Journal of Economic Perspectives—Volume 25, Number 4—Fall 2011—Pages 165–190

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

Peter Diamond and Emmanuel Saez

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 basic research results 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, 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 as a function of individual utilities—the sum of utilities in the utilitarian case. The marginal weight for a given person in the social welfare function 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 combines with constraints arising from limits on resources and behavioral responses to taxation in order to derive specific

■ *Peter Diamond is Professor Emeritus of Economics, Massachusetts Institute of Technology, 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.*

† There is an Appendix at the end of this article. To access an additional online Appendix, visit http://www.aeaweb.org/articles.php?doi=10.1257/jep.25.4.165.

doi=10.1257/jep.25.4.165

tax policy recommendations. In addition, optimal income tax theory can be used to evaluate current policies and suggest avenues for reform. Understanding what would be good policy, if implemented, is a key step in making policy recommendations.

When done well, moving from mathematical results, theorems, or calculated examples to policy recommendations is a subtle process. The nature of a model is 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 models; hence our reliance on trying to draw inferences simultaneously from multiple models.

In our view, a theoretical result can be fruitfully used as part of forming a policy recommendation only if three conditions are met. First, the result should be based on an economic mechanism that is empirically relevant and first order to the problem at hand. Second, the result should 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 intertemporal choice. Therefore, we should view with suspicion results that depend critically on very strong homogeneity or rationality assumptions. Deriving 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—that is, 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. By socially acceptable, we do not mean to limit the choice to currently politically plausible policy options. Rather, we mean there should not be very widely held normative views that make such policies seem implausible and inappropriate at pretty much all times. For example, a policy prescription such as taxing height (Mankiw and Weinzierl, 2010) is obviously not socially acceptable because it violates certain horizontal equity concerns that do not appear in basic models. The complexity constraint can also be an issue when optimal taxes depend in a complex way on the full history of earnings and consumption, as in some recent path-breaking papers on optimal dynamic taxation.

*

We obtain three policy recommendations from basic research that we believe can satisfy these three criteria reasonably well. First, very high earners should be subject to high and rising marginal tax rates on earnings. In particular, we discuss why the famous zero marginal tax rate at the top of the earnings distribution is not policy relevant. 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 follows because labor supply responses of low earners are concentrated along the margin of whether to participate in labor markets at all (the extensive as opposed to the intensive margin). These 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 lump sum "demogrants." Third, we argue that capital income should be taxed. We will review certain theoretical results—in particular, those of Atkinson and Stiglitz (1976), Chamley (1986), and Judd (1985)—implying no capital income taxes and argue that these findings are not robust enough to be policy relevant. In the end, persuasive arguments for taxing capital income are that there are difficulties in practice in distinguishing between capital and labor incomes, that borrowing constraints make full reliance on labor taxes less efficient, and that savings rates are heterogeneous. *

The remainder of the paper is organized as follows: First, we consider the taxation of very high earners, second, the taxation of low earners, and third, the taxation of capital income. We conclude with a discussion of methodology, contrasting optimal tax and mechanism design ("new dynamic public finance") approaches. In an appendix, we contrast our lessons from optimal tax theory with those of Mankiw, Weinzierl, and Yagan (2009), recently published in this journal.

* Recommendation 1: Very high earnings should be subject to rising marginal rates and higher rates than current U.S. policy for top earners.

The share of total income going to the top 1 percent of income earners (those with annual income above about $400,000 in 2007) has increased dramatically from 9 percent in 1970 to 23.5 percent 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, 2011). Although the average federal individual income tax rate of top percentile tax filers was 22.4 percent, the top percentile paid 40.4 percent of total federal individual income taxes in 2007 (IRS, 2009a). Therefore, the taxation of very high earners is a central aspect of the tax policy debate not only for equity reasons but also for revenue raising. For example, setting aside behavioral responses for a moment, increasing the average federal income tax rate on the top percentile from 22.4 percent (as of 2007) to 29.4 percent would raise revenue by 1 percentage point of GDP.[1] Indeed, even increasing the average federal income tax rate of the top percentile to 43.5 percent, which would be sufficient to raise revenue by 3 percentage points of GDP, would still leave the after-tax income share of the top percentile more than twice as high as in 1970.[2] Of course, increasing upper income tax rates can discourage economic activity through behavioral responses, and hence *

