

Taxing Top Incomes in a World of Ideas

Charles I. Jones*

Stanford GSB and NBER

September 7, 2021 — Version 4.0

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 (though people can earn high incomes in other ways as well), 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. These conditions lead to a new force affecting the optimal top tax rate: by slowing the creation of new ideas that raise 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.

*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, the IMF, the NBER Summer Institute, NYU, the St. Louis Fed, Stanford, Vanderbilt, and the World Bank for helpful suggestions. Jack Hirsh and Andrés Yany provided expert research assistance.

1. Introduction

The classic tradeoff in the optimal income tax literature is between redistribution and the incentive effects that determine the “size of the pie.” In much of the literature — starting with Mirrlees (1971), Diamond (1998), Saez (2001), and Diamond and Saez (2011) — these “size of the pie” effects were relatively limited.¹ In particular, when a top earner reduces her effort because of taxes, that reduces her income but may have no or only modest effects on the incomes of everyone else in the economy. Later research relaxed this restriction, incorporating complementary inputs such as the accumulation of human and physical capital or through externalities; recent examples include Rothschild and Scheuer (2013, 2014), Badel and Huggett (2017) and Lockwood, Nathanson and Weyl (2017).

However, what is in some ways the most natural effect on the size of the pie has not been adequately explored. In the idea-based growth theory of Romer (1990), Aghion and Howitt (1992), and Grossman and Helpman (1991), the enormous increase in living standards over the past century is the result of the discovery of new ideas 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.

A standard policy implication in the R&D literature is that it may be optimal to subsidize formal R&D, and one could imagine subsidizing research but taxing top incomes so as 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 world and is readily subsidized with government funding. Applied research turns these fundamental truths into consumer products or firm-level process innovations. This is the realm of entrepreneurs and may not be readily subsidized as formal R&D. Think about the creation of Walmart, Amazon, or any other new firm; organizational innovations in health care and education; the latest Google search software; 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 innovative effort. For example, around 70% of measured R&D

¹Of course, the theoretical literature noted that these effects could be important; see, e.g., Stiglitz (1982).

occurs in manufacturing, and in 2012, only 15% of workers were employed by firms that conducted any official R&D (Wolfe, 2014). According to their 2018 corporate filings, Walmart and Goldman-Sachs reported zero spending on R&D.

The innovation that occurs beyond formal R&D may be distorted by the tax system. High incomes are a prize that partly motivates entrepreneurs to turn basic insights into a product or process that ultimately benefits consumers. High marginal tax rates deter this effort and therefore reduce innovation and overall GDP.

Taking this force into account is important quantitatively. For example, consider raising the top marginal tax rate from 50% to 75%. In the United States, the share of income that this top 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 below, this increase in the top tax rate reduces innovation and lowers GDP per person in the long run by around 7 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 7% decline in their incomes is not offset by the 5% increase from redistribution. In other words, raising the top marginal rate from 50% to 75% reduces social welfare in this example.

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 revenue-maximizing top tax rate that ignores innovation is 91%. In contrast, the rate that incorporates innovation and maximizes the welfare of workers is much lower: the benchmark value is just 9%, while alternative parameter values give a range of -26% to 50%. Subsidizing the discovery of new ideas through low tax rates may be as effective as redistribution in raising worker welfare.

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 effect 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). Section 3 then considers the tax scheme that maximizes the welfare of the “bottom 90%.” In Diamond and Saez (2011), this is the same as revenue maximization, but that is not the case with spillovers: the planner cares about distorting the creation of ideas not merely because it affects the revenue that can be obtained from regular workers, but because it affects their consumption. This distinction matters quantitatively, reducing the top tax rate from 65% to 9% in the baseline. Section 4 goes further and finds the tax system that maximizes utilitarian social welfare. Section 5 discusses additional results, including empirical evidence on growth and top income taxation. Finally, Section 6 builds the full dynamic growth model that nests the model used in the bulk of the paper.

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. Rothschild and Scheuer (2013, 2014), Lockwood, Nathanson and Weyl (2017), Badel, Huggett and Luo (2020), and Sachs, Tsyvinski and Werquin (2020) allow for imperfect substitution in production, increasing returns, general externalities, or human capital spillovers so that the decisions of top earners can affect the wages of others in the economy. These very general setups, especially in Rothschild and Scheuer (2014), in some sense nest the specific framework developed below. The advantage of being specific, however, is that it allows us to leverage the progress that has been made in the endogenous growth literature to inform the calibration of the spillovers that are involved. The papers are complements, not substitutes.

