Taxing Top Incomes in a World of Ideas

Charles I. Jones*

Stanford GSB and NBER

March 26, 2019 — Version 1.0

Abstract

This paper considers the taxation of top incomes when the following conditions apply: (i) new ideas drive economic growth, (ii) the reward for creating a successful innovation is a top income, and (iii) innovation cannot be perfectly targeted by a separate research subsidy — think about the business methods of Walmart, the creation of Uber, or the "idea" of Amazon.com. These conditions lead to a new force affecting the optimal top tax rate: by slowing the creation of the new ideas that drive aggregate GDP, top income taxation reduces everyone's income, not just the income at the top. When the creation of ideas is the ultimate source of economic growth, this force sharply constrains both revenue-maximizing and welfare-maximizing top tax rates. For example, for extreme parameter values, maximizing the welfare of the middle class requires a *negative* top tax rate: the higher income that results from the subsidy to innovation more than makes up for the lost redistribution. More generally, the calibrated model suggests that incorporating ideas as a driver of economic growth cuts the optimal top marginal tax rate substantially relative to the basic Saez calculation.

*John Cochrane and I initiated this project together, but because of other pressing commitments, he had to withdraw. I am extremely grateful for his many insights. I also thank Ufuk Akcigit, Sebastian Di Tella, Ben Hebert, Pete Klenow, Ben Lockwood, Casey Mulligan, Emmanuel Saez, Stefanie Stantcheva, Chris Tonetti, Gabriel Zucman, and seminar participants at Berkeley, BU, Chicago, Columbia, Stanford, and Vanderbilt for helpful suggestions. Andrés Yany provided expert research assistance.

1. Introduction

This paper considers the taxation of top incomes when the following conditions apply: (i) new ideas drive economic growth, (ii) the reward for creating a successful new idea is a top income, and (iii) innovation is broad-based and cannot be perfectly targeted by a separate research subsidy.

The classic tradeoff in the optimal taxation literature is between redistribution and the incentive effects that determine the "size of the pie." But in most of that literature — starting with Mirrlees (1971), Diamond (1998), Saez (2001) and Diamond and Saez (2011) — the "size of the pie" effects are relatively limited. In particular, when a top earner reduces her effort because of a tax, that reduces her income but may have no or only modest effects on the incomes of everyone else in the economy.

In constrast, I embed the optimal tax literature in the idea-based growth theory of Romer (1990), Aghion and Howitt (1992), and Grossman and Helpman (1991). According to this literature, the enormous increase in living standards over the last century is the result of the discovery of new ideas — perhaps by a relatively small number of people. To the extent that top income taxation distorts this innovation, it can impact not only the income of the innovator but also the incomes of everyone else in the economy.

The nonrivalry of ideas is key to this result and illustrates why incorporating physical capital or human capital into the top tax calculation is insufficient. If you add one unit of human capital or one unit of physical capital to an economy — think of adding a computer or an extra year of education for one person — you make one worker more productive, because these goods are rival. But if you add a new idea — think of the computer code for the original spreadsheet or the blueprint for the electric generator — you can make any number of workers more productive. Because ideas are nonrival, each person's wage is an increasing function of the entire stock of ideas. A distortion that reduces the production of new ideas therefore impacts everyone's income, not just the income of the inventor herself.

A standard policy implication in this literature is that it may be optimal to subsidize formal R&D, and one could imagine subsidizing research but taxing top incomes as a way to simultaneously achieve both efficient research and socially-desirable redistribution. Instead, we consider a world with both basic and applied research. Basic research uncovers fundamental truths about the way the world works and is readily subsidized with government funding. Applied research then turns these fundamental truths into consumer products or firm-level process innovations. This is the task of entrepreneurs and may not be readily subsidized as formal R&D. Think about the creation of Walmart or Amazon.com, organizational innovations in the health care and education industries, the latest software underlying the Google search engine, or even the creation of nonrival goods like a best-selling novel or the most recent hit song. Formal R&D is a small part of what economists would like to measure as efforts to innovate. For example, around 70% of measured R&D occurs in the manufacturing industry. In 2012, only 18 million workers (out of US employment that exceeds 130 million) were employed by firms that conducted any official R&D.[1] According to their 2018 corporate filings, Walmart and Goldman-Sachs report doing zero R&D.

The idea creation and implementation that occurs beyond formal R&D may be distorted by the tax system. In particular, high incomes are the prize that motivates entrepreneurs to turn a basic research insight that results from formal R&D into a product or process that ultimately benefits consumers. High marginal tax rates reduce this effort and therefore reduce innovation and the incomes of everyone in the economy.

Taking this force into account is important quantitatively. For example, consider raising the top marginal tax rate from 50% to 75%. As we discuss below, in the United States, the share of income that this top marginal rate applies to is around 10%, so the change raises about 2.5% of GDP in revenue before the behavioral response. In the baseline calibration, such an increase in the top marginal rate reduces innovation and lowers GDP per person in the long run by around 6 percent. With a utilitarian welfare criterion, this obviously reduces welfare. But even redistributing the 2.5% of GDP to the bottom half of the population would leave them worse off on average: the 6% decline in their incomes is not offset by the 5% increase from redistribution. In other words, unless the social welfare function puts disproportionate weight on the poorest people in society, raising the top marginal rate from 50% to 75% reduces social welfare.

We consider various revenue- and welfare-maximizing top tax rate calculations, first ignoring the effect on innovation and then taking it into account. For a broad range of parameter values, the effects are large. For example, in a baseline calculation, the

[1]These numbers are from Wolfe (2014).

revenue-maximizing top tax rate that ignores the innovation spillover is 92%. In contrast, the rate that incorporates innovation and maximizes a utilitarian social welfare function is just 22%. Moreover, if ideas play an even more significant role than assumed in this baseline, it is possible for the optimal top income tax rate to turn negative: the increase in everyone's income associated with subsidizing innovation exceeds the gains associated with redistribution.

Importantly, however, the point of this paper is not to estimate "the" optimal top income tax rate. Such a calculation involves many additional considerations documented in the existing literature (reviewed below) that are omitted from the analysis here. Instead, the point is that future work aimed at calculating such a number will certainly want to explicitly consider the effects of top income taxation on the creation of new ideas. They appear to be quantitatively important.

The remainder of the paper is organized as follows. After a brief literature review, Section 2 lays out the steady-state of a rich dynamic growth model and considers the top tax rate that maximizes revenue, along the lines of Diamond and Saez (2011) and others. Section 3 then considers the tax scheme that maximizes the welfare of the "bottom 90%" or the median voter (they are the same people here). This turns out to matter quantitatively: the planner cares about distorting the creation of ideas not merely because it affects the revenue that can be earned by regular workers, but because it affects their consumption and economic growth directly. This setup is especially convenient for two reasons. First, it yields a nice closed-form solution. Second, it allows us to remain agnostic about the *source* of the behavioral tax elasticity for top earners: whether this comes from an effort choice or an occupational choice or from something else is irrelevant; we just need to know the elasticity.

Section 4 goes further and finds the tax system that maximizes a utilitarian social welfare function. While this objective function is obviously of interest, the solution does not admit a closed-form expression. In addition, we must be explicit about the nature of the behavioral tax elasticity for top earners, which makes the model less general. Section 5 discusses additional results and extensions, including empirical evidence on growth and top income taxation. Finally, Section 6 builds the full dynamic growth model that nests the simple model as a special case in the steady state.

1.1 Literature Review

Partly motivated by the rise in top income inequality documented by Piketty and Saez (2003), there has been an explosion of work on top income taxation in recent years. Badel and Huggett (2015), Rothschild and Scheuer (2015), Sachs, Tsyvinski and Werquin (2016), and Lockwood, Nathanson and Weyl (2017) allow for imperfect substitution in production, general exernalities, or human capital spillovers so that the decisions of top earners can affect the wages of others in the economy. Ales and Sleet (2016) consider a different kind of spillover: the taxation of CEOs in an assignment model like Gabaix and Landier (2008) or Tervio (2008) in which CEO effort affects the profits of the firm. Badel and Huggett (2017) considers the revenue-maximizing top tax rate in a general setting when incomes at lower brackets can be affected by the top tax rate and when top earners possibly have other sources of income taxed at different rates that are themselves affected by the top rate. In all these cases, the additional forces typically lower the optimal top tax rate relative to the Diamond and Saez (2011) numbers.

Among these, Lockwood, Nathanson and Weyl (2017) is closest to the present paper. They consider optimal nonlinear taxation in a full Mirrleesian setup when different people can choose different careers. They suggest that some careers, such as research or engineering, generate positive externalities in the economy while others, such as finance and law, may generate negative externalities. The fully-optimal tax system would assign different tax rates to these different career options, but if forced to pick a single top tax rate, the planner has to average across these externalities in some way. I instead focus explicitly on the research dimension, where leveraging the extensive endogenous growth literature allows us to make more precise statements about this particular externality.

Piketty, Saez and Stantcheva (2014) emphasize that to the extent that some of the standard behavioral elasticity of top incomes with respect to taxes is associated with rent seeking, the relevant elasticity for optimal top income taxation is even lower. Akcigit, Baslandze and Stantcheva (2016b) explore the international mobility of superstar inventors in response to top income tax rate differences and find a large elasticity, particularly for foreign inventors. Moretti and Wilson (2017) show that the migration patterns of star scientists across U.S. states are highly elastic with respect to tax rates. Kindermann and Krueger (2014) consider a rich overlapping generations model with idiosyncratic risk and show that top tax rates as high as 90% may be optimal. Scheuer and Werning (2017) show how the Saez (2001) formula generalizes in the presence of "superstar" effects. In particular, they show that the same basic formula applies, but only when one recognizes that the income distribution is itself endogenous to the tax schedule.

Akcigit, Grigsby, Nicholas and Stantcheva (2018) provide extensive empirical evidence that innovation responds to tax incentives. They use a combination of patent data, state-level corporate and personal income tax rates, and changes in federal tax rates in both "macro" and "micro" level research designs in the U.S. in the 20th century. For example, they find that the elasticity of patents, the number of inventors, and patent citations with respect to either personal income taxes or corporate tax rates (in particular, with respect to the "keep rate" $1-\tau$) are all larger than 2 in magnitude: increasing the keep rate by 10% raises innovation by more than 20 percent(!). These macro elasticities include movements of innovation from one state to the other, and therefore are an upper bound on the net effect most relevant in this paper. But they also report various estimates of the net effect that are at least 0.6 in magnitude — increasing the keep rate by 10% raises innovation by at least 6%.

Another closely related paper is Jaimovich and Rebelo (2017), who study the growth consequences of taxation. They focus on a puzzle in the endogenous growth literature: if long-run growth rates are sensitive to tax policy, then why is it that we see so little evidence of this in time series and cross-sectional evidence? They use a version of the Romer (1990) growth model in which researchers have heterogeneous talent to study the effect of a (linear) tax on firm profits and find that the effect of taxes is nonlinear. When tax rates are low, tax changes have relatively muted effects because the marginal researchers have relatively low talent. But as tax rates get high, they start to distort the research effort of highly-talented researchers (they assume a Pareto talent distribution), and this can have large effects. They show that a social planner in their model would choose a tax rate on profits of around 30-35%, close to the observed corporate profits tax rate, and in the range where the effect of taxes on growth is small.

This paper studies optimal taxation in a growth setting and differs in several ways. First, the focus is explicitly on top income taxation. We embed the growth framework in the setting of the tax literature and characterize the various forces that influence the optimal top tax rate. Second, considering nonlinear taxes rather than a linear tax on profits is important. For example, in Jaimovich and Rebelo (2017), the optimal nonlinear schedule might feature a subsidy for research incomes in the intermediate range followed by a high tax rate on top incomes. The extra bracket in the tax schedule would allow the planner to subsidize research so that the right number of people work in R&D while also using the high rate to transfer income away from top earners. We incorporate an explicit subsidy to formal R&D in the framework and consider the additional role played by nonlinear income taxation. Finally, the growth model used by Jaimovich and Rebelo (2017) features strong growth effects: a change in tax policy can permanently alter the growth rate of the economy. A wealth of evidence suggests that this framework may overstate the effect of policy changes; for example, see Bloom, Jones, Van Reenen and Webb (2019). I instead explore a model in which policies have long-run level effects, and the size of this level effect is a key parameter influencing the top tax rate; Jaimovich and Rebelo (2017) essentially assume this parameter is infinity while we consider values of between 1/8 and 1.