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

[2] The average federal individual tax rate paid by the top percentile was 25.7 percent in 1970 (Piketty and Saez, 2007) and 22.4 percent in 2007 (IRS, 2009a). The overall average federal individual tax rate was 12.5 percent in 1970 and 12.7 percent in 2007. The pre-tax income share for the top percentile of tax filers was 9 percent in 1970 and 23.5 percent in 2007. Hence, the top 1 percent after-tax income share in 1970 was 7.6 percent = 9% × (1 − .257)/(1 − .125), and in 2007 it was 20.9 percent = 23.5% × (1 − .224)/(1 − .127) and, with a tax rate of 43.5 percent on the top percentile (which would increase the average tax rate to 17.7 percent), would have been 16.1 percent = 23.5% × (1 − .435)/(1 − .177).

potentially reduce tax collections, creating the standard equity-efficiency trade-off discussed in the introduction.

The Optimal Top Marginal Tax Rate

For the U.S. economy, the current top income marginal tax rate on earnings is about 42.5 percent,[3] combining the top federal marginal income tax bracket of 35 percent with the Medicare tax and average state taxes on income and sales.[4] As shown in Saez (2001), the optimal top marginal tax rate is straightforward to derive. Denote the tax rate in the top bracket by τ. Figure 1 shows how the optimal tax rate is derived. The horizontal axis of the figure shows pre-tax income, while the vertical axis shows disposable income. The original top tax bracket is shown by the solid line. As depicted, consider a tax reform which increases τ by $\Delta\tau$ above the income level z^*. To evaluate this change we need to consider the effects on revenue and social welfare. Ignoring behavioral responses at first, this reform mechanically raises additional revenue by an amount equal to the change in the tax rate ($\Delta\tau$) multiplied by the number of people to whom the higher rate applies (N^*) multiplied by the amount by which the average income of this group (z_m) is above the cut-off income level (z^*) so that the additional revenue is $\Delta\tau\, N^*[z_m - z^*]$. 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^* is the same for all z^* in the top tail and equal to $a/(a-1)$. For the U.S. economy, the cutoff for the top percentile of tax filers is approximately \$400,000, and the average income for this group is approximately \$1.2 million, so that $z_m/z^* = 3$ and hence $a = 1.5$.

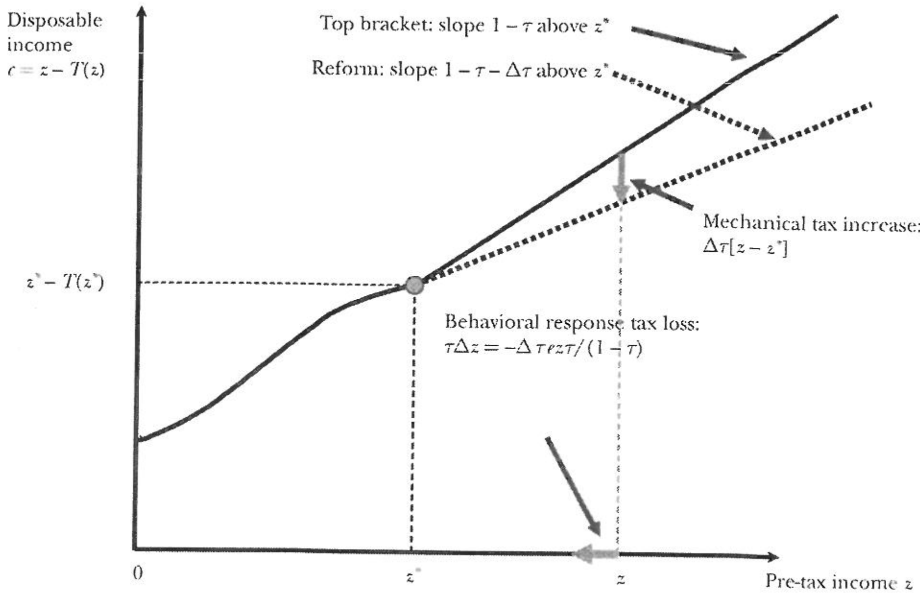
Raising the tax rate on the top percentile obviously reduces the utility of high-income tax filers. If we denote by g the social marginal value of \$1 of consumption for top income earners (measured relative to government revenue), the direct welfare cost is g multiplied by the change in tax revenue collected.[5] Because the government values redistribution, the social marginal value of consumption for top-bracket tax filers is small relative to that of the average person in the economy, and so g is small and as a first approximation can be ignored. A utilitarian social welfare criterion with marginal utility of consumption declining to zero, the most commonly

[3] This top marginal tax rate is much higher than the current average tax rate among top 1 percent earners mentioned above because of deductions and especially lower tax rates that apply to realized capital gains.