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) consider the revenue-maximizing top tax rate when incomes at lower brackets can be affected by the top rate in a generalized fashion. 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 in which 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. I instead focus explicitly on the research dimension, where leveraging an extensive literature allows us to make more precise statements about this particular externality.

Piketty, Saez and Stantcheva (2014) and Rothschild and Scheuer (2016) 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 optimal top income tax rate is higher. Akcigit, Baslandze and Stantcheva (2016) 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 superstar effects increase the effective behavioral elasticity and therefore lower the optimal top tax rate.

Akcigit, Grigsby, Nicholas and Stantcheva (2021) 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 geographic relocation of innovation from one state to another, and therefore are an upper bound on the net effect most relevant in this paper. Importantly, they also report various estimates of the net effect at the individual inventor level controlling for inventor fixed effects and state fixed effects. These elasticities

are large and range from 0.6 to 0.9 for patents and citations.

Another closely related paper is Jaimovich and Rebelo (2017), which studies 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, 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.

The present paper differs in several ways. First, its 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.

A final important distinction raises a broader issue. In particular, 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. Here, instead, we explore a model in which policies have long-run level effects, and the size of these level effects is governed by a key parameter influencing the top tax rate.

To what extent is this alternative growth approach appropriate? First and foremost, the topic of this paper requires an idea-driven growth model, that is, something in the tradition of Romer (1990) and Aghion and Howitt (1992). Within this broad class of

models, the “semi-endogenous” approach taken here is in some sense conservative. With growth effects, it is as if the level effects are infinitely large, whereas in our calibrations we consider values between 1/8 and 1. (It’s slightly more complicated than that because with $\gamma = \infty$ one would integrate to get present values of the growth effects, but the spirit of the point is correct.) As we show below, raising this parameter only reinforces our findings. In other words, the basic force responsible for the results of this paper is not limited to semi-endogenous growth models but instead is present in every idea-driven growth model in the literature.

Models like the one in this paper have been used recently to study quantitatively a range of phenomena. Jones (2022) employs such a model to account for economic growth over the past 75 years, discusses the semi-endogenous growth literature more generally, and highlights the implications for the future of economic growth. Peters and Walsh (2021) explore the consequence of slowing population growth for firm dynamics and growth. Bloom, Jones, Van Reenen and Webb (2020) consider a broad array of micro evidence on idea production functions and suggest that the evidence uniformly supports the “semi-endogenous” specification. Arkolakis, Lee and Peters (2020) highlight the importance of European immigration for U.S. economic growth between 1880 and 1920. Buera and Oberfield (2020) study technology diffusion across countries. Atkeson and Burstein (2019) explore the short-run and long-run impact of policy changes, including a careful consideration of transition dynamics. Ngai and Samaniego (2011) study heterogeneity in idea production functions across industries.²

On the response of inventive effort to taxes, Bell, Chetty, Jaravel, Petkova and Van Reenen (2019b) 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 (2019a) 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

²The older literature includes Jones (1995), Kortum (1997), and Segerstrom (1998), among others.

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.

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 75% 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.

This paper focuses on labor income. But the examples of Amazon and Uber raise a question: how much of the return to entrepreneurship is ordinary income versus capital gains? Smith, Yagan, Zidar and Zwick (2019) document that the typical top earner derives most of her income from entrepreneurial business income (e.g. pass-through profits) that is taxed as ordinary income rather than from financial capital, generally supporting the approach taken here. Nevertheless, it would be interesting to study the extent to which capital gains partly reflect a return to innovation. The existing literature on the taxation of capital income is undergoing a revision. Chamley (1986) and Judd (1985) provide famous arguments that the long-run tax rate on capital should be zero. Bassetto and Benhabib (2006) and Straub and Werning (2020) suggest that the foundations for this result are not nearly as strong as previously thought. The mechanism in this paper is very different and could be usefully incorporated into a study of capital taxation.

Finally, another important consideration is other policies that may be 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 the balanced growth path of a full dynamic model. The model and notation are similar to Jones (2005).

Consider a closed economy containing 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 , and managers, M . This setup therefore abstracts from occupational choice, for the reasons discussed above. Entrepreneurs and managers 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, and then applied research turns these discoveries 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.³

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

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

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

³For a recent application of this distinction in the growth literature, see Akcigit, Hanley and Serrano-Velarde (2020).