On a related point, Bell, Chetty, Jaravel, Petkova and Van Reenen (2017) combine tax records from the IRS and patenting data for individual inventors to suggest that the elasticity of occupational choice for inventors with respect to the top tax rate is small. Their argument is that the distribution of earnings outcomes for inventors is highly skewed, so the decision to become an inventor is like buying a lottery ticket. Because of risk aversion, taxes on top incomes have a small effect on occupational choice in their setting; top tax rates hit only when the marginal utility of consumption is very low. Bell, Chetty, Jaravel, Petkova and Van Reenen (2019) emphasize that exposure to role models who are innovators is empirically an important determinant of who becomes an innovator. It is partly for this reason and because of the points made above related to Jaimovich and Rebelo (2017) that this paper completely abstracts from effects on the extensive margin of occupational choice. Instead, as these papers acknowledge, the conventional effects of taxes on effort choice — the focus here — may still be important. (And that is perhaps one way to reconcile these studies with the evidence cited earlier in this literature review of relative large effects of taxes on inventive effort, patents, and citations.)

Hall and Woodward (2010) provide a different perspective. Using an extensive data set on venture capital funding from 1987 until 2008, they show that the returns to entrepreneurs are extremely skewed: nearly 3/4ths of entrepreneurs receive nothing at exit, while a few receive more than a billion dollars. An entrepreneur with a coefficient of relative risk aversion of two values this lottery with a certainty equivalent of only slightly more than zero. An implication is that the tax rate that applies to the successful outcome can have a substantial influence on entrepreneurial activity.

Finally, another important consideration is other policies that may be more effective at stimulating innovation. For example, perhaps providing a strong social safety net and lenient bankruptcy laws will provide insurance on the downside that encourages entrepreneurs to take risks. Or perhaps high top tax rates encourage talented people to become entrepreneurs instead of hedge fund managers, actually stimulating innovation. Clearly any calculation attempting to determine *the* optimal top tax rate will need to take all these forces into account. This paper takes a first step at showing how the theory of optimal top income taxation can be combined with idea-based growth theory, focusing on a relatively conventional setting and exploring a range of parameter values consistent with the existing literature.

2. The Revenue-Maximizing Tax Rate in a Simple Model

The simple model in this section provides the key intuition for our main result. Later, in Section 6, we show that the simple model is a special case of the balanced growth path of a richer dynamic model. The model and notation are similar to Jones (2005).

The economy contains an exogenous number of entrepreneurs who do applied research, S_a, an exogenous number of "basic" researchers, S_b, and an exogenous number of workers, L. This setup therefore abstracts from occupational choice, for the reasons discussed above. Entrepreneurs get to choose their effort, e; workers and basic researchers each supply one unit of labor inelastically.

There is a long history in the literature on R&D distinguishing between "invention" and "innovation" or between "basic" and "applied" research. Basic science discovers fundamental truths or inventions, and then applied research turns this into goods that benefit firms or consumers. Applied research is essential in that you cannot eat chemical formulas; basic research is essential in that applied research is based on the fundamental discoveries. In the context of optimal income taxation, the funding side of this distinction is also relevant. Much of government funding for research is focused on basic research, while we think of top income taxation as being relevant for the applied innovations that create consumer goods or process innovations.[2]

The production function for consumption uses applied ideas, A, and two types of labor, managers $\tilde{M}$ and workers L:

$$Y = A^{\gamma} \tilde{M}^{\psi} L^{1-\psi}. \tag{1}$$

The purpose of these different types of labor will be explained more later.

Next, we need to specify a production function for ideas. Section 6 develops a full dynamic growth model in which ideas are a stock and researchers produce a flow of new ideas every period. In the long run, the stock of ideas is proportional to the flow, which in turn depends on the number of researchers. For the majority of this paper, we focus on the balanced growth path of the full model; this makes the intuition and the analysis as clear as possible. The tradeoff is that the analysis ignores transition dynamics, but given how little we know about the transition paths in these models, that tradeoff is worth making. Along the balanced growth path, the stock of applied ideas is

$$A = \nu_a \mathbb{E}[ez] S_a B^{\beta}. \tag{2}$$

The long-run stock of applied ideas is proportional to the number of entrepreneurs, S_a. The factor of proportionality, in turn, depends on effort, e, talent, z, and the stock of basic research ideas, B. The expectation is needed because entrepreneurs are heterogeneous in their talent, and the effort choice may depend on this talent. (In equilibrium, it turns out that e is independent of z, so we can eventually write this as just $\mathbb{E}[ez] = \bar{e}\bar{z}$.) Also, the larger the stock of basic research ideas, B, the higher the number of applied ideas in the long run; that is, we assume $\beta > 0$.

Finally, basic researchers produce basic research ideas, where once again stocks and flows are proportional:

$$B = \nu_b S_b. \tag{3}$$

[2]For a recent application of this distinction in the growth literature, see Akcigit, Hanley and Serrano-Velarde (2016a).

To keep things simple for the moment, we assume the stock of basic knowledge, B, is directly proportional to the number of basic researchers.

Combining these equations, final output — and consumption since there is no capital — in steady state is

$$Y = \left(\nu \mathbb{E}[ez] S_a S_b^{\beta}\right)^{\gamma} \tilde{M}^{\psi} L^{1-\psi} \tag{4}$$

where $\nu \equiv \nu_a \nu_b^{\beta}$. In the full dynamic model of Section 6, we consider a much richer structure for the idea production functions, including dynamic feedback from past basic and applied ideas into the production of new ideas as well as diminishing returns to research effort at a point in time. The key equation (4) still turns out to be the solution in this richer structure; for example, see equation (46) in Section 6.

In the dynamic version of this model, all the population variables — S_a, S_b, $\tilde{M}$, and L — grow at a constant population growth rate. The overall degree of increasing returns to scale in the economy is then given by $\gamma(1+\beta)$, and this combination of parameters plays a crucial role in the model.

First and foremost, this increasing returns results from the nonrivalry of ideas, emphasized by Romer (1990). Because this force plays such a crucial role in what follows, it deserves some elaboration. The parameter γ results from the nonrivalry of applied ideas in equation (1). The parameter β results from the nonrivalry of basic ideas as an input in applied research.

How does the nonrivalry of ideas explain economic growth? According to Romer (1990), the key is that nonrivalry gives rise to increasing returns to scale. The *standard replication argument* is a fundamental justification for constant returns to scale in production. If we wish to double the production of computers from a factory, one feasible way to do it is to build an equivalent factory across the street and populate it with equivalent workers, materials, and so on. That is, we replicate the factory exactly. This means that production with rival goods is, at least as a useful benchmark, a constant returns process.

What Romer stressed is that the nonrivalry of ideas is an integral part of this replication argument: firms do not need to reinvent the idea for a computer each time a new computer factory is built. Instead, the same idea — the detailed set of instructions for how to make a computer — can be used in the new factory, or indeed in any number of factories, because it is nonrival. Since there are constant returns to scale in the rival inputs (the factory, workers, and materials), there are therefore *increasing returns* to the rival inputs and ideas taken together: if you double the rival inputs and the quality or quantity of the ideas, you will more than double total production. These insights are embedded in the production function in equation (1): there is constant returns in the rival inputs (here, just labor) and increasing returns to ideas and the rival inputs taken together. The parameter γ measures the overall degree of increasing returns associated with the nonrivalry of applied ideas. A similar argument explains the parameter β and the nonrivalry of basic ideas in applied research.

This discussion hints at why distortions to the creation of ideas can be so costly. If you add one unit of human capital or one unit of physical capital to an economy — think of adding a computer or an extra year of education for one person — you make one worker more productive. But if you add a new idea — think of the computer code for the first spreadsheet, or the design of the iphone, or the blueprint for the latest electric generator — you can make any number of workers more productive. GDP per person is proportional to the *total* stock of ideas, not to ideas per person. A distortion that reduces the production of new ideas therefore impacts everyone's income.

In principle, there are several allocative decisions that need to be made, even in this simple economy. The overall population needs to be divided into entrepreneurs, basic researchers, and workers, and the entrepreneurs need to decide how much to work. In what follows, we take the first allocation as given — for example, we assume that a formal research subsidy puts the right number of people into basic research and incentivizes them to provide the right amount of effort. Instead, we focus on the extent to which the behavior of entrepreneurs is distorted by the top marginal tax rate. Hence, the only allocative decision in this simple economy is the choice of e: how hard do entrepreneurs work to create new ideas and make everyone in the economy richer? Importantly, note that the optimal allocation of people between basic research and labor is invariant to the entrepreneur's effort choice. For example, if $L + S_b = N$ (as in the full model in Section 6), let $b \equiv S_b/N$ be the fraction of people who work in basic research. Then total output can be written as $Y = Const \cdot b^{\beta\gamma}(1 - b)$ and the output-maximizing allocation is $\frac{b}{1-b} = \beta\gamma$, independent of the effort choice.

2.1 Diamond, Saez, and the Revenue-Maximizing Tax Rate

Following the insights of Diamond (1998), Saez (2001) and Diamond and Saez (2011) etc., one can make good progress on this question by keeping the model of effort choice as general as possible. This is particularly true if we take the tax schedule as given and focus on finding the top tax rate that maximizes revenue.

Managers are included in the model to capture the traditional Diamond and Saez (2011) tradeoff. In particular, we assume $\tilde{M} := \mathbb{E}(ez)M$, where M is the exogenous supply of managers, e denotes their effort, and z denotes their (heterogeneous) talent. We abuse notation slightly in using the same notation for the talent and effort of managers that we used for entrepreneurs. Even ignoring ideas completely by setting $\gamma = 0$, the model will feature a "traditional" role for top income taxation working through managers.

We assume a two-part tax schedule: all income below $\bar{w}$ is taxed at rate τ_0, while all income above is taxed at τ. Let w be the wage of unskilled work, w_b the wage of basic research, w_s the wage per unit of effective effort (ez) of entrepreneurs, and w_m the wage per unit of effective effort of managers. We assume basic researchers are not taxed at the top rate and that the supply of managers is sufficiently small that their income is taxed at the top rate.

Total tax revenue in the economy is then

$$T = \underbrace{\tau_0(wL + wS_b + w_a\mathbb{E}(ez)S_a + w_m\mathbb{E}(ez)M)}_{\text{all income pays } \tau_0} + \underbrace{(\tau - \tau_0)[(w_a\mathbb{E}(ez) - \bar{w})S_a + (w_m\mathbb{E}(ez) - \bar{w})M]}_{\text{income above } \bar{w} \text{ pays an additional } \tau - \tau_0}$$

As we show later in Section 6, the share of output paid to entrepreneurs is a constant, ρ_s, that is invariant to the top tax rate, as is the share of output paid to managers. That is

$$\frac{w_a\mathbb{E}(ez)S_a}{Y} = \rho_s \quad \text{and} \quad \frac{w_m\mathbb{E}(ez)M}{Y} = \rho_m. \tag{5}$$

Note that in general, ρ_m could differ from ψ because ideas must be paid for in some way as well, and this means that factors are not typically paid their marginal products. Defining $\rho \equiv \rho_s + \rho_m$, the remainder of GDP, $1 - \rho = \frac{wL + w_bS_b}{Y}$, is paid to workers and basic researchers.

With these definitions, total tax revenue can be rewritten as

$$T = \tau_0 Y(\tau) + (\tau - \tau_0)[\rho Y(\tau) - \bar{w}(S_a + M)]. \tag{6}$$

That is, all income gets taxed at the base rate τ_0, while the income above $\bar{w}$, due to entrepreneurs and managers, generates extra revenue to the extent that the top tax rate exceeds τ_0.

The top tax rate that maximizes tax revenue can be found by setting the derivative of equation (6) equal to zero. This first-order condition can be written as

$$\underbrace{(\rho - \bar{\rho})\, Y}_{\text{mechanical gain}} + \underbrace{\frac{dY}{d\tau} \cdot [(1-\rho)\tau_0 + \rho\tau]}_{\text{behavioral loss}} = 0 \tag{7}$$

where $\bar{\rho} \equiv \frac{\bar{w}(S_a+M)}{Y}$. The first term of this equation is the mechanical revenue gain that comes from raising the top tax rate, holding Y constant. The second term is the loss in revenue that comes from changes in economic behavior leading to a reduction in Y. Maximizing revenue sets the sum of these effects to zero.

This can be rearranged to give an expression for the revenue-maximizing top tax rate τ^*. Letting $\eta_{Y,1-\tau} \equiv \frac{d \log Y}{d \log(1-\tau)}$ and $\Delta\rho \equiv \rho - \bar{\rho}$, we have[3]

$$\tau^*_{rm} = \frac{1 - \tau_0 \cdot \frac{1-\rho}{\Delta\rho} \cdot \eta_{Y,1-\tau}}{1 + \frac{\rho}{\Delta\rho}\, \eta_{Y,1-\tau}}, \tag{8}$$

where the subscript "rm" stands for "revenue maximization."