[4] The top tax rate τ is 42.5 percent for ordinary labor income when combining the top federal individual tax rate of 35 percent, uncapped Medicare taxes of 2.9 percent, and an average combined state top income tax rate of 5.86 percent and average sales tax rate of 2.32 percent. The average across states is computed using state weights equal to the fraction of filers with adjusted gross income above \$200,000 that reside in the state as of 2007 (IRS, 2009a). The 2.32 percent average sales tax rate is estimated as 40 percent of the average nominal sales tax rate across states (as the average sales tax base is about 40 percent of total personal consumption.) As the 1.45 percent employer Medicare tax is deductible for both federal and state income taxes, and state income taxes are deductible for federal income taxes, we have $((1 - .35) \times (1 - .0586) - .0145)/(1.0145 \times 1.0232) = .575$, and hence $\tau = 42.5$ percent.

[5] Formally, g is the weighted average of social marginal weights on top earners, with weights proportional to income in the top bracket.

Figure 1
Optimal Top Tax Rate Derivation



Source: The authors.
Notes: The figure depicts the derivation of the optimal top tax rate $\tau^* = 1/(1 + ae)$ by considering a small reform around the optimum which increases the top marginal tax rate τ by $\Delta\tau$ above z^*. A taxpayer with income z mechanically pays $\Delta\tau[z - z^*]$ extra taxes but, by definition of the elasticity e of earnings with respect to the net-of-tax rate $1 - \tau$, also reduces his income by $\Delta z = ez\Delta\tau/(1 - \tau)$ leading to a loss in tax revenue equal to $\Delta\tau e z \tau/(1 - \tau)$. 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 $\tau^* = 1/(1 + ae)$. This is the optimum tax rate when the government sets zero marginal welfare weights on top income earners.

used specification in optimal tax models, has this implication. For example, if the social value of utility is logarithmic in consumption, then social marginal welfare weights are inversely proportional to consumption. In that case, the social marginal utility at the \$1,364,000 average income of the top 1 percent in 2007 (Piketty and Saez, 2003) is only 3.9 percent of the social marginal utility of the median family, with income \$52,700 (U.S. Census Bureau, 2009).

Behavioral responses can be captured by the elasticity e of reported income with respect to the net-of-tax rate $1 - \tau$. By definition, e measures the percent increase in average reported income z_m when the net-of-tax rate increases by 1 percent.[6] At the optimum, the marginal gain from increasing tax revenue with no behavioral response and the marginal loss from the behavioral reaction must be equal to each

[6] 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.

other. Ignoring the social value of marginal consumption of top earners, the optimal top tax rate τ^* is given by the formula

$$\tau^* = 1/(1 + ae).$$

*

The optimal top tax rate τ^* is the tax rate that maximizes tax revenue from top bracket taxpayers.[7] Since the goal of the marginal rates on very high incomes is to get revenue in order to hold down taxes on lower earners, this equation does not depend on the total revenue needs of the government. Any top tax rate above τ^* would be (second-best) Pareto inefficient as reducing tax rates at the top would both increase tax revenue and the welfare of top earners.

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. τ^* is decreasing with the elasticity e (which affects the deadweight burden) and the Pareto parameter a, which measures the thinness of the top of the income distribution and so the ratio of those above a tax level to the income of those at the tax level.