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; this is the key relationship we exploit in the next two equations. 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 as a starting point for intuition, that tradeoff is worth making. Along the balanced growth path, the stock of applied ideas is

$$A = \nu_a \mathbb{E}[\theta e] S_a B^\beta. \quad (2)$$

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 θ , 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 θ , so we can eventually write this as just $\mathbb{E}[\theta e] = \bar{\theta} \bar{e}$.) Top taxation matters in this setting because it distorts the effort of entrepreneurs. For the first half of this paper, we follow Saez (2001) and Diamond and Saez (2011) and treat the elasticity of effort with respect to taxation as a parameter; later we will model this formally. Finally, 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:

$$B = \nu_b S_b. \quad (3)$$

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

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

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 (50) 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.

The increasing returns in this setup derives from the nonrivalry of ideas. 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.

Romer (1990) explains how nonrivalry gives rise to increasing returns to scale. The *standard replication argument* is a fundamental justification for constant returns. 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 is constant returns to scale in the rival inputs (the factory, workers, and materials), there is 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, managers and workers) 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.

2.1 Why the Nonrivalry of Ideas Matters for Taxation

Managers are included in the model to capture the traditional Diamond and Saez (2011) tradeoff. In particular, we assume $\tilde{M} := \mathbb{E}[\theta e]M$, where M is the exogenous supply

of managers, θ denotes their (heterogeneous) talent, and e denotes their effort. Total output in equation (4) then becomes

$$Y = \left(\nu \mathbb{E}[\theta e] S_a S_b^\beta \right)^\gamma (\mathbb{E}[\theta e] M)^\psi L^{1-\psi} \quad (5)$$

We abuse notation slightly in using the same notation for the talent and effort of managers that we used for entrepreneurs.

Top income taxation therefore affects overall GDP in two ways, suggested by the presence of the two effort terms $\mathbb{E}[\theta e]$ in equation (5). There is a “traditional” effect working through managers, governed by the ψ exponent, as well as the “idea” effect working through applied researchers, governed by the γ exponent.

This view highlights why the nonrivalry of ideas is so important. Managers enter the production function *inside* the constant returns to scale. The exponent on managers and workers (and capital if we added it to the model) must sum to one, so ψ essentially “competes” with the factor exponents on capital and labor. This means that plausible calibrations of ψ in the literature are typically small, around 0.15 or less, which limits the extent to which the taxation of managers can reduce GDP.⁴ In contrast, the idea effect enters outside the constant returns portion of the production function because of nonrivalry. The parameter γ is not “part of a sum that has to equal one” and therefore can potentially be large. This is a key part of the intuition for why the nonrivalry of ideas can substantially affect top income taxation.

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 key 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

⁴For example, see Lucas (1978) and Garicano, Lelarge and Van Reenen (2016).

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 $Y = Const \cdot b^{\beta\gamma}(1-b)^{1-\psi}$ and the output-maximizing allocation of basic research is $\frac{b}{1-b} = \frac{\beta\gamma}{1-\psi}$, independent of the effort choice.⁵

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

Following a tradition in public economics — also followed by Saez (2001), Diamond and Saez (2011), and many others — one can make useful progress by keeping the model of effort unspecified for now and treating the elasticity of effort with respect to taxes as a parameter.

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 (θe) 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 then⁶

$$T = \underbrace{\tau_0(wL + w_b S_b + w_s \mathbb{E}[\theta e] S_a + w_m \mathbb{E}[\theta e] M)}_{\text{all income pays } \tau_0} + \underbrace{(\tau - \tau_0)[(w_s \mathbb{E}[\theta e] - \bar{w}) S_a + (w_m \mathbb{E}[\theta e] - \bar{w}) M]}_{\text{income above } \bar{w} \text{ pays an additional } \tau - \tau_0}$$

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

$$\frac{w_s \mathbb{E}[\theta e] S_a}{Y} = \rho_s \quad \text{and} \quad \frac{w_m \mathbb{E}[\theta e] M}{Y} = \rho_m. \quad (6)$$

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, $1 - \rho = \frac{wL + w_b S_b}{Y}$, is paid to workers and basic researchers.

⁵A similar argument explains why adding capital to the Cobb-Douglas setup here would not affect the basic results: once the steady-state capital-output ratio is replaced by a term proportional to the investment rate, an equation such as (4) will still describe the steady state of the model.