This equation can be compared to the basic result in Saez (2001) and Diamond and Saez (2011), which is $\tau_{ds} = \frac{1}{1+\alpha\zeta}$. The result here differs in two key ways. First, the term involving τ_0 is absent from the numerator in the Diamond-Saez formula: the reason is that these papers do not consider any interaction effects between the efforts of top earners and the wages earned by workers outside the top. Such a term has been introduced by follow-up research by Badel and Huggett (2015), Rothschild and Scheuer (2015), and Lockwood, Nathanson and Weyl (2017), among others. Those papers allow for general exernalities or human capital "spillovers" while Badel and Huggett (2017) derives such a formula in a very general setting. The approach here differs in that we

[3] A more complete derivation is available in the appendix.

focus explicitly on the effects that arise in idea-based growth models. This could be viewed as a special case of some of the more general formulations. But the value is that this is a special case that we understand well because of the large amount of research on idea-based growth models. We therefore obtain stronger intuitions and quantitative estimates as a result.

The second difference relative to the Diamond-Saez formula is in the nature of the elasticity, $\eta_{Y,1-\tau}$. In Diamond and Saez, the fundamental elasticity (called ζ above) is that of average top income with respect to the take-home rate. Here, instead, it is economy-wide income, again reflecting the fact that top taxes can affect the economy more broadly. The entirety of the growth model is embedded in $\eta_{Y,1-\tau}$, and this elasticity will be discussed further in a moment.

It is helpful to keep in mind the following intuition for the various terms that appear in the solutions. The intuition can typically be found by looking back at equation (7). For example, $\Delta\rho \equiv \rho - \bar{\rho}$ is the tax base to which the top marginal rate applies, capturing the mechanical effect. Similarly, $1 - \Delta\rho$ is the tax based to which the τ_0 rate applies. Next, $\rho\eta_{Y,1-\tau}$ is the effect of top taxes on the revenue of top earners via the behavioral response. And $\tau_0(1-\rho)\eta_{Y,1-\tau}$ is the effect on bottom tax revenue working through the behavioral response of top earners.

The term $\rho/\Delta\rho$ in the denominator parallels the Pareto distribution exponent in the Diamond-Saez formula (which they call "α"). It is the ratio of the total income of top earners divided by the share of that income that is taxed at the top rate. The smaller is this ratio, the higher is the income above the top rate, and the higher is the revenue-maximizing top tax rate.

To go further with this equation, of course, we now need to characterize $\eta_{Y,1-\tau}$, which is itself an equilibrium object. From the production function in (4) — recognizing that $\tilde{M} := \mathbb{E}(ez)M$ — we have

$$\eta_{Y,1-\tau} = (\gamma + \psi)\zeta \tag{9}$$

where ζ is defined to be the elasticity of $\mathbb{E}[ez]$ with respect to $1-\tau$. The elasticity ζ is the key elasticity that enters the Diamond and Saez formula: what is the elasticity of a top earner's income with respect to the take-home rate? For now, we will just leave this as a parameter and calibrate it the way Diamond and Saez do — in particular, we will consider a low value of $\zeta \approx 0.2$ and a higher value of $\zeta \approx 0.5$.

The parameter ψ captures the importance of "managers" — more accurately, top earners who are not idea creators — in the economy. We calibrate it to a small but seemingly reasonable value of 0.15. As we explain below, this parameter plays a minor role in the results that follow.

The other key parameter, γ, captures the increasing returns associated with applied ideas: when an entrepreneur creates and implements a new idea, it raises *everyone's* income. The more important are ideas, the more important this "spillover" will be.

There is no consensus value of γ in the growth literature. As shown in the full model in Section 6 — see especially equation (48) — along a balanced growth path, the growth rate of income per person satisfies:

$$g_y = \gamma(1+\beta)g_S. \tag{10}$$

That is, the long-run growth rate of income per person is proportional to the growth rate of researchers. The factor of proportionality is $\gamma(1+\beta)$, the overall degree of increasing returns to scale. One might feed in 1% to 2% for g_y. Values for g_S range from a low of around 1% (corresponding to population growth) to a high of around 4% (if one uses the NSF/OECD definition of researchers). Taking ratios, this suggests that $\gamma(1+\beta)$ lies somewhere between 1/4 and 2.

In the simple model, the top marginal tax rate only affects the effort of applied researchers, so we need to separate γ from $1+\beta$. It is far from obvious how to do this. Moreover, one could easily imagine cases in which the top marginal tax rate affects the effort of basic research or that basic research is imperfectly subsidized by a formal R&D tax credit. Given that we already have a wide range of uncertainty surrounding the value of $\gamma(1+\beta)$, we consider a set of parameter values for γ between 1/8 and 1. Since what matters for the top tax rate is the sum $\gamma+\psi$, the fact that we consider a large range of values for γ is what makes the precise value of ψ relatively unimportant.

The one parameter not yet discussed is $\Delta\rho \equiv \rho - \bar{\rho}$. This parameter equals the amount of income taxed at the top rate as a share of the economy's GDP. When top incomes obey a Pareto distribution with tail parameter $1/\xi$, it is straightforward to show that

$$\Delta\rho = \xi \times \begin{array}{l}\text{Income share of earners taxed} \\ \text{at the top marginal rate } (\rho)\end{array}$$

In the U.S. economy, Pareto inequality is approximately $\xi = 2/3$ and according to the Internal Revenue Service (2017, p. 30), the share of taxable income from returns with a marginal tax rate at the top is just under 20% in 2015. This suggests a first estimate of $\Delta\rho \approx .67 \times 20\% \approx 13\%$. However, in 2015, total taxable income of $7.4tr was only 41% of the $18.2tr GDP. Multiplying by this 41% provides a second estimate of $\Delta\rho \approx .41 \times 13\% \approx 5.3\%$. We use an intermediate value of $\Delta\rho = 10\%$ in the calculations since it is not obvious which number to prefer and since this value makes the intuition for the calculations easier to appreciate. We similarly choose $\rho = 15\%$, which makes $\rho/\Delta\rho = 1.5$, which is the value that Diamond and Saez typically use for the Pareto parameter that enters the denominator of their revenue-maximizing top tax rate. Alternatively, the expression above implies that $\rho/\Delta\rho = 1/\xi$. This is another way of saying that this term is precisely the Pareto parameter that enters the standard Saez (2001) expression. Finally, given that we choose $\rho = 0.15$, it is helpful to choose $\psi = 0.15$ as well: when we show results with $\gamma = 0$ to shut down the role of ideas, this means that managers are paid a share of GDP commensurate with their economic importance.

Table 1 shows the revenue-maximizing top tax rate for a range of cases and parameter values. The first line is from the simple Diamond and Saez formula $1/(1+\alpha\zeta)$ where $\alpha = 1.5 = \rho/\Delta\rho$. The remaining lines are from equation (8).

Several findings stand out. First, the baseline parameter values deliver high tax rates of 0.80 and 0.67 with the Diamond and Saez formula. Interestingly, the second and third lines of the table show that equation (8) implies even higher top rates when $\gamma = 0$. The reason for this can be seen in the fact that $\eta_{Y,1-\tau} = (\gamma + \psi)\zeta$, whereas the Diamond and Saez formula only involves ζ. The difference is that $\gamma + \psi$ captures the general equilibrium effect that taxes have on wages, whereas the Diamond-Saez formula can be interpreted as ignoring this GE effect. In our Cobb-Douglas setting, managers are paid the fraction ψ of output, regardless of how much they work: lower effort the leads to an offsetting increase in the wage w_m. Therefore, an increase in the top tax rate that reduces work effort increases the top wage, somewhat mitigating the change in effort: $\psi\zeta < \zeta$ since $\psi < 1$, which tends to raise the top tax rate. For example, the second row of the table finds a revenue-maximizing top tax rate of 96.4% instead of 80%.

The third row of the table preserves the basic GE forces, but keeps the idea channel

Table 1: Revenue-Maximizing Top Tax Rates

	Behavioral Elasticity	
Case	0.20	0.50
Diamond-Saez formula:	0.800	0.667
$\gamma = 0$ and $\tau_0 = 0$:	0.964	0.930
$\gamma = 0$ and $\tau_0 = 0.20$:	0.923	0.851
Degree of IRS, γ		
1/8	0.863	0.742
1/4	0.806	0.644
1/2	0.702	0.477
1	0.524	0.221

Note: The table reports the revenue-maximizing top tax rate for various cases. The first line is from the simple Diamond and Saez formula $1/(1+\alpha\zeta)$ where $\alpha = 1.5$. The remaining lines are from equation (8). The parameter ζ is calculated from the behavioral elasticity $\frac{\zeta}{1-\zeta}$ which implies $\zeta = 1/6$ and $\zeta = 1/3$ for the 0.2 and 0.5 cases, respectively (see Section 4 for more details). Other baseline parameter values are $\Delta\rho = 0.1$, $\rho = \psi = 0.15$, and $\tau_0 = 0.20$.

turned off by setting $\gamma = 0$. With $\tau_0 = 0.2$, there is still an $\eta_{Y,1-\tau}$ force in the numerator of the revenue-maximizing top tax rate formula: when managers work harder, that raises the wages of workers, a la Badel and Huggett (2017). But these effects are relatively small in this calibration because the share of managers in the economy, ψ, is small, and the revenue-maximizing top tax rate is 92.3% for $\gamma = 1/2$.

The remainder of the table shows how these calculations change when we turn on the idea channel. The more important are ideas, the lower is the revenue-maximizing top tax rate.

2.2 Intuition

The intuition for many of the results in this paper can be found by thinking about a simple question: suppose we double the "keep rate" $1-\tau$. What is the long-run effect on GDP?

The answer to this question is determined by the key elasticity $\eta_{Y,1-\tau}$. In particular, if you double the keep rate, then GDP goes up by a factor of $2^{\eta_{Y,1-\tau}}$. To keep things simple, suppose the effort of managers is completely unaffected by taxes and focus on the effect working through ideas, so that $\psi = 0$ and $\eta_{Y,1-\tau} = \gamma\zeta$.

For the baseline calibration, consider $\gamma = 1/2$ and a behavioral elasticity $\frac{\zeta}{1-\zeta} = 0.2$ which implies $\zeta = 1/6$. (Nothing substantive changes if you just set $\zeta = 0.2$ directly; we explain later why the Frisch elasticity is $\frac{\zeta}{1-\zeta} = 0.2$.) In this case, $\eta_{Y,1-\tau} = 1/12$ and $2^{1/12} \approx 1.06$. In other words, going from a top tax rate of 75% to 50% — which doubles the keep rate from 25% to 50% — raises GDP in the long run by only 6%! This is surprisingly small. If anything, it suggests we might want values of γ that are higher, not lower.

Now consider increasing the top tax rate from 50% to 75%. The top income tax base $\Delta\rho$ is 10% of GDP, so without any change in economic behavior, the policy changes top revenue from 5% to 7.5% of GDP, for a gain of 2.5% of GDP in revenue. However, the cost of this policy is 6% of GDP with our baseline parameter values. Focusing on the bottom 90%: their incomes go down by 6% because of the reduction in GDP and then go up by 2.5%/0.9 ≈ 2.8% because of redistribution, for a net loss of about 3.2% of their consumption. The next section drives this point home explicitly by considering the choice of both τ and τ_0 to maximize a social welfare function.

One other thing to appreciate about the 6% gain in GDP from lowering the tax rate from 75% to 50%: while this seems small, notice that it is achieved by a potentially small number of people. How many researchers are there in the economy? Maybe 1% or 5% or 10%? Their effort is increasing by $2^{\zeta} = 1.15$, or by 15%. So a small group of talented researchers working 15% harder raises GDP by 6% in the long run. But recall, that is in some sense the entire point of Romer: a relatively small number of researchers is responsible for the bulk of economic growth for the last 150 years!

3. Maximizing Worker Welfare

In this section, we consider the choice of τ and τ_0 to maximize a social welfare function (instead of taking τ_0 as given and choosing τ to maximize tax revenue). We begin by considering only the welfare of workers, as this yields a clean, closed-form solution.

3.1 Workers

On the worker side, we assume workers are below the top tax threshold and supply one unit of labor inelastically:

$$c^w = w(1-\tau_0) \tag{11}$$

$$u_w(c) = \theta \log c. \tag{12}$$

3.2 The Government Budget Constraint

Suppose the government budget constraint requires raising a fraction Ω of final output in tax revenue, so that $T = \Omega Y$. For example, this revenue may be used in part to pay for the basic research. Using equation (6), we now have

$$\tau_0 Y + (\tau - \tau_0)[\rho Y - \bar{w}(S_a + M)] = \Omega Y. \tag{13}$$

Alternatively, it is sometimes useful to express this equation as

$$[(1-\rho)\tau_0 + \rho\tau]Y - (\tau - \tau_0)[\bar{w}(S_a + M)] = \Omega Y. \tag{14}$$

When Y changes, the effect on tax revenue depends on $(1-\rho)\tau_0 + \rho\tau$.

