy implications we have drawn.

[33] Thus we disagree with Hahn (1973), "that it is a mistake to import unexplained second-best constraints into a model which leaves no room for their justification." We think that model tractability makes it appropriate to assume rather than derive plausible conditions when one thinks the two approaches would lead to the same central conclusion, even though, of course, some other conclusions would not carry over.

References:

Aiyagari, S. Rao. 1995. “Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting.” *Journal of Political Economy*, 103(6): 1158-75.

Atkinson, Anthony, Thomas Piketty, and Emmanuel Saez. 2010. “Top Incomes in the Long Run of History”, forthcoming *Journal of Economic Literature*.

Atkinson, Anthony B. and Agnar Sandmo. 1980. “Welfare Implications of the Taxation of Savings.” *Economic Journal*, 90 (September): 529-549.

Atkinson, Anthony, and Joseph E. Stiglitz. 1976. “The Design of Tax Structure: Direct Versus Indirect Taxation.” *Journal of Public Economics*, 6(1–2): 55–75.

Auerbach, Alan. 1988. "Capital Gains Taxation in the United States." *Brookings Papers on Economic Activity*, 2: 595-631.

Banks, James and Peter Diamond. 2009 “The Tax Base,” in *Dimensions of Tax Design: The Mirrlees Review*, Oxford University Press for Institute for Fiscal Studies (Oxford).

Chamley, Christophe. 1986. “Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives.” *Econometrica*, 54(3): 607–22.

Christiansen, Vidar, and Matti Tuomala. 2008. “On taxing capital income with income shifting.” *International Tax and Public Finance*, 15: 527–545.

Cremer, Helmuth, and Pierre Pestieau. 2004. “Wealth Transfer Taxation: A Survey of the Theoretical Literature**.”** in S.C. Kolm and J. Mercier Ythier, eds, *Handbook on Altruism, Giving and Reciprocity*, North Holland, Amsterdam.

Diamond, Peter. 1980. "Income Taxation with Fixed Hours of Work," *Journal of Public Economics* 13, February, 101-110.

Diamond, Peter. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review*, 88(1): 83-95.

Diamond, Peter. 2003, *Taxation, Incomplete Markets and Social Security, the 2000 Munich Lectures*, Cambridge: MIT Press.

Diamond, Peter A., and James A. Mirrlees. 1971. “Optimal Taxation and Public Production I: Production Efficiency.” *American Economic Review*, 61(1): 8–27.

Diamond, Peter and James Mirrlees. 1982. "Social Insurance with Variable Retirement and Private Saving", MIT wp#296 Feb 1982.

Diamond, Peter and James Mirrlees. 1986. "Payroll-Tax Financed Social Insurance with Variable Retirement," *Scandinavian Journal of Economics* 88 (1): 25-50.

Diamond, Peter and James Mirrlees. 2000. "Adjusting One's Standard of Living: Two Period Models," in P. J. Hammond and G. D. Myles (eds.), Incentives, Organization, and Public Economics, papers in Honour of Sir James Mirrlees, Oxford; Oxford University Press.

Diamond, Peter and Johannes Spinnewijn. 2010. "Capital Income Taxes with Heterogeneous Discount Rates", MIT working paper.

Golosov, Mikhail, Narayana Kocherlakota and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." *Review of Economic Studies*, 70(3): 569-87.

Golosov, Mikhail and Aleh Tsyvinski. 2006. "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy*, 114(2): 257-69.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2009. "A Quantitative Exploration in the Theory of Dynamic Optimal Taxation," Yale University working paper.

Golosov, Mikhail, Aleh Tsyvinski, and Matthew Weinzierl. 2009. "Preference Heterogeneity and Optimal Commodity Taxation" Yale Working Paper.

Goolsbee, Austan. 2000. "What Happens When You Tax the Rich? Evidence from Executive Compensation." *Journal of Political Economy*, 108(1): 352-378.

Gordon, Roger H., and Jeffrey K. MacKie-Mason. 1995. "Why is There Corporate Taxation in a Small Open Economy?" In *The Effects of Taxation on Multinational Corporations*, eds. Martin Feldstein, James R. Hines, Jr., and R. Glenn Hubbard. Chicago: University of Chicago Press.

Gordon, Roger and Joel Slemrod. 2000. "Are 'Real' Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?" In Joel Slemrod, ed., *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, Harvard University Press: Cambridge.

Gruber, Jonathan and Emmanuel Saez. 2002. "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics*, 84: 1-32.

Hahn, Frank H. 1973. "On Optimum Taxation," *Journal of Economic Theory* 6(1): 96-106.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability," *The Bell Journal of Economics*, 10(1): 74-91.

Internal Revenue Service. 2009. *Statistics of Income: Individual Income Tax Returns for Year 2007*, Publication No. 1304, (Washington, DC: Government Printing Press).