⁶For simplicity, we assume that the lower support of the talent distribution is such that all entrepreneurs and managers face the top marginal tax rate.

With these definitions, total tax revenue can be rewritten as

$$T = \tau_0 Y(\tau) + (\tau - \tau_0)[\rho Y(\tau) - \bar{w}(S_a + M)]. \quad (7)$$

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 .

Some Intuition. Now is a good time to develop some additional intuition. Recall that total output in the economy can be expressed as

$$Y = \left(\nu \mathbb{E}[\theta e] S_a S_b^\beta \right)^\gamma (\mathbb{E}[\theta e] M)^\psi L^{1-\psi}.$$

Equation (6) earlier says that entrepreneurs and managers get paid ρ_s and ρ_m as shares of GDP. There is a fundamental tension here. Efficiency argues for paying entrepreneurs and managers factor shares equal to the exponents in the production function, γ and ψ . But in an economy with increasing returns, Euler's theorem ensures that this cannot happen for all inputs: the exponents sum to greater than one.

This intuition can be developed further. Entrepreneurs are paid the share ρ_s and at the margin receive compensation tied to the after-tax share $\rho_s(1 - \tau)$. In Section 5.3, we study the social rate of return to research in this economy and show that equating the social return to the real interest rate — one version of efficiency — requires $\rho_s(1 - \tau) \approx \gamma$, consistent with the intuition given in the previous paragraph: efficiency points to entrepreneurs receiving the share γ of output. Two points then follow. First, even if $\rho_s = \gamma$, there is still an efficiency-equity tradeoff. Taxing top incomes will introduce a wedge into the efficiency margin. Second, to the extent that ρ_s is less than γ — as is likely because of increasing returns and Euler's theorem — an envelope-type result will not apply and the efficiency margin may be even more sensitive to top taxation. A large literature discussed in Section 5.3 argues that social rates of return to research are high, so this underinvestment scenario is plausible.

Alternatively, one can solve for the first-best allocation in this economy using a social welfare function and a utility cost of effort; we do this formally in Section 5.4. First-best entrepreneurial effort requires a tax that sets $s(1 - \tau) = \gamma$, where s is the

population share of entrepreneurs. This expression nicely illustrates one of the points in the introduction: the fact that a relatively small share of the population, s , drives economic growth via the nonrivalry of ideas, γ , constrains the top tax rate, τ . For example, when $s < \gamma$, achieving first-best effort requires a *negative* top tax rate. The social planner trades off this efficiency consideration with a concern for equity. But this makes clear why the importance of ideas relative to the size of the population generating them (either in number or in income share) matters for the top rate.

The Revenue-Maximizing Top Rate. The top tax rate that maximizes revenue can be found by setting the derivative of equation (7) equal to zero. The key tradeoff is that a higher top rate reduces Y via the effort terms in equation (5).

This first-order condition for maximizing revenue can be written as

$$\underbrace{(\rho - \bar{\rho}) Y}_{\text{mechanical gain}} + \underbrace{\frac{dY}{d\tau} \cdot [(1 - \rho)\tau_0 + \rho\tau]}_{\text{behavioral loss}} = 0 \quad (8)$$

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^* . Defining $\eta_{Y,1-\tau} \equiv \frac{d \log Y}{d \log(1-\tau)}$ and $\Delta\rho \equiv \rho - \bar{\rho}$, we have⁷

$$\tau_{rm}^* = \frac{1 - \tau_0 \cdot \frac{1-\rho}{\Delta\rho} \cdot \eta_{Y,1-\tau}}{1 + \frac{\rho}{\Delta\rho} \eta_{Y,1-\tau}}, \quad (9)$$

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 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. The second difference is in the nature of the elasticity, $\eta_{Y,1-\tau}$. In Diamond and Saez, the fundamental elasticity

⁷A more complete derivation is available in the [appendix](#).

(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.

Formulas very similar to (9) have been derived in the recent literature, including Rothschild and Scheuer (2014), Lockwood, Nathanson and Weyl (2017), and Badel, Huggett and Luo (2020), among others. Those papers allow for general externalities or human capital “spillovers” while Badel and Huggett (2017) derives such a formula in a very general setting. What differs here is that we focus explicitly on the effects that arise in idea-based growth models. The entirety of the growth model is embedded in $\eta_{Y,1-\tau}$ and this allows us to make progress in two ways. First, we leverage the extensive literature on growth and R&D to help us understand this elasticity. We therefore obtain stronger intuitions and quantitative estimates. Second, having a growth model explicitly in mind makes clear that revenue maximization and welfare maximization are not identical, and many of our insights come from thinking about welfare instead of revenue; this is pursued further in Sections 3 and 4.