3.3 Maximizing Utility of Workers

The consumption of a representative worker equals her wage, proportional to $(1 - \tau_0)Y/L$, so the tax system that maximizes the welfare of workers solves

$$\max_{\tau,\tau_0} \ \log(1-\tau_0) + \log Y(\tau) \quad \text{s.t.} \tag{15}$$

$$\tau_0 Y + (\tau - \tau_0)(\rho Y - \bar{w} S_a) = \Omega Y. \tag{16}$$

Some straightforward calculation then gives the following result (a derivation is in the appendix):

Proposition 1 (Maximizing Worker Welfare): *The top tax rate that maximizes the wel-*

fare of workers subject to the government budget constraint satisfies

$$\tau^*_{ww} = \frac{1 - \eta_{Y,1-\tau}\left(\frac{1-\rho}{\Delta\rho}\cdot\tau^*_0 + \frac{1-\Delta\rho}{\Delta\rho}\cdot(1-\tau^*_0) - \frac{\Omega}{\Delta\rho}\right)}{1 + \frac{\rho}{\Delta\rho}\eta_{Y,1-\tau}}. \tag{17}$$

Combining this condition with the government budget constraint gives the explicit solution:

$$\tau^*_{ww} = \frac{1 - \eta_{Y,1-\tau}\left[\frac{1-\Delta\rho}{\Delta\rho} - \frac{\Omega}{\Delta\rho}\left(1 + \frac{\bar\rho}{1-\Delta\rho}\right)\right]}{1 + \frac{\rho}{\Delta\rho}\eta_{Y,1-\tau} + \frac{\bar\rho}{1-\Delta\rho}\eta_{Y,1-\tau}}, \tag{18}$$

where $\eta_{Y,1-\tau} = (\gamma+\psi)\zeta$.

Equation (17) in the proposition has a form that is most easily compared to the revenue-maximizing top tax rate we found in the earlier section in equation (8). Many of the terms are similar.

Relative to the Diamond and Saez result, there are now three new terms in the numerator rather than just one — all multiplied by the key elasticity $\eta_{Y,1-\tau}$. The first term is $\frac{1-\rho}{\Delta\rho}\cdot\tau^*_0$. This term was the novelty in the revenue-maximizing top tax rate back in (8). Recall that it captures the fact that increasing the top tax rate reduces GDP, and therefore reduces the revenue that accrues via the lower tax rate τ_0.

The second new term in the numerator is $\frac{1-\Delta\rho}{\Delta\rho}\cdot(1-\tau^*_0)$. While it is not obvious from these symbols, this term captures the direct effect of a higher tax rate on the welfare of workers: the higher tax reduces GDP which directly reduces the worker's consumption. The revenue-maximizing tax rate considered earlier only captured the effect of lower GDP on lowering tax revenue; this new term captures the fact that lower GDP lowers the consumption of workers as well. The term $1-\tau^*_0$ appears because it is related to the marginal utility of consumption of the worker, and the $\frac{1-\Delta\rho}{\Delta\rho}$ term is a scaling factor that adjusts for the fraction of income taxed at the base rate, $1-\Delta\rho$, versus the fraction taxed at the top rate, $\Delta\rho$.

Finally, the last new term in the numerator is $\Omega/\Delta\rho$. This term appears because we require our tax system to raise an amount of revenue equal to Ω as a share of aggregate income. Other things equal, a higher Ω requires a higher top tax rate.

The second equation in the proposition, equation (18), uses the government budget constraint to eliminate the now-endogenous τ^*_0 from the solution and provides an explicit solution for the top tax rate that maximizes worker welfare.[4]

A useful intuition from this equation comes from studying the point at which a "flat tax" is optimal — that is, where $\tau = \tau_0 = \Omega$, or, equivalently, the keep rates are equal, $\kappa = \kappa_0 = 1 - \Omega$. Equation (18) implies that[5]

$$\tau \leq \tau_0 \text{ and } \kappa \geq \kappa_0 \iff \eta_{Y,1-\tau} \geq \frac{\Delta\rho}{1-\Delta\rho}. \tag{19}$$

Suppose we are considering increasing κ by lowering the top tax rate. The percent gain to GDP, and therefore to a worker's consumption, is $\eta_{Y,1-\tau} \cdot d\log\kappa$. The amount of revenue this requires is $\Delta\rho \cdot d\log\kappa$ since the top tax base is proportional to $\Delta\rho$. Alternatively, we could take that same amount of revenue and redistribute it directly. The redistribution gets divided among $1-\Delta\rho$ "people," so the per worker gain is $\frac{\Delta\rho}{1-\Delta\rho} d\log\kappa$. If these two ways of increasing a worker's consumption yield the same gain, then the two keep rates are the same. If one yields more, then its keep rate will be higher.

For example, consider our baseline parameter values of $\Delta\rho = 0.1$ and a Frisch labor supply elasticity of 0.20, so that $\zeta = 1/6$. In this case, $\frac{\Delta\rho}{1-\Delta\rho} = \frac{1}{9}$ and $\eta_{Y,1-\tau} = \frac{1}{6}(\gamma+\psi)$. Therefore, the flat tax will maximize worker welfare when $\gamma+\psi = 2/3$. And if $\gamma+\psi > 2/3$, then the top tax rate will be *below* τ_0.

Table 2 shows some numerical examples for the baseline parameter values including $\Omega = 0.2$. Even with a Frisch labor supply elasticity of 0.20, the top tax rate falls to 21.7% for $\gamma = 1/2$. Notice that with $\psi = 0.15$, this case means that $\gamma + \psi = 0.65$, which is close to the flat tax value of 2/3. When $\gamma = 1$, the top tax rate is solidly into negative territory at -24.7%: subsidizing idea creation is even more effective than redistribution

[4] A higher $\eta_{Y,1-\tau}$ (and therefore a higher γ or ζ) lowers the top tax rate, provided Ω is not too large. A higher Ω raises the top tax rate. The main term that is novel in this expression is $\frac{\bar{\rho}}{1-\Delta\rho}$. Looking back at equation (17), one can see that τ_0 entered twice. Combining those two terms reveals that the net effect depends on $[(1-\rho)-(1-\Delta\rho)]$. The $1-\Delta\rho$ term captures the tax base to which τ_0 applies, while the $1-\rho$ term captures the extra revenue that comes from the base tax rates when Y changes; recall equation (14). It is the net of these two effects that matters, and $[(1-\rho)-(1-\Delta\rho)] = \Delta\rho - \rho = -\bar{\rho}$. In other words, a higher τ_0, other things equal, lowers τ^*_{ww}, as the behavioral effect is larger than the mechanical tax base effect. The $\bar{\rho}$ expression enters the tax rate that maximizes worker welfare twice. The reason for this is that the government budget constraint means that $\tau_0 = \frac{\Omega-\Delta\rho\tau}{1-\Delta\rho}$, so there is an additional Ω term (in the numerator of 18) and an additional multiplier effect coming through the fact that τ_0 depends on τ.

[5] It is easier to see this result using the following solution for the top keep rate that maximizes worker welfare:

$$\kappa^*_{ww} = \frac{\eta_{Y,1-\tau}\frac{1-\Omega}{\Delta\rho}\left(1+\frac{\bar{\rho}}{1-\Delta\rho}\right)}{1+\frac{\rho}{\Delta\rho}\eta_{Y,1-\tau}+\frac{\bar{\rho}}{1-\Delta\rho}\eta_{Y,1-\tau}}.$$

Table 2: Tax Rates that Maximize Worker Welfare

Degree of IRS, γ	Behavioral elast. = 0.2 τ^*_{ww}	τ^*_0	Behavioral elast. = 0.5 τ^*_{ww}	τ^*_0
1/8	0.639	0.151	0.323	0.186
1/4	0.490	0.168	0.067	0.215
1/2	0.217	0.198	-0.368	0.263
1	-0.247	0.250	-1.028	0.336

Note: Shows the values of τ and τ_0 that maximize worker welfare based on Proposition 1. Assumes $\Omega = 0.2$ in addition to the other baseline parameter values.

as raising worker welfare.

4. Maximizing Utilitarian Social Welfare

In order to incorporate the welfare of entrepreneurs and managers, we have to specify their utility function and how they choose effort.

4.1 Preferences and Equilibrium Effort

Assume entrepreneurs and managers have utility functions $u(c,e)$ that depend on consumption c and effort e with the following functional form:

$$u(c,e) = \theta \log c - \zeta e^{1/\zeta}. \tag{20}$$

This specification is a simple form of "constant Frisch elasticity" preferences and delivers constant effort in the presence of exponential growth in consumption. The (compensated) Frisch elasticity of effort with respect to the wage is then $\frac{\zeta}{1-\zeta}$. This is what we've called the "behavioral elasticity" in the earlier calculations.

We focus in what follows on entrepreneurs, but managers are treated symmetrically. An entrepreneur with talent z solves the following problem:

$$\max_{c,e} u(c,e) \quad \text{s.t.} \tag{21}$$

$$
\begin{aligned}
c =& \bar{w}(1-\tau_0) + [w_s e z - \bar{w}](1-\tau) + R \\
=& \bar{w}(1-\tau_0) - \bar{w}(1-\tau) + w_s e z(1-\tau) + R \\
=& \bar{w}(\tau-\tau_0) + w_s e z(1-\tau) + R
\end{aligned}
$$

where R is a lump-sum rebate of tax revenue, discussed further shortly, that the agent takes as given.

Given our functional form on utility, this leads to the following FOC:

$$
e^{\frac{1}{\zeta}-1} = \frac{\theta w_s z(1-\tau)}{c}. \tag{22}
$$

And the Frisch elasticity of effort with respect to the wage or the keep rate, holding consumption constant, is obviously $\frac{\zeta}{1-\zeta}$ as noted earlier.

To complete the solution of the entrepreneur's problem, we need to specify the rebate, R, and an important issue arises. Recall that in general with log preferences, the SE and IE cancel, so that in the absence of a rebate, a linear tax will have no effect on effort. This is not true for the nonlinear tax we have here. Instead, the top tax rate will reduce effort, but in a heterogeneous fashion that asymptotes to zero for highly talented people.[6]

That heterogeneity might be interesting to explore, but it is not especially clean. Instead, we take a different approach below. In particular, we assume that the rebates are heterogenous and such that equilibrium consumption of a researcher with talent z is given by

$$
c_z = w_s e z(1-\tau)^{1-\alpha} \tag{23}
$$

where α parameterizes the extent to which the rebate depends on the tax rate.[7]

[6]Setting $R = 0$ and combining the FOC with the entrepreneur's budget constraint gives

$$
e^{1/\zeta-1}\left(e + \frac{\bar{w}(\tau-\tau_0)}{w_s z(1-\tau)}\right) = \theta
$$

and certainly if ζ is not too large, an increase in τ will reduce e. But as $z \to \infty$, e becomes independent of τ.

[7]That is, the researcher solves the problem in equation (21) taking the following rebates as exogenous:

$$
R(z) = w_s e^* z(1-\tau)^{1-\alpha} - \bar{w}(\tau-\tau_0) - w_s e^* z(1-\tau)
$$

where e^* does not depend on z and is given in equation (24). Note that these rebates differ according to the entrepreneur's skill, z. This raises an interesting question about how this could be microfounded in a Mirrleesian framework in which it is typically assumed that the government cannot observe individuals' skills. This issue does not arise in the revenue-maximizing or the maximize-worker-welfare tax systems

The FOC for effort then reduces to

$$e^* = [\theta(1-\tau)^{\alpha}]^{\zeta}, \tag{24}$$

meaning that entrepreneurs of differing talent put in the same effort in equilibrium. Importantly, notice that the elasticity of effort with respect to the keep rate is $\alpha\zeta$, so that $\eta_{Y,1-\tau} = \alpha(\gamma+\psi)\zeta$. The parameter α therefore influences the top tax rate.

Consumption, however, does depend on talent, z:

$$c_z^* = \theta^{\zeta} w_s z (1-\tau)^{1-\alpha(1-\zeta)}. \tag{25}$$

If $\alpha = 1$, both consumption and effort are proportional to $(1-\tau)^{\zeta}$. Alternatively, a smaller α moves us closer to the situation where the SE and IE cancels, making effort less sensitive to the tax rate and consumption more sensitive so that taxes distort less and transfer more.

4.2 Utilitarian Social Welfare

A social welfare function of interest is the utilitarian or "behind the veil of ignorance" specification:

$$SWF \equiv Lu(c^w) + S_b u(c^b) + S_a \int u(c_z^s, e_z^s) dF(z) + M \int u(c_z^m, e_z^m) dF(z).$$

Substituting in the expressions for $c^w, c^b, c_z^s, e_z^s, c_z^m, e_z^m$ and the income share parameters (e.g. ρ), this expression can be rewritten as

$$\begin{aligned} \overline{SWF} \equiv & \frac{SWF - \varphi}{\theta} \cdot \frac{1}{L + S_b + S_a + M} \\ = & \log Y + \ell \log(1-\tau_0) + s[(1-\alpha)\log(1-\tau) - \zeta(1-\tau)^{\alpha}], \end{aligned} \tag{26}$$

where $s \equiv \frac{S_a+M}{L+S_b+S_a+M}$, $\ell \equiv 1-s$, and φ is a term that does not depend on τ or τ_0.[8]

and so is not central to the present paper. I therefore leave this interesting question to future research.