Internal Revenue Service. 2009b. "The 400 Individual Income Tax Returns Reporting the Highest Adjusted Gross Incomes Each Year, 1992-2007", (Washington, DC: Government Printing Press).

Judd, Kenneth. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." *Journal of Public Economics*, 28(1): 59–83.

Judd, Kenneth L. 1999. "Optimal taxation and spending in general competitive growth models," *Journal of Public Economics*, 71: 1-26.

Kaplow, Louis. 2006. "On the Undesirability of Commodity Taxation Even When Income Taxation Is Not Optimal." *Journal of Public Economics*, 90(6-7): 1235-50.

Katz, Michael B. 1996. *In the Shadow of the Poorhouse: A Social History of Welfare in the United States*. New York, NY, Basic Books, 2nd edition.

Kocherlakota, Narayana. 2005. "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation." *Econometrica*, 73(5): 1587-621.

Kocherlakota, Narayana R. 2009. *The New Dynamic Public Finance*. www.econ.umn.edu/~nkocher/ndpf_0608.pdf

Kopczuk, Wojciech. 2005. "Tax Bases, Tax Rates and the Elasticity of Reported Income," *Journal of Public Economics*, 89(11-12): 2093-2119.

Laroque, Guy R. 2005. "Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof." *Economics Letters*, 87(1): 141-4.

Mankiw, N. Gregory, and Matthew Weinzierl. 2010. "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution." *American Economic Journal: Economic Policy*, forthcoming.

Mankiw, N. Gregory, Matthew C. Weinzierl, and Danny Yagan. 2009. "Optimal Taxation in Theory and Practice." *Journal of Economic Perspectives* 23(4): 147-174.

Marshall, Alfred. 1948. *Principles of Economics*, 8th ed. New York: Macmillan.

Meade, James Edward. 1978. *The Structure and Reform of Direct Taxation, Report of a Committee chaired by Professor J. E. Meade*. London: George Allen & Unwin.

Meyer, Bruce. 2010. "The Effects of the EITC and Recent Reforms", *Tax Policy and the Economy*, ed. Jeffrey Brown, (Cambridge: MIT Press), Volume 24.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic studies*, 38, 175-208.

Moffitt, Robert and Mark Wilhelm. 2000. "Taxation and the Labor Supply Decisions of the Affluent." *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, Joel Slemrod, ed., Harvard University Press: Cambridge.

Pareto, Vilfredo. 1896. "La courbe de la répartition de la richesse", *Ecrits sur la courbe de la répartition de la richesse*, (writings by Pareto collected by G. Busino, Librairie Droz, 1965), 1-15.

Piketty, Thomas. 1997. "La Redistribution Fiscale face au Chomage." *Revue Francaise d'Economie*, 12(1), 157-201.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, 118(1), 1-39, series updated to 2007 in August 2009.

Piketty, Thomas and Emmanuel Saez. 2007 "How Progressive is the U.S. Federal Tax System? A Historical and International Perspective", Journal of Economic Perspectives, 21(1), 3-24.

Pirttilä, Jukka, and Håkan Selin. 2007. Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993. unpublished.

Reis, Catarina. 2007. "Entrepreneurial Labour Taxation." Dissertation chapter, MIT.

Sadka, Efraim. 1976. "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies*, 43(1): 261-268.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economic Studies*, 68: 205-229.

Saez, Emmanuel. 2002. "Optimal Income Transfer Programs: Intensive Versus Extensive Labour Supply Responses." *Quarterly Journal of Economics*, 117(2): 1039-73.

Saez, Emmanuel. 2002b. "The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes." *Journal of Public Economics*, 83(2): 217–230.

Saez, Emmanuel. 2004. "Reported Incomes and Marginal Income Tax Rates, 1960-2000: Evidence and Policy Implications", *Tax Policy and the Economy*, ed. James Poterba, (Cambridge: MIT Press), Volume 18.

Saez, Emmanuel. 2004b. "The Optimal Treatment of Tax Expenditures", *Journal of Public Economics*, 88(12): 2657-2684.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" forthcoming *American Economic Journal: Economic Policy*.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz. 2010. “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review”, forthcoming in the *Journal of Economic Literature*.

Seade, Jesus K. 1977. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, 7(1): 203-236.

Slemrod, Joel. 1995. "Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986." *American Economic Review*, 85(1): 175-80.

Slemrod, Joel, ed. 2000. *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, Harvard University Press: Cambridge.

Slemrod, Joel and Wojciech Kopczuk. 2002. "The Optimal Elasticity of Taxable Income", *Journal of Public Economics*, 2002, 84(1): 91-112.

US Census Bureau. 2009. "Median Household Income for States: 2007 and 2008", American Community Surveys, September 2009.

Vickrey, William. 1947. *Agenda for Progressive Taxation*, Ronald Press Company: New York.

Zeldes Stephen P. 1989. “Consumption and Liquidity Constraints: An Empirical Investigation,” *Journal of Political Economy*, 97(2): 305-346.