It is helpful to keep in mind the following intuition for the various terms that appear in the solutions, such as in equation (9). The intuition can typically be found by looking back at equation (8). 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 base to which τ_0 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 the ratio, the higher the income above the top rate, resulting in a higher revenue-maximizing top tax rate.

Calibration. To go further with this equation, we need to choose values for various parameters, starting with those that determine $\eta_{Y,1-\tau}$. From the production function in (5), we have

$$\eta_{Y,1-\tau} = \gamma \zeta^s + \psi \zeta^m \tag{10}$$

where ζ is defined to be the elasticity of $\mathbb{E}[\theta e]$ with respect to $1 - \tau$, and we now explicitly consider the possibility that entrepreneurs (ζ^s) have a different elasticity than managers (ζ^m). Each of the parameters in this expression merits discussion.

The ζ 's are the key elasticities that enter many optimal tax formulas: what is the elasticity of a top earner's income with respect to the take-home rate? For now, we treat each ζ as a parameter and calibrate it based on the empirical literature to a range of values. This approach is very common in the tax literature; for example, Diamond and Saez (2011) consider a low value of $\zeta = 0.2$ and a higher value of $\zeta = 0.5$.⁸

A related question about this elasticity is whether it is a compensated elasticity or an uncompensated elasticity. The public economics literature often emphasizes compensated elasticities, i.e. holding consumption constant. Saez (2001) is clear that both elasticities can matter in general; the simple Diamond and Saez formula given earlier assumes no income effects so that these elasticities are the same. Later, in Section 4.1, we develop a model in which it turns out to be the uncompensated elasticity — taking into account income effects from changes in consumption — that matters, so this is what we calibrate to here. This is not surprising given that the paper studies general equilibrium effects that lead to changes in incomes throughout the economy. This force pushes for a lower elasticity: as top earners are taxed more, the income effect means they consume less of all goods, including leisure, and therefore work more.

With these factors in mind, we consider values for ζ^m — the elasticity for top earners who are not entrepreneurs — of $\{0.1, 0.2, 0.3\}$. These are appropriately smaller than the compensated elasticity range of $(0.2, 0.6)$ reported by Chetty (2012) while still capturing a wide range of values, consistent with the literature's general uncertainty about this elasticity.

Next, there is the question of whether the elasticity for entrepreneurs is the same as that of a typical top earner: how different are ζ^s and ζ^m ? Many innovative efforts fail to produce a substantial return, so a new entrepreneur's chance of facing the top tax rate could be low, for example, leading to a much smaller elasticity. The literature

⁸Saez, Slemrod and Giertz (2012) survey the evidence on this elasticity and document the substantial uncertainty that exists about its average value. Chetty (2012) attempts to reconcile the evidence surveyed by Saez, Slemrod, and Giertz (2012) using the presence of small frictions that distort individual behavior. He finds that evidence from a wide range of studies using different methods is consistent with a 0.33 intensive margin estimate for the compensated labor supply elasticity, with a 95% confidence interval of $(0.23, 0.61)$.

review above discussed broadly a range of evidence suggesting that innovation responds to taxes. We now consider this evidence more carefully.

Akcigit, Grigsby, Nicholas and Stantcheva (2021) is currently the most thorough paper estimating this elasticity for inventors, and they robustly find high elasticities. For example, using panel data for U.S. states since 1940, they find “macro” elasticities of invention (patents, citations, and number of inventors) with respect to top marginal keep rates ranging from 0.5 to 1.5, or even higher. In many models, income would be proportional to idea production, so it is plausible to consider these as the elasticity of effort with respect to $1 - \tau$.

To the extent that people leave states with high top tax rates and move to states with low tax rates, mobility will cause these estimates to overstate the true elasticity of overall invention. After working hard to control for mobility, Akcigit et al. conclude “the majority of the macro effect of personal taxation appears to result from reduced innovation at the individual level rather than through shifting the location of innovation from one state to another.” That is, their evidence overall supports high values for individual elasticities.