[8] In particular,

$$\varphi \equiv \theta L \log \frac{\rho_\ell}{L} + \theta S_b \log \frac{\rho_b}{S_b} + \theta \left(S_a \log \frac{\rho_s}{\tilde{z} S} + M \log \frac{\rho_m}{\tilde{z} M} \right) + \theta (S_a + M)\tilde{z},$$

where $\tilde{z} \equiv \int \log z \, dF(z)$.

As shown in the appendix, we then have the following result:

Proposition 2 (Maximizing Social Welfare): *The tax rates that maximize the social welfare function in (26) subject to the government budget constraint satisfy the following two equations in two unknowns, written in terms of the keep rates κ and κ_0 (i.e. $\kappa \equiv 1-\tau$):*

$$\alpha\zeta s\kappa^{\alpha} + \frac{\kappa}{\kappa_0} \cdot \frac{\ell}{1-\Delta\rho}\left(\Delta\rho + \bar{\rho}\eta_{Y,1-\tau}\right) = \eta_{Y,1-\tau}\left(1 + \frac{\bar{\rho}\ell}{1-\Delta\rho}\right) + s(1-\alpha) \tag{27}$$

$$\kappa_0(1-\Delta\rho) + \kappa\Delta\rho = 1 - \Omega. \tag{28}$$

Unfortunately, these equations do not admit a closed-form solution. However, the proposition is easy to analyze graphically. For example, in the special case of $\alpha = 1$, the first equation in (27) can be rewritten as

$$\kappa = \frac{\eta_{Y,1-\tau}\left(1 + \frac{\bar{\rho}\ell}{1-\Delta\rho}\right)}{s\zeta + \frac{1}{\kappa_0}\frac{\ell}{1-\Delta\rho}\left(\Delta\rho + \bar{\rho}\eta_{Y,1-\tau}\right)} \tag{29}$$

And then our two equations — (28) and (29) — can be analyzed graphically, as in Figure 1. In fact, inspecting equation (27) reveals that it delivers κ as an increasing function of κ_0, so Figure 1 applies more generally even when α differs from one.

Comparative statics can be inferred by totally differentiating the two equations. In particular, a sufficient condition for the "expected" comparative statics is that the tax system is weakly progressive, i.e. $\kappa \leq \kappa_0$. In that case:

- $\uparrow\gamma \Rightarrow \uparrow\kappa$ and $\downarrow\kappa_0$: The more important are ideas, the lower the top tax rate.
- $\uparrow\zeta \Rightarrow \uparrow\kappa$ and $\downarrow\kappa_0$: The more elastic are top earners, the lower the top tax rate.
- $\uparrow\Omega \Rightarrow \downarrow\kappa$ and $\downarrow\kappa_0$: The more the government needs to raise, the higher the tax rates.

4.3 Numerical Examples

We can also proceed numerically. The examples shown next assume $\Omega = 0.2$, $s = 0.1$, and $\Delta\rho = 0.1$. Table 3 shows results for $\alpha = 1$, while Table 4 shows results for $\alpha = 1/2$.

Two properties of these examples stand out. First, for $\gamma \geq 1/2$, the top marginal tax rate is quite low, at 23% or lower. Despite the utilitarian impetus to redistribute from

Figure 1: Maximizing Social Welfare: $\alpha = 1$

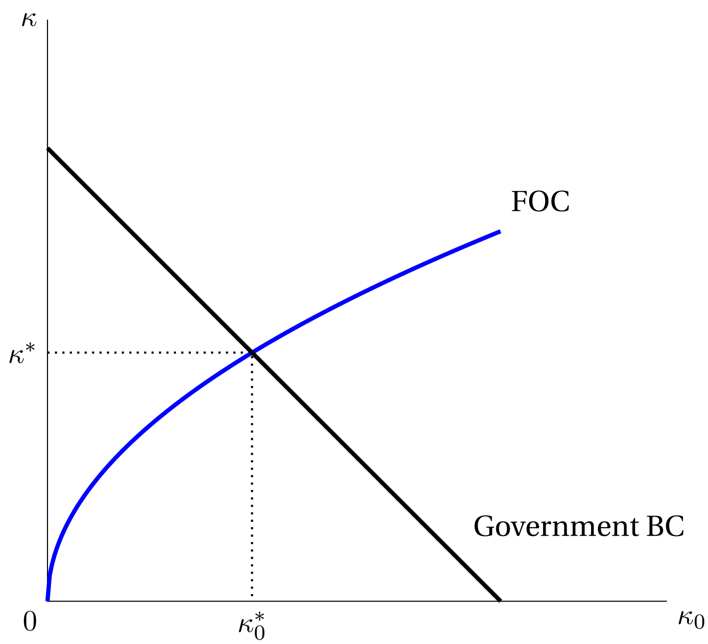

Table 3: Tax Rates to Maximize Utilitarian Welfare: $\alpha = 1$

	Behavioral elast. = 0.2			Behavioral elast. = 0.5		
Degree of IRS, γ	τ^*	τ_0^*	GDP loss if $\tau = 0.75$	τ^*	τ_0^*	GDP loss if $\tau = 0.75$
1/8	0.649	0.150	0.7%	0.400	0.178	3.6%
1/4	0.502	0.166	2.8%	0.163	0.204	9.6%
1/2	0.231	0.197	8.9%	-0.255	0.251	23.6%
1	-0.238	0.249	23.4%	-0.919	0.324	49.3%

Note: Shows the SWF-maximizing values of τ and τ_0, calculated numerically based on Proposition 2. Assumes $\Omega = 0.2$ and $s = 0.1$, in addition to the other baseline parameter values. "GDP loss if $\tau = 0.75$" is computed by noting that $Y \propto (1-\tau)^{\alpha(\gamma+\psi)\zeta}$ and reports the percentage decline in GDP from moving from the welfare-maximizing tax system to one with a top tax rate of 75%.

Table 4: Tax Rates to Maximize Utilitarian Welfare: $\alpha = 1/2$

	Behavioral elast. = 0.2			Behavioral elast. = 0.5		
Degree of IRS, γ	τ^*	τ_0^*	GDP loss if $\tau = 0.75$	τ^*	τ_0^*	GDP loss if $\tau = 0.75$
1/8	0.445	0.173	0.8%	0.328	0.186	2.0%
1/4	0.369	0.181	1.9%	0.189	0.201	4.8%
1/2	0.222	0.198	4.6%	-0.070	0.230	11.4%
1	-0.047	0.227	11.3%	-0.517	0.280	26.0%

Note: See notes to Table 3.

high earners to low earners, the fact that the discovery of ideas lies at the heart of the economic growth experienced by all earners restrains top income taxation.

Second, these examples show that if ideas are sufficiently important, this force can result in a *negative* top tax rate. That is, it can be welfare-maximizing to subsidize the incomes of higher earners at the margin.

The tables also report the "GDP loss if $\tau = 0.75$." These columns show the percentage decline in GDP from moving from the welfare-maximizing tax system to one with a top tax rate of 75%, computed by noting that $Y \propto (1-\tau)^{\alpha(\gamma+\psi)\zeta}$.[9] For example, when $\gamma = 1/2$, $\alpha = 1$, and the behavioral elasticity is 0.2, the welfare-maximizing top tax rate is 0.23, and GDP declines by 8.9% if instead the top tax rate is 0.75.

4.4 Summary

We summarize our calibration exercises in Table 5. As we showed at the start of this paper, our basic setup delivers extremely high revenue-maximizing top tax rates in the absence of the idea channel ($\gamma = 0$). When the idea channel is present, however, top tax rates fall considerably. Revenue-maximizing top rates fall from 92% to 70% with $\gamma = 1/2$. However, once the idea channel is present, revenue-maximization is no longer even approximately the same as welfare maximization: the welfare of everyone in the economy is affected by the discovery of new ideas. The top tax rate that maximizes

[9] This can be seen by combining equation (4) with (24).

Table 5: Summary of Calibration Exercises

Exercise	Top rate, τ	
No ideas, $\gamma = 0$		
Revenue-maximization, $\tau_0 = 0$	0.96	
Revenue-maximization, $\tau_0 = 0.20$	0.92	
With ideas	$\gamma = 1/2$	$\gamma = 1$
Revenue-maximization	0.70	0.52
Maximize worker welfare	0.22	-0.25
Maximize utilitarian welfare	0.22	-0.05

Note: This table collects the results from the various calibration exercises for the default parameter values of $\Delta\rho = 0.10$, $s = 0.10$, and a Frisch elasticity of labor supply for top earners of 0.2. In the second and third rows, we add the consideration that $\tau_0 = 0.2$. In the final two rows, we instead require the tax system to raise 20% of GDP in revenue ($\Omega = 0.2$), and the last row sets $\alpha = 1/2$.

worker or utilitarian welfare is considerably lower, falling to 22% for $\gamma = 1/2$, for example. Taking ideas into account has first-order effects on optimal top income taxation in this setting.

5. Discussion

In this section, we discuss several issues related to economic growth and the taxation of top incomes.

5.1 Empirical Evidence on Growth and Taxes

Stokey and Rebelo (1995) first observed a remarkable fact related to growth and taxes: the growth rate of the U.S. economy in the 20th century was relatively stable while taxes as a share of GDP increased substantially. Similarly, there are large movements in top marginal tax rates that are not associated with systematic changes in the growth rate. For example, the top marginal tax rate in the U.S. reached 94% at the end of World War II and exceeded 90% during much of the 1950s and early 1960s [Internal Revenue

Service (2018)]. Yet these years exhibit, if anything, some of the fastest growth in the U.S. economy, not the slowest. Top marginal rates declined in the next two decades, but so did growth rates. These facts must be confronted by any theory of economic growth.

Semi-endogenous growth theory (in which policy has level effects as in Solow, rather than long-run growth effects) helps, but it is not a complete resolution. In particular, other things equal, one would expect large declines in top marginal tax rates to raise growth rates at least along a transition path to a higher steady state.

Instead, I see two possible ways to reconcile the basic facts with growth theory. First, the "other things equal" qualification is clearly not met. What the government does with the tax revenue is important, and in the decades during and after World War II, there were massive investments in basic research. Similarly, the decline in marginal tax rates in recent decades has been paired with a decline in funding for basic research as a share of GDP. We do not know what the counterfactual path of economic growth in the 20th century would look like had taxes not changed.

Second, reductions in the top tax rate may shift people from producing goods to producing ideas; while this extensive margin is not modeled in this paper, it features in other papers such as Jaimovich and Rebelo (2017). The impact effect of this change is to *reduce* the production of goods, and this may show up as lower GDP growth in the near term before the positive effects through increased innovation materialize.[10]

5.2 The Social Rate of Return

At some basic level, there is an optimal amount that needs to be spent as compensation for research effort, and the question in the model boils down to "How big is the gap between the equilibrium share and the optimal amount to pay for research?" If there is a big gap, then the implied social return to research is high, and we want to subsidize research. In this section, we show how to connect this interpretation to our analysis.

Jones and Williams (1998) discuss how to calculate a social rate of return to R&D. In particular, they consider the following variational argument: suppose you reduce consumption today by one unit, invest it in R&D to create new ideas, and then re-

[10]There are two channels for lower GDP growth in the near term. First, in the old national income accounts, R&D was treated as an intermediate good and so would not be counted. Second, to the extent that there is a lag between research and innovation, the GDP gains may take longer to appear.

duce R&D tomorrow, consuming the proceeds, so as to leave the future time path of ideas unchanged. What is the percentage gain in consumption that this variation can achieve? They show that the optimal R&D investment rate in an endogenous growth model occurs where this social rate of return equals the social planner's interest rate as implied by a standard Euler equation for consumption.

Applying their analysis here indicates that the social return to entrepreneurial effort in our model is given by[11]

$$\tilde{r} = g_Y + \lambda g_y \left(\frac{1}{\rho_s(1-\tau)} - \frac{1}{\gamma} \right) \tag{30}$$

where $\rho_s \equiv w_s \tilde{S}_a / Y$ is the share of income paid to entrepreneurs. At the margin, after taxes, entrepreneurs earn $\rho_s(1-\tau)$. Speaking loosely, this equation suggests that the optimal subsidy is approximately such that $\rho_s(1-\tau) = \gamma$. That is, you want entrepreneurs to earn a share of income roughly equal to the importance of applied ideas in the economy.[12]

To put some numbers on this, suppose $g_y = 2\%$, $g_Y = 3\%$, $\gamma = 1/2$, and $\rho_s = 10\%$. For $\tau = 1/2$, equation (30) implies that the social rate of return to applied research is 39%. If the share of income earned by entrepreneurs, ρ_s, equals 5% instead, the social return rises to 79%; with $\rho_s = 20\%$, the return is 16%.