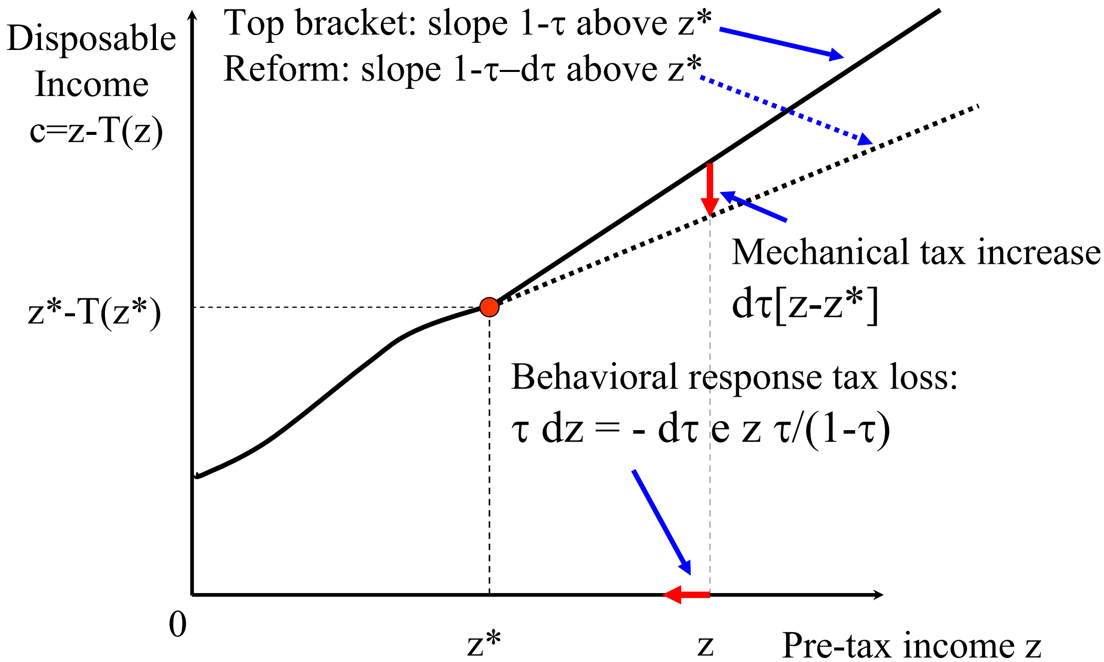

Figure 1
Optimal Top Tax Rate Derivation
Note. The figure depicts the derivation of the optimal top tax rate formula (1), $\tau^*=1/(1+ae)$ by considering a small reform around the optimum which increases the top marginal tax rate τ by $d\tau$ above z^*. A taxpayer with income z mechanically pays $d\tau\ [z-z^*]$ extra taxes but also reduces his income by $dz=e\ z\ d\tau/(1-\tau)$ leading to a loss in tax revenue equal $d\tau\ e\ z\ \tau/(1-\tau)$. Summing across all top bracket taxpayers and denoting by z_m the average income above z^* and $a=z_m/(z_m-z^*)$), we obtain the revenue maximizing tax rate $\tau^*=1/(1+ae)$, which is the optimum tax rate when the government sets zero marginal welfare weights on top income earners.