Most of the direct evidence on individual elasticities in the published paper of Akcigit et al. concerns the response to inventors’ own marginal tax rates rather than to top marginal rates. However, Table C.37 from their online appendix provides the estimates we need. Consistent with the intuition provided above, the elasticity for “regular inventors” of individual inventive effort to the marginal rate at the 90th percentile of the income distribution is noticeably smaller than the other estimates in the paper, on the order of 0.18 or lower. As suggested by Bell, Chetty, Jaravel, Petkova and Van Reenen (2019b), most inventors are unlikely to ever face the top rate, so it matters less to them. Importantly, however, the elasticity for star inventors (those in the top 10% of the cumulative patent distribution) is quite high, on the order of 0.8 or 0.9. This leads to average elasticities — averaging across “regular” inventors with low elasticities and star inventors with high elasticities — that are substantial. For example, the overall elasticity of patents by individuals to $1 - \tau$ is 0.3 and the elasticity of citations is 0.15. Appendix B develops an extension of our model with star and regular inventors confirming that it is the weighted average of the tax elasticities of the different types that matters, where the weights reflect the share of total ideas produced. That analysis

suggests even slightly higher elasticities, such as 0.24 to 0.40.

Taken together, this evidence supports the following approach. While there remains uncertainty about the right values for ζ^s and future research will surely want to provide better estimates, all the evidence we know of suggests that 0.1 to 0.3 — the range indicated by the literature for top earners in general — is also reasonable for inventors. In the interest of clarity, we will therefore assume $\zeta^s = \zeta^m \equiv \zeta$ in what follows, and consider values of ζ in $\{0.1, 0.2, 0.3\}$.

The parameter ψ summarizes 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.

While there is no consensus on the exact value of γ in the growth literature, there is a limited range of plausible values. As shown in the full model in Section 6 — see especially equation (52) — along a balanced growth path, the growth rate of income per person satisfies:

$$g_y = \gamma(1 + \beta)g_S. \quad (11)$$

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 2.⁹

In the model, we need to separate γ from $1 + \beta$ because only applied (not basic) researchers have their effort distorted by taxes. 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. Given that we

⁹The degree of increasing returns to scale resulting from the nonrivalry of ideas has not been estimated precisely in the literature. Jones (2002) reports estimates using time series methods of between 0.17 and 0.32. Arkolakis, Lee and Peters (2020) use European immigration to the U.S. between 1880 and 1920 and estimate a range of values of 0.7 to 1.3. Peters (2019) uses the pseudo-random settlement of East Germans into West Germany after World War II and finds a value of 0.89.

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 $\Delta\rho \approx .41 \times 13\% \approx 5.5\%$. 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.

Results. 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 (9).

Several findings stand out. First, using the Diamond and Saez formula, the baseline parameter values deliver the familiar high tax rates ranging from 0.69 to 0.87. Interestingly, the second line of the table shows that equation (9) implies *even higher* top rates when $\gamma = 0$, i.e. when the idea channel is turned off. The reason for this can be seen in

Table 1: Revenue-Maximizing Top Tax Rates

Case	Behavioral Elasticity (ζ)		
	0.1	0.2	0.3
Diamond-Saez	0.87	0.77	0.69
No ideas, $\gamma = 0$			
$\tau_0 = 0$	0.98	0.96	0.94
$\tau_0 = 0.20$	0.95	0.91	0.87
Degree of IRS, γ			
1/8	0.92	0.84	0.77
1/4	0.88	0.77	0.67
1/2	0.81	0.65	0.52
1	0.69	0.45	0.27

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 (9). Other baseline parameter values are $\Delta\rho = 0.1$, $\rho = \psi = 0.15$, and $\tau_0 = 0.20$.

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 ignores this GE effect. In our Cobb-Douglas setting, managers are paid a constant fraction of output, regardless of how much they work: lower effort leads to an offsetting increase in the wage w_m working through the Cobb-Douglas exponent ψ . This somewhat mitigates the effect on revenue ($\psi\zeta < \zeta$ since $\psi < 1$) and therefore raises the top tax rate. For example, the second row of the table finds a revenue-maximizing top tax rate of 96% instead of 77% for $\zeta = 0.2$ (which is the intermediate case we will focus on for clarity).

The third row of the table preserves the basic GE forces, but keeps the idea channel 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 91%.