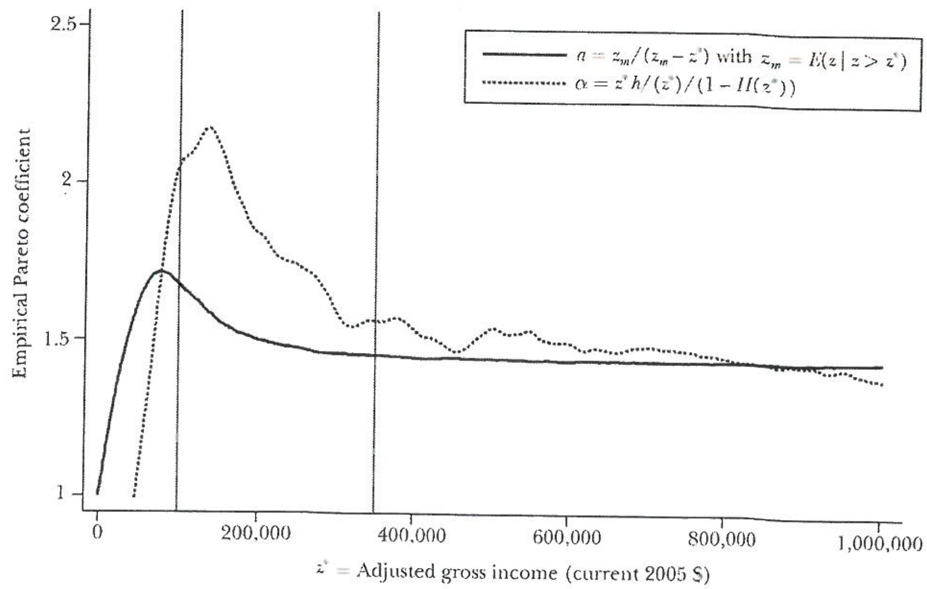
The solid line in Figure 2 depicts the empirical ratio $a = z_m/(z_m - z^*)$ with z^* ranging from \$0 to \$1,000,000 in annual income using U.S. tax return micro-data for 2005. We use "adjusted gross income" from tax returns as our income definition. The central finding is that a is extremely stable 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, as summarized in Atkinson, Piketty, and Saez (2011).

If we assume that the elasticity e is roughly constant across earners at the top of the distribution, the formula $\tau = 1/(1 + ae)$ 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). That is, the optimal marginal tax rate is approximately the same over the range of very high incomes where the distribution is Pareto and the marginal social weight on consumption is small.[8] This makes the optimal tax formula quite general and useful.

[7] If a positive social weight $g > 0$ is set on top earners' marginal consumption, then the optimal rate is $\tau = (1 - g)/(1 - g + ae) < \tau^*$. With plausible weights that are small relative to the weight on an average earner, the optimal tax does not change much.

[8] If the elasticity e does not vary by income level, then the Pareto parameter a does not vary with τ. If the elasticity varies by income, the Pareto parameter a might depend on the top tax rate τ. The formula $\tau^* = 1/(1 + ae)$ is still valid in that case, but determining τ^* would require knowing how a varies with τ.

Figure 2
Empirical Pareto Coefficients in the United States, 2005



Source: The authors using public use tax return data.
Notes: The figure depicts in solid line the ratio $a = z_m/(z_m - z^*)$ with z^* ranging from \$0 to \$1,000,000 annual income and z_m the average income above z^* using U.S. 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 $\tau^* = 1/(1 + ae)$. Denoting by $h(z)$ the density and by $H(z)$ the cumulative distribution function 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 $e(z)$ implies that the optimal marginal tax rate $T'(z) = [1 - G(z)]/[1 - G(z) + \alpha(z)e(z)]$ increases with z.

The Tax Elasticity of Top Incomes

The key remaining empirical ingredient to implement the formula for the optimal tax rate is the elasticity e of top incomes with respect to the net-of-tax rate. With the Pareto parameter $a = 1.5$ if $e = .25$, a mid-range estimate from the empirical literature, then $\tau^* = 1/(1 + 1.5 \times .25) = 73$ percent, substantially higher than the current 42.5 percent top U.S. marginal tax rate (combining all taxes).[9] *

[9] Using g^* of .04, the optimal tax rate decreases by about 1 percentage point.