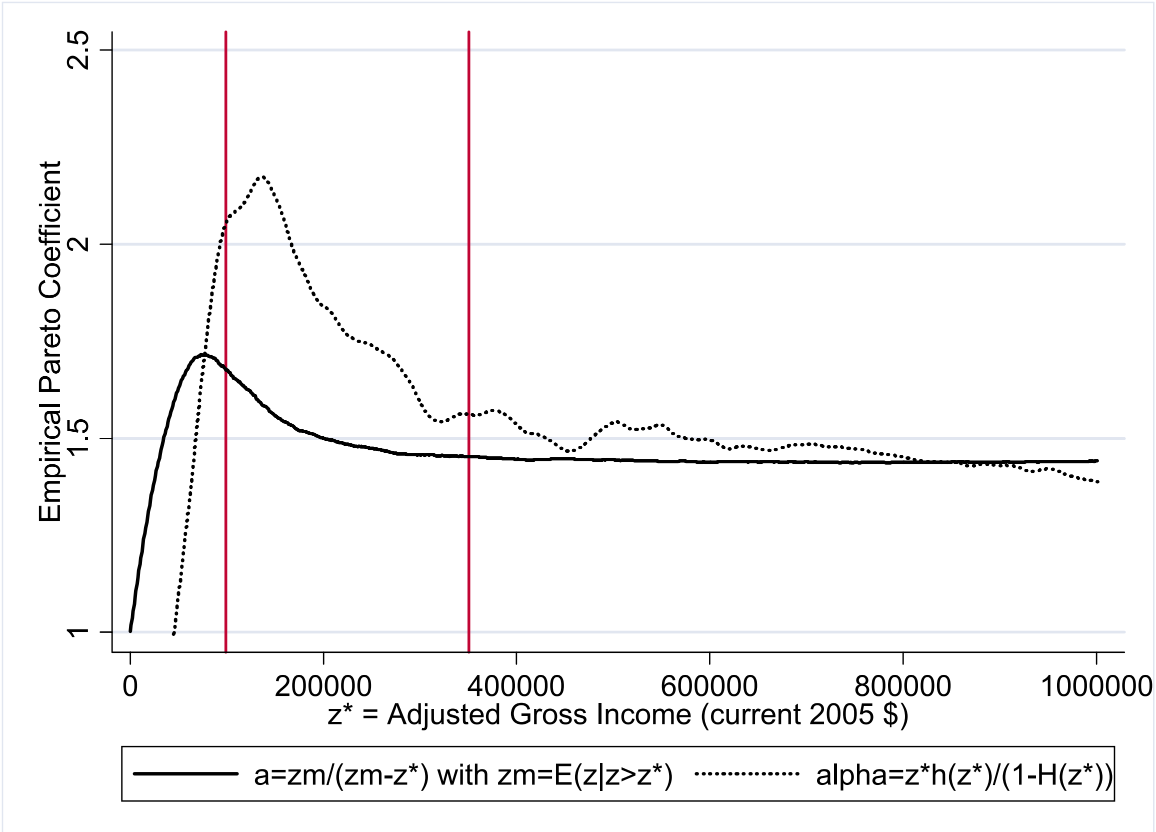

Figure 2

Empirical Pareto Coefficients in the United States, 2005

Note. The figure depicts in solid line the ratio a=zm/(zm-z*) with z* ranging from $0 to $1,000,000 annual income and zm the average income above z* using US tax return micro data for 2005. Income is defined as Adjusted Gross Income reported on tax returns and is expressed in current 2005 dollars. Vertical lines depict the 90th percentile ($99,200) and 99th percentile ($350,500) nominal thresholds as of 2005. The ratio a is equal to one at z*=0, and is almost constant above the 99th percentile and slightly below 1.5, showing that the top of the distribution is extremely well approximated by a Pareto distribution for purposes of implementing the optimal top tax rate formula (1), $\tau^*=1/(1+a\ e)$. Denoting by h(z) the density and by H(z) the cdf of the income distribution, the figure also displays in dotted line the ratio $\alpha(z^*)=z^*h(z^*)/(1-H(z^*))$ which is also approximately constant, around 1.5, above the top percentile. A decreasing (or constant) $\alpha(z)$ combined with a decreasing G(z) and a constant e implies that the optimal marginal tax rate from formula (2), $T'(z)=[1-G(z)]/[1-G(z) + \alpha(z)\ e]$ increases with z.