The remainder of the table shows how these calculations change when we turn on the idea channel. With $\gamma = 1/2$, for example, the revenue-maximizing top tax rate falls to 65% instead of 91%.

2.3 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” $1 - \tau$, say by reducing the top tax rate from 75% to 50%. 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 $\eta_{Y,1-\tau} = \gamma\zeta$.

For a baseline, consider $\gamma = 1/2$ and a behavioral elasticity $\zeta = 0.2$. In this case, $\eta_{Y,1-\tau} = 1/10$ and $2^{1/10} \approx 1.07$. 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 7%! This is perhaps surprisingly small. It is reassuring in that it suggests that the values of γ we are using are not implausible.

Now consider the revenue side. The top income tax base $\Delta\rho$ is 10% of GDP, so without any change in economic behavior, the policy reduces top revenue from 7.5% to 5% of GDP, for a loss in revenue equal to 2.5% of GDP. However, the benefit of this policy is 7% of GDP with our baseline parameters. Focusing on the bottom 90%: their incomes go up by 7% because of innovation and then go down by $2.5\%/0.9 \approx 2.8\%$ because of lost redistribution. On net, workers gain from reducing the top tax rate from 75% to 50% in this example. This hints at the importance of welfare maximization instead of revenue maximization — explored in the next section.

One other thing to appreciate about the 7% 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%? 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 7% in the long run. But recall, that is in some sense the entire point of the growth literature: a relatively small number of researchers is responsible for the bulk of economic growth for the last 150 years.

3. Maximizing Worker Welfare

In the original Saez (2001) approach, the revenue-maximizing calculation can be viewed as characterizing a welfare-maximizing top tax rate with a zero Pareto weight on top earners. However, when the top tax rate affects the income of workers directly, the two are no longer equivalent. It is the results from this section that are most naturally compared to the Diamond and Saez calculation, and as we see next, maximizing the welfare of workers typically requires an even lower top tax rate.

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. Assume workers are below the top tax threshold and supply one unit of labor inelastically:

$$c^w = w(1 - \tau_0) \quad (12)$$

$$u_w(c) = \varphi \log c. \quad (13)$$

3.1 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 basic research, a good legal system, and education — things that also contribute to economic growth. Using equation (7), we now have

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

Alternatively, it is sometimes useful to express this equation as

$$[(1 - \rho)\tau_0 + \rho\tau]Y - (\tau - \tau_0)[\bar{w}(S_a + M)] = \Omega Y. \quad (15)$$

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

3.2 Maximizing the Welfare 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

$$\max_{\tau, \tau_0} \log(1 - \tau_0) + \log Y(\tau) \quad \text{s.t.} \quad (16)$$

$$\tau_0 Y + (\tau - \tau_0)[\rho Y - \bar{w}(S_a + M)] = \Omega Y. \quad (17)$$

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 welfare of workers subject to the government budget constraint satisfies*

$$\tau_{ww}^* = \frac{1 - \eta_{Y,1-\tau} \left(\frac{1-\rho}{\Delta\rho} \cdot \tau_0^* + \frac{1-\Delta\rho}{\Delta\rho} \cdot (1 - \tau_0^*) - \frac{\Omega}{\Delta\rho} \right)}{1 + \frac{\rho}{\Delta\rho} \eta_{Y,1-\tau}}. \quad (18)$$

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

$$\tau_{ww}^* = \frac{1 - \eta_{Y,1-\tau} \left[\frac{1-\Delta\rho}{\Delta\rho} - \frac{\Omega}{\Delta\rho} \left(1 + \frac{\bar{\rho}}{1-\Delta\rho} \right) \right]}{1 + \frac{\rho}{\Delta\rho} \eta_{Y,1-\tau} + \frac{\bar{\rho}}{1-\Delta\rho} \eta_{Y,1-\tau}}, \quad (19)$$

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

Equation (18) 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 (9). 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 (9). 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 (19), 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.¹⁰

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 (19) implies that¹¹

$$\tau \leq \tau_0 \text{ and } \kappa \geq \kappa_0 \iff \eta_{Y,1-\tau} \geq \frac{\Delta\rho}{1 - \Delta\rho}. \quad (20)$$

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

¹⁰A 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 (18), 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 (15). 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 (19)) and an additional multiplier effect coming through the fact that τ_0 depends on τ .