These numbers are in line with estimates of the social rate of return to research in the literature — for example, see Jones and Williams (1998) and Hall, Mairesse and Mohnen (2010) for surveys. In one of the most recent and best-identified studies, Bloom, Schankerman and Van Reenen (2013) estimate a social rate of return of 59%, identified from state-specific changes in R&D tax policy in a panel of U.S. firms.[13] Of course it is not at all clear that social returns to formal R&D should be compared to the social

[11] A full derivation is presented in Appendix B.

[12] Interestingly, the "optimal" subsidy does not equate the social rate of return to the real interest rate, partly because there are different definitions of "optimal" that we consider in this paper: revenue maximization, maximizing worker welfare, and maximizing social welfare. These are evidently not the same as the variational argument regarding consumption. Also, we say "roughly" because the planner's interest rate is not g_Y, but this is close.

[13] Two other recent papers focus on careful identification of the benefits of R&D policy. Howell (2017) uses a regression-discontinuity design to identify the effect of the Department of Energy's SBIR grant program. Budish, Roin and Williams (2015) provide a detailed analysis of clinical trials for cancer treatments. While neither paper formally computes an estimate of the social return to research, both find evidence suggestive of large social returns. Lucking and Bloom (2017) update the analysis of Bloom, Schankerman and Van Reenen (2013) so that the latest year is 2015 instead of 2001 and find a social return to research of 58%.

return to effort in our model, particularly since one of our key points is that applied research is not easily subsized through formal R&D subsidies, while the social return estimate in Bloom et al is identfied precisely from state-specific R&D subsidies. Still, the calculation and comparison does seem useful at some level — one might expect the social returns to non-formal research to be even higher, perhaps, because it does not benefit from the existing subsidies — and suggests that our baseline parameter values are not unreasonable.

6. The Full Growth Model

This section is basically housekeeping. It presents a fully-worked out dynamic growth model and shows that the allocation along the balanced growth path is characterized by the simple static model that we've used in the main body of the paper up until now. The basic economic environment for the full growth model is summarized in Table 6.

6.1 Equilibrium with Imperfect Competition and Taxes

In this section, we study the equilibrium with imperfect competition first described for a model like this by Romer (1990). Briefly, the economy consists of three sectors. A final goods sector produces the consumption-output good using managers and a collection of intermediate goods. The intermediate goods sector produces a variety of different intermediate goods using ideas and labor. Finally, talented entrepreneurs produce new ideas, which in this model are represented by new kinds of intermediate goods. These entrepreneurs use basic research ideas when they innovate, and basic research is funded entirely by the government. The final goods sector and the research sector are perfectly competitive and characterized by free entry, while the intermediate goods sector is the place where imperfect competition is introduced. When a new design for an intermediate good is discovered, the design is awarded an infinitely-lived patent. The owner of the patent has the exclusive right to produce and sell the particular intermediate good and therefore acts as a monopolist in competition with the producers of other kinds of intermediate goods. The monopoly profits that flow to this producer ultimately constitute the compensation to the entrepreneurs who discovered the new design in the first place.

Table 6: Economic Environment for Growth Model

Final output	$Y_t = \int_0^{A_t} x_{it}^{1-\psi} di\, [\mathbb{E}(ez) M_t]^{\psi}$
Production of variety i	$x_{it} = \ell_{it}$
Resource constraint (ℓ)	$\int \ell_{it} di = L_t$
Resource constraint (N)	$L_t + S_{bt} = N_t$
Population growth	$N_t = \bar{N} \exp(nt)$
Entrepreneurs	$S_{at} = \bar{S}_a \exp(nt)$
Managers	$M_t = \bar{M} \exp(nt)$
Applied ideas	$\dot{A}_t = \bar{a}[\mathbb{E}(ez) S_{at}]^{\lambda} A_t^{\phi_a} B_t^{\alpha}$
Basic ideas	$\dot{B}_t = \bar{b} S_{bt}^{\lambda} B_t^{\phi_b}$
Talent heterogeneity	$z_i \sim F(z)$
Utility (S_a, M)	$u(c, e) = \theta \log c - \zeta e^{1/\zeta}$
Utility (L, S_b)	$u(c) = \theta \log c$

We begin by stating the key decision problems that have to be solved by the various agents in the economy and then we put these together in our formal definition of equilibrium.

Problem (HH). Households solve a standard optimization problem, choosing a time path of consumption and an allocation of time. There are three types of households: workers (who can make goods or work as basic researchers), entrepreneurs, and managers. The problem for workers is assumed to be simple — they just work and consume their income after taxes:

$$c^w = w(1-\tau_0) \tag{31}$$

The problems for entrepreneurs and managers are essentially identical, apart from the fact that their wage per efficiency unit of effort can vary. For entrepreneurs: taking the time path of $\{w_{st}, r_t, R_t\}$ as given, they solve

$$\max_{\{c_t, e_t\}} \int_0^\infty u(c_t, e_t) \exp(-\delta t)\, dt, \tag{32}$$

subject to

$$\dot{v}_t = (r_t - n)v_t + \bar{w}(\tau - \tau_0) + w_s e z(1-\tau) + R_t - c_t, \;\; v_0 \;\; given, \tag{33}$$

$$u(c,e) = \theta \log c - \zeta e^{1/\zeta} \tag{34}$$

$$\lim_{t\to\infty} v_t \exp(-\int_0^t (r_s - n) ds) \geq 0. \tag{35}$$

where v_t is the financial wealth of an individual, r_t is the interest rate, and R_t is the lump-sum rebate for people of type z, which they take as given. We are suppressing the dependence of v_t and R_t on type z to simplify the notation.

Problem (FG). A perfectly competitive final goods sector takes the variety of intermediate goods in existence as given and produces output according to

$$Y_t = \tilde{M}_t^{\psi} \int_0^{A_t} x_{it}^{1-\psi} di. \tag{36}$$

That is, at each point in time t, taking the manager's wage rate w_{mt}, the measure of intermediate goods A_t, and the prices of the intermediate goods p_{it} as given, the representative firm solves

$$\max_{\{x_{it}\},\tilde{M}_t} \tilde{M}_t^{\psi} \int_0^{A_t} x_{it}^{1-\psi}\, di - w_{mt}\tilde{M}_t - \int_0^{A_t} p_{it}x_{it}di. \tag{37}$$

Problem (IG). Each variety of intermediate good is produced by a monopolist who owns a patent for the good, purchased at a one-time price P_{At}. As discussed in describing the economic environment, one unit of the intermediate good can be produced with one unit of labor. The monopolist sees a downward-sloping demand curve for her product from the final goods sector and chooses a price to maximize profits. That is, at each point in time and for each intermediate good i, a monopolist solves

$$\max_{p_{it}} \pi_{it} \equiv (p_{it} - w_t)x(p_{it}) \tag{38}$$

where $x(p_{it})$ is the demand from the final goods sector for intermediate good i if the price is p_{it}. This demand curve comes from a first-order condition in Problem (FG).[14]

Problem (R&D). The production function for new ideas depends on entrepreneurs, the existing stock of applications, A_t, and the stock of fundamental ideas, B_t:

$$\dot{A}_t = \bar{a}\tilde{S}_{at}^{\lambda}A_t^{\phi_a}B_t^{\alpha} \tag{39}$$

where $\tilde{S}_{at}$ is the efficiency units of applied research effort. Each individual research firm is small and takes the productivity of the idea production function as given. In particular, each research firm assumes that the idea production function is

$$\dot{A}_t = \bar{\nu}_t\tilde{S}_{at}. \tag{40}$$

That is, the duplication effects associated with λ and the knowledge spillovers associated with ϕ_a and α in equation (39) are assumed to be external to the individual

[14]To be more specific, the demand curve $x(p_i)$ is given by

$$x(p_{it}) = \left((1-\psi)\tilde{M}_t^{\psi} \cdot \frac{1}{p_{it}}\right)^{1/\psi}.$$

research firm. In this perfectly competitive sector, the representative firm solves

$$\max_{\tilde{S}_{at}} P_{At}\bar{\nu}_t\tilde{S}_{at} - w_{st}\tilde{S}_{at}, \tag{41}$$

taking the price of ideas P_{At}, entrepreneurial productivity $\bar{\nu}_t$, and the wage w_{st} as given.

The Government and Basic Research. To keep things simple, we assume the government hires basic researchers to produce the fundamental research ideas. If the government spends G_{bt} on hiring basic researchers at the wage w_t, then $G_{bt} = w_t S_{bt}$ implicitly determines the number of researchers that get hired, and these researchers produce basic ideas according to

$$\dot{B}_t = \bar{b} S_{bt}^{\lambda} B_t^{\phi_b}. \tag{42}$$

We've assumed for simplicity that applied ideas A_t do not feed back into basic research, but this could be easily relaxed.

The government collects taxes from everyone in the economy and uses these revenues to fund basic research, the tax rebates, and other spending, G_{ot}, which together equal a fraction Ω of final output:

$$\begin{aligned}\tau_0[wL + wS_b + w_a\mathbb{E}(ez)S_a + w_m\mathbb{E}(ez)M] + (\tau - \tau_0)[(w_a\mathbb{E}(ez) - \bar{w})S_a + (w_m\mathbb{E}(ez) - \bar{w})M]\\ = G_{bt} + \int R(z)dF(z) + G_{ot} \equiv \Omega Y_t.\end{aligned} \tag{43}$$

Now that these decision problems have been described, we are ready to define the equilibrium.

An *equilibrium* in this economy consists of time paths for the allocations $\{c^s_{zt}, c^m_{zt}, c^w_t, e^s_{zt}, e^m_{zt}, v^s_{zt}, v^m_{zt}, \{x_{it}\}, Y_t, \tilde{M}_t, \tilde{S}_{at}, S_{bt}, L_t, A_t, B_t, \bar{\nu}_t, G_{bt}, G_{ot}\}_{t=0}^{\infty}$ and prices $\{w_t, w_{st}, w_{mt}, r_t, \{p_{it}\}, \{\pi_{it}\}, R_t(z), P_{At}\}_{t=0}^{\infty}$ such that for all t:

1. $c^s_{zt}, c^m_{zt}, c^w_t, e^s_{zt}, e^m_{zt}, v^s_{zt}$, and v^m_{zt} solve Problem (HH).
2. $\{x_{it}\}$ and $\tilde{M}_t$ solve Problem (FG).
3. p_{it} and π_{it} solve Problem (IG) for all $i \in [0, A_t]$.
4. $\tilde{S}_{at}$ solves Problem (R&D).

5. S_{bt} is determined by the government: $S_{bt} = G_{bt}/w_t$.

6. (r_t) The capital market clears: $\int (S_{at}v^s_{zt} + M_t v^m_{zt})dF(z) = P_{At}A_t$.

7. (w_t) The labor market clears: $L_t + S_{bt} = N_t$.

8. (w_{st}) The market for entrepreneurs clears: $\mathbb{E}(ez)S_{at} = \tilde{S}_{at}$

9. (w_{mt}) The manager market clears: $\mathbb{E}(ez)M_t = \tilde{M}_t$

10. $(\bar{\nu}_t)$ The idea production function is satisfied: $\bar{\nu}_t = \nu \tilde{S}_{at}^{\lambda-1} A_t^{\phi_a} B_t^{\alpha}$.

11. (P_{At}) Assets have equal returns: $r_t = \frac{\pi_{it}}{P_{At}} + \frac{\dot{P}_{At}}{P_{At}}$.

12. The tax rebates are $R_t(z) = w_s e^* z(1-\tau)^{1-\alpha} - [\bar{w}_t(\tau-\tau_0) + w_s e^* z(1-\tau) + (r-g_Y)v_t]$.

13. Y_t is given by the production function in (36).

14. L_t satisfies the resource constraint for labor: $\int x_{it}di = L_t$

15. A_t is given by the production function in (39).

16. B_t is given by the production function in (42).

17. G_{bt} is exogenously given.

18. (G_{ot}) The government budget constraint in (43) is satisfied.

Notice that, roughly speaking, there are twenty-six equilibrium objects that are part of the definition of equilibrium and there are twenty-six equations described in the conditions for equilibrium that determine these objects at each point in time. Not surprisingly, one cannot solve in general in closed form for the equilibrium outside of the balanced growth path, but along a balanced growth path the solution is relatively straightforward, and we have the following results.