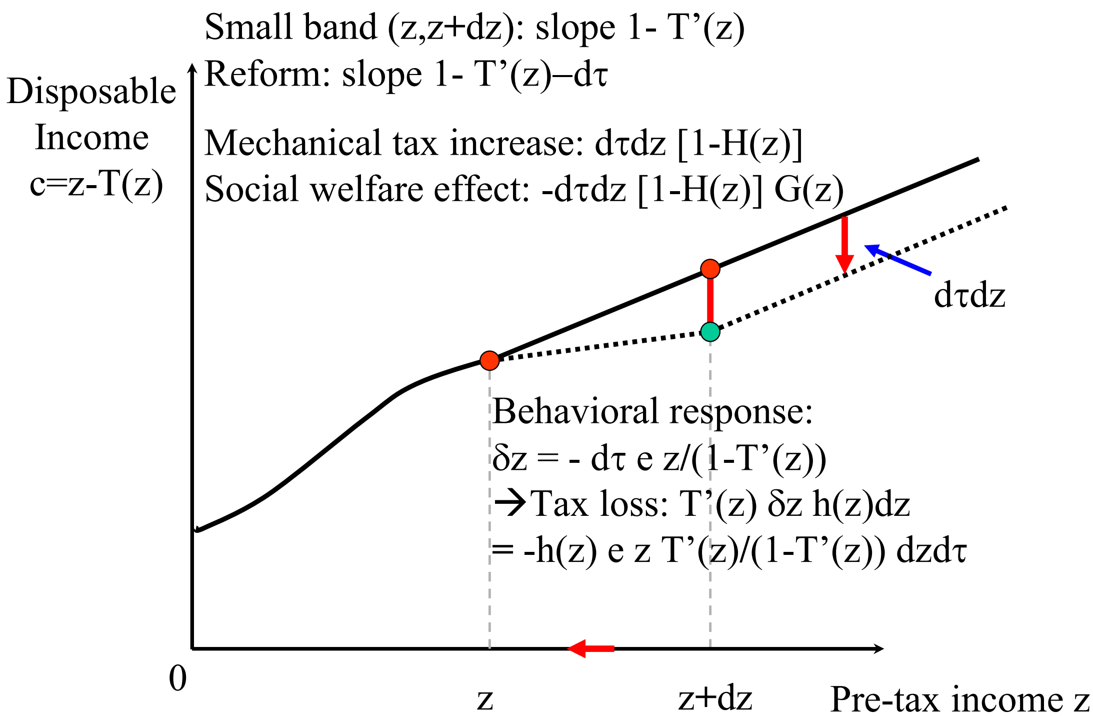

Figure 3

Derivation of the Optimal Marginal Tax Rate at Income Level z

Note. The figure depicts the optimal marginal tax rate derivation at income level z by considering a small reform around the optimum, whereby the marginal tax rate in the small band (z,z+dz) is increased by dτ. This reform mechanically increases taxes by dτdz for all taxpayers above the small band, leading to a mechanical tax increase dτdz[1-H(z)] and a social welfare cost of -dτdz[1-H(z)]G(z). Assuming away income effects, the only behavioral response is a substitution effect in the small band: The h(z)dz taxpayers in the band reduce their income by δz=-dτ e z/(1-T'(z)) leading to a tax loss equal to -h(z) e z T'(z)/(1-T'(z))dz dτ. At the optimum, the three effects cancel out leading to the optimal tax formula (2), T'(z)/(1-T'(z))=(1/e) (1-G(z)) (1-H(z))/(zh(z)), or equivalently T'(z)=[1-G(z)]/[1-G(z)+α(z) e] after introducing α(z)=zh(z)/(1-H(z)).

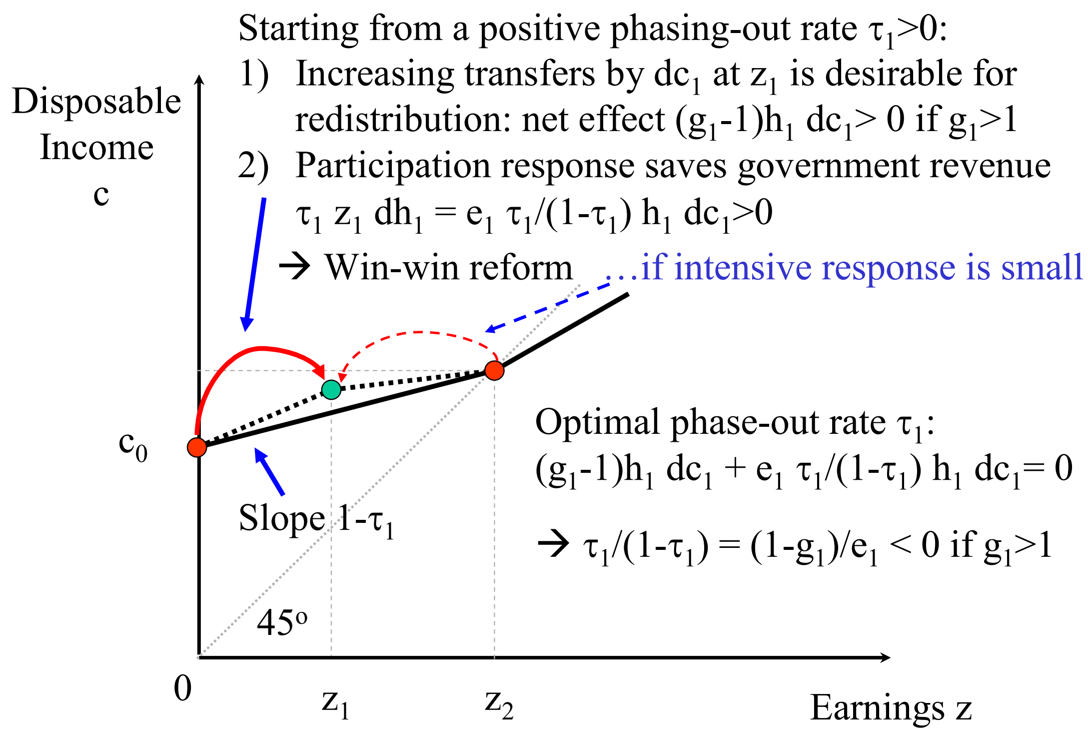

Figure 4

Optimal Bottom Marginal Tax Rate with Extensive Labor Supply Responses

Note. The figure depicts the derivation of the optimal marginal tax rate at the bottom in the discrete model with labor supply responses along the extensive margin only. Starting with a positive phasing out rate $\tau_1>0$, the government introduces a small in-work benefit dc_1. Let h_1 be the fraction of low income workers with earnings z_1, and elasticity e_1 with respect to the participation net-of-tax rate $1-\tau_1$. The reform has three standard effects: mechanical fiscal cost $dM=-h_1dc_1$, social welfare gain, $dW=g_1h_1dc_1$, and tax revenue gain due to behavioral responses $dB= \tau_1z_1dh_1= e_1\tau_1/(1-\tau_1)h_1dc_1$. If $g_1>1$, then $dW+dM>0$. If $\tau_1>0$, then $dB>0$ implying that $\tau_1>0$ cannot be optimal. The optimal τ_1 is such that $dM+dW+dB=0$ implying that $\tau_1/(1-\tau_1)=(1-g_1)/e_1$.