The current rate, $\tau = 42.5$ percent, would be optimal only if the elasticity e were extremely high, equal to 0.9.[10]

Before turning to empirical estimates, we review some of the interpretation issues that arise when moving beyond the simplest version of the Mirrlees (1971) model. In the Mirrlees model, there is a single tax on each individual. With many taxes, for example, in many periods, the key measure is the response of the present discounted value of all taxes, not the response of revenue in a single year. This observation matters given significant control by some people over the timing of taxes and over the forms in which income might be received. Also, because the basic Mirrlees model has no tax-deductible charitable giving, a tax-induced change in taxable income involves only distortions from reduced earnings. However, when an increase in marginal tax rates leads to an increase in charitable giving, the gain to the recipients needs to be incorporated in the efficiency measure (Saez, 2004). Other tax deductions are more difficult to consider. In the Mirrlees model, compensation equals the marginal product. In bargaining settings or with asymmetric information, people may not receive their marginal products. Thus, effort is responding to a price that is higher or lower than marginal product, and the tax rate itself may affect the gap between compensation and marginal product.

The large literature using tax reforms to estimate the elasticity relevant for the optimal tax formula has focused primarily on the response of reported income, either "adjusted gross income" or "taxable income," to net-of-tax rates. Saez, Slemrod, and Giertz (forthcoming) offer a recent survey, while Slemrod (2000) looks at studies focusing on the rich. The behavioral elasticity is due to real economic responses such as labor supply, business creation, or savings decisions, but also tax avoidance and 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. 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, exercises of stock options surged in 1992 before the 1993 top rate increase took place (Goolsbee, 2000). The Tax Reform Act of 1986 also led 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). The paper Gruber and Saez (2002) is often cited for its substantial taxable income elasticity estimate ($e = 0.57$) at the top of the distribution. However, its authors 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 in a given year is quite sensitive to tax rates, so the elasticity e is large, and the optimal top tax rate is correspondingly low. Two important qualifications must be made.

[10] Alternatively, if the elasticity is $e = .25$, then $\tau = 42.5$ percent is optimal only if the marginal consumption of very high-income earners is highly valued, with $g = .72$.

First, as mentioned above, many of the tax avoidance channels such as retiming or income shifting produce changes in tax revenue in other periods or other tax bases—called "tax externalities"—and hence do not decrease the optimal tax rate. Saez, Slemrod, and Giertz (forthcoming) provide formulas showing how the optimal top tax rate should be modified in such cases. 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 enforcement (Slemrod and Kopczuk, 2002; Kopczuk, 2005). Thus, the distinction between real responses and tax avoidance responses is critical for tax policy. As an illustration using the different 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 $\tau^* = 1/(1 + 1.5 \times 0.57) = 54$ percent, while the optimal top tax rate using a broader income base with no deductions would be $\tau^* = 1/(1 + 1.5 \times 0.17) = 80$ percent. Taking as fixed state and payroll tax rates, such rates correspond to top federal income tax rates equal to 48 and 76 percent, respectively. Although considerable uncertainty remains in the estimation of the long-run behavioral responses to top tax rates (Saez, Slemrod, and Giertz, forthcoming), the elasticity $e = 0.57$ is a conservative upper bound estimate of the distortion of top U.S. tax rates. Therefore, the case for higher rates at the top appears robust in the context of this model.

Additional Considerations

To some readers, proposing marginal income tax rates on the top percentile of earners, along with a broadened tax base, in a range from 48 to 76 percent may seem implausibly high. One way to judge how seriously to take such numbers is to consider whether elements left out in the derivation push for a significantly different answer. Two key omitted elements are the presence of capital income and a longer-run dynamic perspective.

Does the presence of capital income mean that earnings should be taxed significantly differently? When we discuss taxation of capital income in a later section, we note that the ability to convert some labor income into capital income is a reason for limiting the difference between tax rates on the two types of income—that is, an argument for taxing capital income. Plausibly, it is also an argument for a somewhat lower labor income tax, assuming that labor income should be taxed more heavily than capital income.

Perhaps most critically, does an estimate based on a single period model still apply when recognizing that people earn and pay income taxes year after year? First, earlier decisions such as education and career choices affect later earnings opportunities. It is conceivable that a more progressive tax system could reduce incentives to accumulate human capital in the first place. The logic of the equity–efficiency tradeoff would still carry through, but the elasticity e should reflect not only short-run labor supply responses but also long-run responses through education and career choices. While there is a sizable multiperiod optimal tax literature using life-cycle models and generating insights, we unfortunately have little compelling empirical evidence to assess whether taxes affect earnings through those long-run channels.

Second, there is significant uncertainty in future earnings. Such uncertainty gives an insurance role for earnings taxation and, as we shall see, also has consequences for the taxation of savings.[14] However, the applicability of results for policy seems unclear to us.