(a) Because of symmetry considerations, the production function for final output can be written as

$$Y = A^{\psi} \tilde{M}^{\psi} L^{1-\psi}. \tag{44}$$

The stock of ideas along the BGP is in turn given by[15]

$$A_t^* = \left(\nu_a \tilde{S}_{at} S_{bt}^{\frac{\alpha}{1-\phi_b}}\right)^{\frac{\lambda}{1-\phi_a}}. \tag{45}$$

Putting these two equations together,

$$Y_t^* = \left(\nu_a \mathbb{E}[ez] S_a S_b^{\beta}\right)^{\gamma} \tilde{M}_t^{\psi} L_t^{1-\psi} \tag{46}$$

where $\gamma \equiv \frac{\psi\lambda}{1-\phi_a}$ and $\beta \equiv \frac{\alpha}{1-\phi_b}$. This is exactly the specification of the production function for our simple model in (4).

Letting $y \equiv Y/(N + M + S_a)$ denote per capita income,

$$y_t^* \propto \left(\tilde{S}_{at} S_{bt}^{\beta}\right)^{\gamma}. \tag{47}$$

Taking logs and derivatives, growth rates along the BGP are given by

$$g_y = \gamma(1+\beta) g_S = \gamma(1+\beta) n. \tag{48}$$

(b) The Euler equation for consumption and the effort choice for a person of talent z are

$$\frac{\dot{c}_t(z)}{c_t(z)} = \frac{\dot{c}_t}{c_t} = (r_t^{eq} - n - \delta) \tag{49}$$

As we did earlier, we've chosen the lump-sum rebates (in the definition of equilibrium) to deliver a steady-state consumption of

$$c_t^*(z) = \theta^{\zeta} w_s z (1-\tau)^{1-\alpha(1-\zeta)}. \tag{50}$$

and then equilibrium effort is independent of z:

$$e^* = [\theta(1-\tau)^{\alpha}]^{\zeta} \tag{51}$$

[15]Define

$$\nu_a^{\frac{\lambda}{1-\phi_a}} \equiv \left(\frac{\bar{a}}{g_A}\right)^{\frac{1}{1-\phi_a}} \left(\frac{\bar{b}}{g_B}\right)^{\frac{\alpha}{1-\phi_b}\cdot\frac{1}{1-\phi_a}}$$

(c) Profit maximization. The solution to Problem (FG) implies that managers and intermediate goods receive constant shares of factor payments:

$$\frac{w_m \tilde{M}}{Y} = \psi \equiv \rho_m \quad \text{and} \quad \frac{Apx}{Y} = 1 - \psi \tag{52}$$

The solution to Problem (IG) involves a monopoly markup over marginal cost that depends on the CES parameter in the usual way:

$$p_{it}^{eq} = p_t^{eq} = \frac{w_t}{1 - \psi}. \tag{53}$$

These equations then imply that total profits of intermediate goods firms satisfy

$$\frac{A\pi}{Y} = \psi(1 - \psi). \tag{54}$$

Also, recall from the resource constraint for labor that $Ax = L$. Substituting this into (52) together with $p = w/(1 - \psi)$ yields

$$\frac{wL}{Y} = (1 - \psi)^2 \equiv \rho_\ell. \tag{55}$$

(d) Research. The perfectly competitive research sector ensures that the wage per unit of research satisfies $w_s = P_A \cdot \frac{\dot{A}}{\tilde{S}_a}$ which implies

$$\frac{w_s \tilde{S}_a}{Y} = g_A \frac{P_A A}{Y}. \tag{56}$$

The arbitrage equation pins down the price of an idea as

$$P_A = \frac{\pi}{r - g_\pi} = \frac{\pi}{r - (g_Y - g_A)}. \tag{57}$$

Therefore

$$\frac{P_A A}{Y} = \frac{1}{r - (g_Y - g_A)} \cdot \frac{A\pi}{Y} = \frac{1}{r - (g_Y - g_A)} \cdot \psi(1 - \psi). \tag{58}$$

And substituting this back into (56) gives

$$\frac{w_s \tilde{S}_a}{Y} = \frac{g_A}{r - (g_Y - g_A)} \cdot \psi(1-\psi) \equiv \rho_s \tag{59}$$

Notice that because the interest rate and growth rates of the economy are invariant to policy in the long run, the share of income paid to entrepreneurs ρ_s is as well.

In equilibrium, raw labor is paid less than its marginal product so that some of final output can be used to compensate entrepreneurs. As Romer (1990) pointed out, because of the increasing returns associated with the nonrivalry of ideas, all factors cannot be paid their marginal products. There is not enough final output to go around.

Notice that $\rho_\ell + \rho_m + \rho_s \neq 1$. In particular, what is true is that $Y = wL + w_m\tilde{M} + A\pi$ and $A\pi > w_s\tilde{S}_a$. That is, there are some "profits" left over associated with owning patents in this economy (that get paid to entrepreneurs and managers through the claims on profits in asset markets). Nevertheless, it remains true that $wL + w_m\tilde{M} + w_s\tilde{S}_a$ is proportional to Y with a constant (and invariant to taxes) factor of proportionality in the steady state, so everything in the simple model carries through ignoring this subtlety.[16]

Overall, then, this section explains how the steady state of the full model captures the simple model that we used at the start of the paper.

7. Conclusion

This paper considers the taxation of top incomes when the following conditions apply: (i) new ideas drive economic growth, (ii) the reward for creating a successful innovation is a top income, and (iii) innovation cannot be perfectly targeted by a separate research subsidy. These conditions lead to a new force affecting the optimal top tax rate: by slowing the creation of the new ideas that drive aggregate GDP, top income taxation reduces everyone's income, not just the income at the top. When the creation of ideas is the ultimate source of economic growth, this force sharply constrains both revenue-

[16] One might wonder about where the payments to basic researchers wS_b show up. The equation $Y = wL + w_m\tilde{M} + A\pi$ describes the payments made to factors, not the income received. The income received gets taxed, and these taxes go, in part, to pay for basic research.

maximizing and welfare-maximizing top tax rates. For example, for extreme parameter values, maximizing the welfare of the middle class requires a *negative* top tax rate: the higher income that results from the subsidy to innovation more than makes up for the lost redistribution. More generally, the calibrated model suggests that incorporating ideas and economic growth cuts the optimal top marginal tax rate substantially relative to the basic Saez calculation.

The point of this paper is not to provide a precisely calculated value of the "correct" optimal top tax rate. As an extensive literature points out, there are many considerations that need to be taken into account to provide such a number. Instead, the point is that the existing literature has underemphasized a consideration that appears to have a first-order effect on the calculation. When the return to the discovery of new ideas is an important part of the compensation of top earners, the nonrivalry of these ideas means that distortions to innovation can have large effects on other people's incomes, and therefore on the optimal top tax rate itself.

A. Proofs

The Revenue-Maximizing Top Tax Rate in Equation (8). Starting from equation (7), note that $\frac{dY}{d\tau} = -\frac{dY}{d(1-\tau)}$ so that

$$
\begin{aligned}
(\rho - \bar{\rho}) &= \frac{dY}{d(1-\tau)} \frac{1-\tau}{Y} \cdot \frac{(1-\rho)\tau_0 + \rho\tau}{1-\tau} \\
\Rightarrow \quad \Delta\rho(1-\tau) &= \eta_{Y,1-\tau} \cdot [(1-\rho)\tau_0 + \rho\tau] \\
\Rightarrow \quad \Delta\rho - \eta_{Y,1-\tau}(1-\rho)\tau_0 &= (\Delta\rho + \eta_{Y,1-\tau}\rho)\tau \qquad \text{(A1)}
\end{aligned}
$$

which can be solved for τ to give the solution in the text:

$$\tau^*_{rm} = \frac{1 - \tau_0 \cdot \frac{1-\rho}{\Delta\rho} \cdot \eta_{Y,1-\tau}}{1 + \frac{\rho}{\Delta\rho}\eta_{Y,1-\tau}}.$$

Proof of Proposition 1. Maximizing Worker Welfare

The Lagrangian for this problem is

$$\mathscr{L} = \log(1-\tau_0) + \log Y + \lambda[\tau_0 Y + (\tau - \tau_0)(\rho Y - \bar{w} S_a) - \Omega Y]. \tag{A2}$$

The first order condition with respect to τ_0 is

$$\begin{aligned} \frac{1}{1-\tau_0} &= \lambda[Y - (\rho Y - \bar{w} S_a)] \\ &= \lambda Y[1 - (\rho - \bar{\rho})] \\ &= \lambda Y (1 - \Delta\rho). \end{aligned} \tag{A3}$$

The first order condition with respect to τ is

$$\begin{aligned} &\frac{1}{Y}\frac{\partial Y}{\partial \tau}\frac{\tau}{\tau} + \lambda[\tau_0 + (\tau - \tau_0)\rho - \Omega]\frac{\partial Y}{\partial \tau}\frac{\tau}{Y}\frac{Y}{\tau} + \lambda(\rho Y - \bar{w} S_a) = 0 \\ \Rightarrow \quad &\eta_{Y,\tau}\frac{1}{\tau} + \lambda Y[(1-\rho)\tau_0 + \rho\tau - \Omega]\eta_{Y,\tau}\frac{1}{\tau} + \lambda Y(\rho - \bar{\rho}) = 0. \end{aligned}$$

Now use the fact that $\eta_{Y,\tau} = -\eta_{Y,1-\tau} \cdot \frac{\tau}{1-\tau}$ and to simplify notation, let $\eta := \eta_{Y,1-\tau}$. Then, rewrite this FOC as

$$\begin{aligned} -\frac{\eta}{1-\tau} - \lambda Y[(1-\rho)\tau_0 + \rho\tau - \Omega]\frac{\eta}{1-\tau} + \lambda Y \Delta\rho = 0 \\ \Rightarrow \quad \eta\left(\frac{1}{\lambda Y} + (1-\rho)\tau_0 + \rho\tau - \Omega\right) = \Delta\rho(1-\tau) \end{aligned}$$

Substituting from (A3) for λY and rearranging gives

$$\begin{aligned} \eta[(1-\Delta\rho)(1-\tau_0) + (1-\rho)\tau_0 - \Omega] &= \Delta\rho(1-\tau) - \eta\rho\tau \\ &= \Delta\rho - \tau(\Delta\rho + \rho\eta). \end{aligned}$$

Finally, solve for τ:

$$\begin{aligned} \tau &= \frac{\Delta\rho - \eta[(1-\Delta\rho)(1-\tau_0) + (1-\rho)\tau_0 - \Omega}{\Delta\rho + \rho\eta} \\ &= \frac{1 - \eta\left[\frac{1-\rho}{\Delta\rho}\tau_0 + \frac{1-\Delta\rho}{\Delta\rho}(1-\tau_0) - \frac{\Omega}{\Delta\rho}\right]}{1 + \frac{\rho}{\Delta\rho}\eta} \end{aligned} \tag{A4}$$

which proves the proposition. QED.

Proof of Proposition 2. Maximizing Social Welfare

The problem is to choose τ_0 and τ to maximize

$$\overline{SWF} = \log Y + \ell \log(1-\tau_0) + s(1-\alpha)\log(1-\tau) - s\zeta(1-\tau)^{\alpha} \tag{A5}$$

subject to the government budget constraint (GBC):

$$\tau_0 Y + (\tau - \tau_0)(\rho Y - \bar{w} S_a) = \Omega Y. \tag{A6}$$

Let λ be the Lagrange multiplier on the GBC, denoting the shadow value of an extra unit of tax revenue. Then the first order condition with respect to τ_0 is

$$\begin{aligned}\frac{\ell}{1-\tau_0} &= \lambda[Y - (\rho Y - \bar{w} S_a)] \\ &= \lambda Y[1 - (\rho - \bar{\rho})] \\ &= \lambda Y(1 - \Delta\rho). \end{aligned} \tag{A7}$$

The first order condition with respect to τ is

$$\begin{aligned}\frac{1}{Y}\frac{\partial Y}{\partial \tau}\frac{\tau}{\tau} - \frac{s(1-\alpha)}{1-\tau} + \alpha\zeta(1-\tau)^{\alpha-1} + \lambda[\tau_0 + (\tau-\tau_0)\rho - \Omega]\frac{\partial Y}{\partial \tau}\frac{\tau}{Y}\frac{Y}{\tau} \\ + \lambda(\rho Y - \bar{w} S_a) = 0 \\ \eta_{Y,\tau}\frac{1}{\tau} - \frac{s(1-\alpha)}{1-\tau} + \alpha\zeta(1-\tau)^{\alpha-1} + \lambda Y[(1-\rho)\tau_0 + \rho\tau - \Omega]\eta_{Y,\tau}\frac{1}{\tau} \\ + \lambda Y(\rho - \bar{\rho}) = 0.\end{aligned}$$

Now use the fact that $\eta_{Y,\tau} = -\eta_{Y,1-\tau}\cdot\frac{\tau}{1-\tau}$ and to simplify notation, let $\eta := \eta_{Y,1-\tau}$. Then rewrite this FOC as

$$\begin{aligned}-\frac{\eta}{1-\tau} - \frac{s(1-\alpha)}{1-\tau} + \frac{\alpha\zeta(1-\tau)^{\alpha}}{1-\tau} - \lambda Y[(1-\rho)\tau_0 + \rho\tau - \Omega]\frac{\eta}{1-\tau} \\ + \lambda Y\Delta\rho = 0.\end{aligned}$$

Now use the FOC with respect to τ_0 in equation (A7) to note that $\lambda Y = \frac{\ell}{(1-\tau_0)(1-\Delta\rho)}$.