Appendix:

1. Comparison with Mankiw, Weinzierl, and Yagan (2009)

Mankiw, Weinzierl and Yagan (2009) (hereafter MWY) present 8 lessons that they draw from the optimal tax literature. Our paper, agrees with some of their lessons but also draws some very different conclusions. In this appendix, we discuss some of the discrepancies between our interpretations following the order of the 8 lessons presented in MWY.

Lesson 1: Optimal Marginal Tax Rate Schedules Depend on the Distribution of Ability: We agree.

Lesson 2: The Optimal Marginal Tax Schedule Could *Decline* at High Incomes: Major disagreement.

MWY derive this lesson based on two arguments. First, they present the famous zero top marginal tax rate result which, combined with positive marginal tax rates below the top, implies that the tax rate should decline as it approaches the top. Second, they discuss evidence from numerical simulations using log-normal skill distributions showing that rates are modest and sometimes decreasing in the upper part of the distribution. They dismiss the results which use Pareto distributions and which obtain high tax rates on upper incomes on two grounds. First, they claim that we cannot infer the ability distribution without making unduly strong assumptions. Second, they examine the right tail of the wage density distribution using Current Population Survey (CPS) data (Figure 1, p. 154) and conclude that it is not possible to distinguish Pareto vs. log-normal distributions from such data. We find both of those arguments invalid – the zero at the top is not relevant for policy, as discussed above, and the evidence is strongly supportive of the Pareto distribution[34] and does not require unusually strong assumptions.[35]

[34] For example, all the recent top income share studies summarized in Atkinson, Piketty, Saez (2010) make use of the Pareto fit to construct estimates using published tax statistics tabulated by brackets for 22 countries over long-time periods. For any distribution with a thinner top tail that the Pareto distribution, such as a log-normal distribution, the parameter $a=z_m/(z_m-z^*)$ diverges to infinity. The test of Pareto vs.

Lesson 3: A Flat Tax, with a Universal Lump-Sum Transfer, Could Be Close to Optimal: Major disagreement.

The analysis we presented showed that, at the bottom, transfers initially increase with earnings to preserve incentives to participate in the labor force and then are phased-out with income at a high rate. MWY does not discuss the participation margin. After transfers are phased-out, marginal tax rates should be lower for the broad middle class and then rise in the upper income groups due to declining marginal welfare weights and the Pareto shape of the income distribution toward the top. Therefore, the optimal system appears quite different from a flat tax with a universal lump-sum transfer advocated by MWY.

Lesson 4: The Optimal Extent of Redistribution Rises with Wage Inequality: We agree.

Lesson 5: Taxes Should Depend on Personal Characteristics As Well As Income: Some disagreement

While a model ignoring both complexity and social acceptability would reach this conclusion for many observable characteristics, we think that these two issues should get due respect. In practice, taxes and transfers depend significantly on only few characteristics (besides income) and those characteristics such as family structure or disability status are related to need.

Lesson 6: Only Final Goods Ought to be Taxed, and Typically They Ought to be Taxed Uniformly: Some disagreement

Limiting variation in commodity (or VAT) taxes is appropriate, but some variation seems well justified, although too much variation seems present in some systems. A central part

---

lognormal right tails presented by Mankiw et al. (2009) in their Figure 1, p. 154 lacks power because it uses CPS data that is thin at the top and it plots density fits when the statistic of central interest is $a=z_m/(z_m-z^*)$ as depicted on Figure 2.

[35] MWY cite the simulation results of Saez (2001) which by necessity require making functional form assumptions but fail to note that the general theoretical tax rate formula $\tau=1/(1+ae)$ is much more general than the numerical illustration.

of our presentation is our disagreement with their inference that this line of argument supports not taxing capital income.

Lesson 7: Capital Income Ought to Be Untaxed, At Least in Expectation: Major Disagreement

MWY invoke the three arguments for zero capital income taxes that we have addressed. As we discussed in the paper, the Diamond-Mirrlees result does not imply that capital income should not be taxed; the Chamley-Judd result relies heavily on empirically implausible bequest behavior; and the Atkinson-Stiglitz result ignores future earnings uncertainty and needs strong assumptions about savings behavior that do not hold in practice.

Lesson 8: In Stochastic Dynamic Economies, Optimal Tax Policy Requires Increased Sophistication: Some disagreement

We agree that stochastic elements call for more sophisticated analysis and justify more sophisticated structures. However, we disagree with their emphasis on the optimality of a regressive interaction between capital income taxation and labor income that raises no revenue in expectation (referred to in the title of Lesson 7). Their focus on the zero expected revenue refers to one implementation of the mechanism design optimum, which calls for discouraging savings, but ignores the presence of a different implementation that has positive taxation of capital (Werning, 2009) and ignores the issue of designing optimal taxes when limited complexity implies that the full mechanism design optimum is not being implemented.

2. Robustness of the Top Rate Formula $\tau=1/(1+a\ e)$ with a finite distribution

Let us consider a finite population of income earners drawn from an underlying distribution with a Pareto upper tail. More precisely, let us assume that realized earnings of individual i are equal to $z_i=n_i\ (1-\tau)^e$ where τ is the marginal tax rate, e the elasticity of earnings with respect to $1-\tau$, and n_i potential earnings (realized earnings when there is no marginal tax rate).[36] We assume that potential earnings n_i are drawn from an underlying distribution with Pareto upper tail with Pareto parameter a. We realistically assume that the government does not know the realized draw of the n_i when setting its tax policy. In this model, the realized earnings distribution will be finite but the level of the very top earner is unknown ex-ante to the government.

We further assume that the government tries to maximize expected tax revenue collected from the very top of the distribution. Revenue maximization corresponds to setting a zero welfare weight on the marginal consumption of those with top incomes (this assumption could easily be relaxed without affecting the results relative to the continuum model). The maximization of expected revenue is a sensible shortcut to avoid exploring possible relations between the revenue from the very top of the earnings distribution and the value of resources to the government across states of nature.

Theorem: Suppose the starting tax system has a top marginal tax rate equal to $\tau=1/(1+a\ e)$ (say above z* where the distribution is Pareto). Then for any $z>z^*$, any small change in the marginal tax rate above z decreases expected tax revenue. Therefore, $\tau=1/(1+a\ e)$ remains the top tax rate that maximizes expected revenue even with a finite distribution.

Proof: Assume that the government changes the marginal tax rate from τ to $\tau+d\tau$ above z. An individual with realized earnings $z_i<z$ is not affected, assuming n_i is known to the individual before determining z_i. An individual with realized earnings $z_i>z$ mechanically pays $d\tau(z_i-z)$ in extra taxes but changes his earnings by $dz_i=-z_i\ e\ d\tau/(1-\tau)$. Therefore, the net change in taxes paid by individual i is $dR_i=d\tau\ [z_i-z-\ z_i\ e\ \tau/(1-\tau)]$. Because the n_i-and hence the $z_i=n_i(1-t)^e$-are Pareto distributed among those with earnings above z, and because the average earnings above any given threshold z is $z\ a/(a-1)$ for a Pareto distribution, the expected value of z_i is $z\ a/(a-1)$. Therefore, the expected change in tax revenue is $dR=N(z)\ z/(a-1)\ d\tau\ [1-\ a\ e\ \tau/(1-\tau)]$, where $N(z)$ is the expected number of individuals above z. Because $\tau=1/(1+a\ e)$, we have $dR=0$, which proves the theorem.

[36] Quasilinear utility functions of the form $u(c,z)=c-n(z/n)^{1+1/e}/(1+1/e)$ would generate such earnings supply functions.

Remark: For simplicity, we have assumed homogeneity of the elasticity e across individuals. We conjecture that the theorem could be extended to heterogeneous populations assuming that the expected elasticity e among top earners above z is constant with z. Basically, whenever the asymptotic tax rate $\tau=1/(1+a\ e)$ applies in the continuum, it also applies in the case of the finite draw.

Considering a known finite distribution, rather than a bounded continuum, we do have the result that the *marginal* tax rate for the highest earner is zero, but the average tax rate over the range of earnings from the second-highest earner to the highest earner is large enough so that the highest earner has the same utility with the chosen earnings and with matching the earnings of the next highest earner. With a sizable gap we would expect from draws from a Pareto distribution the *average* tax rate over this range would need to be significant.

3. Derivation of the optimal marginal tax rate at the bottom in the Mirrlees model

For expositional simplicity, let us consider a discrete version of the Mirrlees (1971) model developed in Piketty (1997) and Saez (2002). As illustrated on Figure A1, suppose that low ability individuals can choose either to work and earn z_1 or not work and earn zero. The government offers a transfer c_0 to those not working phased out at rate τ_1 so that those working receive on net $c_1=(1-\tau_1)\ z_1 + c_0$. In words, non-workers keep a fraction $1-\tau_1$ of their earnings should they work and earn z_1. Therefore, increasing τ_1 discourages some low income workers from working. Let us denote by H_0 the fraction of non-workers in the economy and by $e_0=-(1-\tau_1)/H_0\ dH_0/d(1-\tau_1)$ the elasticity of non-workers H_0 with respect to the net-of-tax rate $1-\tau_1$, where the minus sign is used so that $e_0>0$.