Substituting this in above and using some algebra gives

$$
\begin{aligned}
-\eta - s(1-\alpha) + \alpha s\zeta(1-\tau)^{\alpha} + \frac{\eta\ell}{(1-\tau_0)(1-\Delta\rho)}[(1-\rho)(1-\tau_0) + \rho(1-\tau) - (1-\Omega)]& \\
+ \ell\frac{\Delta\rho}{1-\Delta\rho}\frac{1-\tau}{1-\tau_0} = 0&
\end{aligned}
$$

which can be simplified further to

$$
\begin{aligned}
-\eta - s(1-\alpha) + \alpha s\zeta(1-\tau)^{\alpha} + \frac{\eta\ell}{1-\Delta\rho}\left[1-\rho-\frac{1-\Omega}{1-\tau_0}\right]& \\
+ \ell\frac{\Delta\rho}{1-\Delta\rho}\frac{1-\tau}{1-\tau_0}\left(1+\frac{\eta\rho}{\Delta\rho}\right) = 0.&
\end{aligned}
$$

Next, rearrange this equation and write in terms of the keep rate κ and κ_0 as

$$
\begin{aligned}
\eta + s(1-\alpha) - \frac{\eta\ell(1-\rho)}{1-\Delta\rho} &= \alpha s\zeta\kappa^{\alpha} + \ell\frac{\Delta\rho}{1-\Delta\rho}\frac{\kappa}{\kappa_0}\left(1+\frac{\eta\rho}{\Delta\rho}\right) - \frac{\eta\ell}{1-\Delta\rho}\frac{1-\Omega}{\kappa_0} \\
&= \alpha s\zeta\kappa^{\alpha} + \frac{1}{\kappa_0}\frac{\ell}{1-\Delta\rho}\left[\kappa(\Delta\rho+\eta\rho) - \eta(1-\Omega)\right] \\
&= \alpha s\zeta\kappa^{\alpha} + \frac{1}{\kappa_0}\frac{\ell}{1-\Delta\rho}\left[\kappa\Delta\rho + \eta[\rho\kappa - (1-\Omega)]\right] \qquad \text{(A8)}
\end{aligned}
$$

Now recall that the government budget constraint can be manipulated:

$$
\begin{aligned}
(1-\Delta\rho)\kappa_0 + \Delta\rho\kappa = 1-\Omega& \qquad \text{(A9)} \\
\Rightarrow\ (\rho - \bar{\rho}\kappa - (1-\Omega) = -(1-\Delta\rho)\kappa_0& \\
\Rightarrow\ \rho\kappa - (1-\Omega) = \bar{\rho}\kappa - (1-\Delta\rho)\kappa_0.& \qquad \text{(A10)}
\end{aligned}
$$

Substituting this expression into (A8) gives

$$
\begin{aligned}
\eta + s(1-\alpha) - \frac{\eta\ell(1-\rho)}{1-\Delta\rho} &= \alpha s\zeta\kappa^{\alpha} + \frac{1}{\kappa_0}\frac{\ell}{1-\Delta\rho}\left[\kappa\Delta\rho + \eta[\bar{\rho}\kappa - (1-\Delta\rho)\kappa_0]\right] \\
&= \alpha s\zeta\kappa^{\alpha} + \frac{\kappa}{\kappa_0}\frac{\ell}{1-\Delta\rho}[\Delta\rho + \eta\bar{\rho}] - \eta\ell
\end{aligned}
$$

Rearranging and collecting terms:

$$\begin{aligned}\alpha s\zeta\kappa^{\alpha}+\frac{\kappa}{\kappa_0}\frac{\ell}{1-\Delta\rho}[\Delta\rho+\eta\bar{\rho}] &= \eta+s(1-\alpha)+\eta\ell\left(1-\frac{1-\rho}{1-\Delta\rho}\right)\\ &= \eta+s(1-\alpha)+\eta\ell\left(\frac{\bar{\rho}}{1-\Delta\rho}\right)\\ &= s(1-\alpha)+\eta\left(1+\frac{\bar{\rho}\ell}{1-\Delta\rho}\right)\end{aligned}$$

Together with the government budget in (A9), this proves the proposition. QED.

B. Deriving the Social Rate of Return

Following Jones and Williams (1998), begin with a discrete-time version of our model, where the production function for ideas is

$$\Delta A_{t+1} = G(\tilde{S}_t, A_t, B_t) = \tilde{S}_t^{\lambda} A_t^{\phi_a} B_t^{\alpha} \tag{A11}$$

and the production function for final output is $Y = A^{\sigma}\tilde{M}^{\psi}L^{1-\psi}$.

Consider the following variation: reduce consumption by one unit today, and use the proceeds to hire additional entrepreneurs to produce ideas. Then tomorrow, reduce the number of entrepreneurs by the appropriate amount to leave the time path of A_t unchanged from period $t+2$ onward. Denote the additional consumption can be gained by this variation by $\tilde{r}$, which is the social rate of return.

Mathematically, this return is then given by

$$1+\tilde{r} = \frac{1}{\tilde{w}_t}\cdot\frac{\partial G_t}{\partial \tilde{S}_t}\cdot\frac{\partial Y_{t+1}}{\partial A_{t+1}} + \frac{\tilde{w}_{t+1}}{\tilde{w}_t}\cdot\frac{\partial G_t/\partial \tilde{S}_t}{\partial G_{t+1}/\partial \tilde{S_{t+1}}}\left(1+\frac{\partial G_{t+1}}{\partial A_{t+1}}\right)$$

where $\tilde{w}_t$ is that shadow price that converts one unit of research effort into one unit of consumption. Substituting for the derivatives, we have

$$\begin{aligned}1+\tilde{r} &= \frac{1}{\tilde{w}_t}\cdot\lambda\frac{\Delta A_{t+1}}{\tilde{S}_t}\cdot\sigma\frac{Y_{t+1}}{A_{t+1}} + \frac{\tilde{w}_{t+1}}{\tilde{w}_t}\cdot\frac{\lambda\Delta A_{t+1}/\tilde{S}_t}{\lambda\Delta A_{t+2}/\tilde{S}_{t+1}}\left(1+\phi_a\frac{\Delta A_{t+2}}{A_{t+1}}\right)\\ &\approx \lambda\sigma g_A\cdot\frac{Y_t}{\tilde{w}_t\tilde{S}_t}+1+g_{\tilde{w}}+g_{\tilde{S}}+\phi_a g_A - g_A\end{aligned}$$

where the approximation comes from ignoring second order terms (products of growth rates).

Finally, we need to consider $\tilde{w}$. This is the rate at which one unit of research input can be converted into consumption. Recall $\tilde{S}_t = \bar{e}\bar{z}S_t$, and the key distortion in the economy is on research effort. In particular, the researcher's first order condition is

$$-\frac{1}{z}\cdot\frac{u_e}{u_c} = w_s(1-\tau) \equiv \tilde{w}. \tag{A12}$$

That is, the rate at which entrepreneurs trade off a unit of research input for consumption is $w_s(1-\tau)$. Substituting this into the social rate of return equation, we have

$$\tilde{r} \approx \lambda\sigma g_A \cdot \frac{1}{\rho_s(1-\tau)} + g_Y - (1-\phi_a)g_A. \tag{A13}$$

Collecting terms and noting that $\gamma \equiv \lambda\sigma/(1-\phi_a)$ and $g_y = \sigma g_A$ yields the result in the text in equation (30).

References

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, *60* (2), 323–351.

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde, "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth," December 2016. University of Chicago manuscript.

— , John Grigsby, Tom Nicholas, and Stefanie Stantcheva, "Taxation and Innovation in the 20th Century," Working Paper 24982, National Bureau of Economic Research September 2018.

— , Salome Baslandze, and Stefanie Stantcheva, "Taxation and the International Mobility of Inventors," *American Economic Review*, October 2016, *106* (10), 2930–81.

Ales, Laurence and Christopher Sleet, "Taxing Top CEO Incomes," *American Economic Review*, November 2016, *106* (11), 3331–66.

Badel, Alejandro and Mark Huggett, "Taxing Top Earners: A Human Capital Perspective," Working Papers, Georgetown University, Department of Economics September 2015.

— and — , "The Sufficient Statistic Approach: Predicting the Top of the Laffer Curve," *Journal of Monetary Economics*, 2017, *87* (C), 1–12.

Bell, Alexander M., Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," Working Paper 24062, National Bureau of Economic Research November 2017.

__ , __ , __ , __ , and __ , "Do Tax Cuts Produce More Einsteins? The Impacts of Financial Incentives vs. Exposure to Innovation on the Supply of Inventors," Working Paper 25493, National Bureau of Economic Research January 2019.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," 2019. Stanford University manuscript.

__ , Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 07 2013, *81* (4), 1347–1393.

Budish, Eric, Benjamin N. Roin, and Heidi Williams, "Do firms underinvest in long-term research? Evidence from cancer clinical trials," *The American Economic Review*, 2015, *105* (7), 2044–2085.

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

Diamond, Peter and Emmanuel Saez, "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 2011, *25* (4), 165–90.

Gabaix, Xavier and Augustin Landier, "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 2008, *123* (1), 49–100.

Grossman, Gene M. and Elhanan Helpman, *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press, 1991.

Hall, Bronwyn H., Jacques Mairesse, and Pierre Mohnen, "Chapter 24 - Measuring the Returns to R&D," in Bronwyn H. Hall and Nathan Rosenberg, eds., *Handbook of the Economics of Innovation, Volume 2*, Vol. 2 of *Handbook of the Economics of Innovation*, North-Holland, 2010, pp. 1033 – 1082.

Hall, Robert E. and Susan E. Woodward, "The Burden of the Nondiversifiable Risk of Entrepreneurship," *American Economic Review*, June 2010, *100* (3), 1163–94.

Howell, Sabrina T., "Financing Innovation: Evidence from R&D Grants," *American Economic Review*, April 2017, *107* (4), 1136–64.

Internal Revenue Service, "Statistics of Income 2015: Individual Income Tax Returns," Technical Report September 2017. Publication 1304.

— , "SOI Tax Stats - Historical Table 23," Technical Report 2018.

Jaimovich, Nir and Sergio Rebelo, "Nonlinear Effects of Taxation on Growth," *Journal of Political Economy*, 2017, *125* (1), 265–291.

Jones, Charles I., "Growth and Ideas," in Philippe Aghion and Steven A. Durlauf, eds., *Handbook of Economic Growth*, New York: North Holland, 2005, pp. 1063–1111.

— and John C. Williams, "Measuring the Social Return to R&D," *Quarterly Journal of Economics*, November 1998, *113* (4), 1119–1135.

Kindermann, Fabian and Dirk Krueger, "High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk," Working Paper 20601, National Bureau of Economic Research October 2014.

Lockwood, Benjamin, Charles Nathanson, and E. Glen Weyl, "Taxation and the Allocation of Talent," *Journal of Political Economy*, 2017, *125* (5), 1635–1682.

Lucking, Brian and Nicholas Bloom, "Have R&D Spillovers Changed over Time?," 2017. Stanford University manuscript.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38*, 175–208.

Moretti, Enrico and Daniel J. Wilson, "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists," *American Economic Review*, July 2017, *107* (7), 1858–1903.

Piketty, Thomas and Emmanuel Saez, "Income Inequality In The United States, 1913–1998," *Quarterly Journal of Economics*, February 2003, *118* (1), 1–39.

— , — , and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, February 2014, *6* (1), 230–71.

Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, October 1990, *98* (5), S71–S102.

Rothschild, Casey and Florian Scheuer, "A Theory of Income Taxation under Multidimensional Skill Heterogeneity," manuscript 5165, Stanford University 2015.

Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin, "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," NBER Working Papers 22646, National Bureau of Economic Research, Inc September 2016.

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

Scheuer, Florian and Iván Werning, "The Taxation of Superstars," *Quarterly Journal of Economics*, 2017, *132* (1), 211–270.

Stokey, Nancy L. and Sergio Rebelo, "Growth Effects of Flat-Rate Taxes," *Journal of Political Economy*, June 1995, *103*, 519–550.

Tervio, Marko, "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 2008, *98* (3), 642–668.

Wolfe, Raymond M., "Business R&D Performance in the United States Tops $300 Billion in 2012," *NCSES Info Brief, NSF*, October 2014, *15-303*.