Suppose now that the government increases both the maximum transfer by dc_0 and the phasing-out rate by $d\tau_1$ leaving the tax schedule unchanged for those with income above z_1 so that $dc_0= z_1\ d\tau_1$ as depicted on Figure A1. The fiscal cost is $-H_0\ dc_0$ but the welfare benefit is $H_0\ g_0\ dc_0$ where g_0 is the social welfare weight on non-workers. If the government values redistribution, then $g_0>1$ and g_0 is potentially large as non-workers are the most disadvantaged. Because behavioral responses take place along the intensive margin only in the Mirrlees model, with no income change above z_1, the labor supply of those above z_1 are not affected by the reform. A number $dH_0=d\tau_1\ e_0\ H_0\ /\ (1-\tau_1)$ of low income workers stop working creating a revenue loss of $\tau_1\ z_1\ dH_0 = dc_0\ H_0\ e_0\ \tau_1/(1-\tau_1)$. At the optimum, the three effects sum to zero leading to the optimal bottom rate formula:

$$\text{(A1)} \quad \tau_1/(1-\tau_1)=(g_0-1)/e_0 \text{ or } \tau_1=(g_0-1)/(g_0-1+e_0)$$

Because g_0 is large, τ_1 will also be large. For example, if $g_0=3$ and $e_0=0.5$ (an elasticity sitting in the mid-range of empirical estimates), then $\tau_1=2/2.5=80\%$--a very high phasing-out rate. Formula (A1) is the optimal marginal tax rate at zero earnings in the standard Mirrlees (1971) model when there is an atom of non-workers, which is the most realistic case (this result does not seem to have been noticed in the literature). As is well known since Seade (1977), the optimal bottom tax rate is zero when everybody works and bottom earnings are strictly positive but this case is not practically relevant.

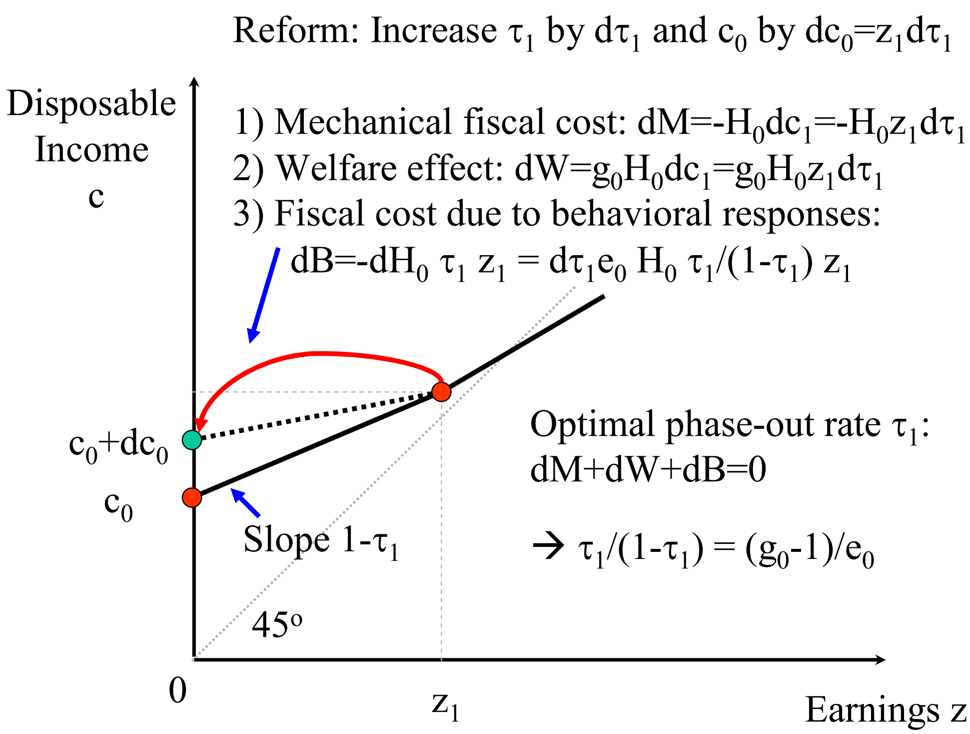

Figure A1

Optimal Bottom Marginal Tax Rate with only Intensive Labor Supply Responses

Note. The figure depicts the derivation of the optimal marginal tax rate at the bottom in the discrete Mirrlees (1971) model with labor supply responses along the intensive margin only. Let H_0 be the fraction of the population not working. This is a function of $1-\tau_1$, the net-of-tax rate at the bottom, with elasticity e_0. We consider a small reform around the optimum where the government increases the maximum transfer by dc_0 by increasing the phasing-out rate by $d\tau_1$ leaving the tax schedule unchanged for those with income above z_1, this creates three effects which cancel out at the optimum. At the optimum, we have $\tau_1/(1-\tau_1)=(g_0-1)/e_0$ or $\tau_1=(g_0-1)/(g_0-1+e_0)$. Under standard redistributive preferences, g_0 is large implying that τ_1 is large.