¹¹It is easier to see this result using the following solution for the top keep rate that maximizes worker welfare:

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

Table 2: Tax Rates that Maximize Worker Welfare

Degree of IRS, γ	— $\zeta = 0.2$ —		$\zeta = 0.1$	$\zeta = 0.3$
	τ_{ww}^*	τ_0^*	τ_{ww}^*	τ_{ww}^*
0	0.76	0.14	0.88	0.64
1/8	0.57	0.16	0.78	0.38
1/4	0.40	0.18	0.68	0.15
1/2	0.09	0.21	0.50	-0.26
1	-0.43	0.27	0.18	-0.90

Note: The table 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.

two keep rates are the same. If one yields more, then its keep rate will be higher.

For example, consider the case of $\Delta\rho = 0.1$ and $\zeta = 0.1$. In this case, $\frac{\Delta\rho}{1-\Delta\rho} = \frac{1}{9}$ and $\eta_{Y,1-\tau} = \frac{1}{10}(\gamma + \psi)$. Therefore, the flat tax will maximize worker welfare when $\gamma + \psi = 10/9$. If the behavioral elasticity is $\zeta = 0.2$ instead, the flat tax point is halved at $\gamma + \psi = 5/9 \approx 0.56$. And if $\gamma + \psi > 5/9$, then the top tax rate will be *below* τ_0 .

Table 2 shows some numerical examples for the baseline parameter values including $\Omega = 0.2$. With a behavioral elasticity of 0.2, the top tax rate falls to 40% for $\gamma = 1/4$ and just 9% for $\gamma = 1/2$. This compares to the rates of 77% and 65% for these two cases in Table 1 for revenue maximization. Notice that with $\psi = 0.15$, the case of $\gamma = 1/2$ means that $\gamma + \psi = 0.65$, which exceeds the flat tax value of 0.56, so the top tax rate is lower than τ_0 . And when $\gamma = 1$, the top tax rate is solidly into negative territory at -43%. If ideas are sufficiently important, subsidizing idea creation can be more effective than redistribution at raising worker welfare.¹²

¹²Rothschild and Scheuer (2013) provide a related example that highlights two important GE effects in the constant returns to scale case. On the one hand, increasing the top tax rate causes high earners (managers when $\gamma = 0$) to work less, which raises their wage and therefore tends to raise the optimal top tax rate. On the other hand, if managers are complementary to low-earning workers, taxing managers will lower their effort and lower the wage of workers. Rothschild and Scheuer (2013) suggest that these forces roughly offset in the Cobb-Douglas case. This can be seen in the first row of Table 2: when $\gamma = 0$, the optimal top rate is 0.76, very close to the Diamond and Saez version of 0.77 reported in Table 1.

4. Maximizing Utilitarian Social Welfare

To incorporate the welfare of entrepreneurs and managers, we have to specify their utility functions and how they choose effort. This section also delivers the promised discussion of the economics underlying the ζ elasticity.

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:

$$u(c, e) = \varphi \log c - \frac{\varepsilon}{1 + \varepsilon} e^{\frac{1+\varepsilon}{\varepsilon}}. \quad (21)$$

This specification is a simple form of “constant Frisch elasticity” preferences (Trabandt and Uhlig, 2011) and delivers constant effort in the presence of exponential growth in consumption.

In what follows, we focus on entrepreneurs, but managers are treated symmetrically. An entrepreneur with talent θ solves the following problem:

$$\begin{aligned} \max_{c, e} u(c, e) \quad \text{s.t.} \quad & (22) \\ c = & \bar{w}(1 - \tau_0) + [w_s \theta e - \bar{w}](1 - \tau) + R \\ = & \bar{w}(1 - \tau_0) - \bar{w}(1 - \tau) + w_s \theta e(1 - \tau) + R \\ = & \bar{w}(\tau - \tau_0) + w_s \theta e(1 - \tau) + R \end{aligned}$$

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:

$$e = \left(\frac{\varphi w_s \theta (1 - \tau)}{c} \right)^\varepsilon. \quad (23)$$

The Frisch elasticity of effort with respect to the wage or the keep rate — that is, holding consumption constant — equals ε .

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 substitution effect and the income effect cancel, so that in the absence of a rebate, a linear tax will have no effect on effort. For the nonlinear tax here, things are more complicated, but the effect of the top tax rate on effort still asymptotes to zero for highly talented people.¹³

This raises the interesting question of *why* the effort of top earners is estimated in the public economics literature to respond to top tax rates. The Prescott (2004) solution of rebating the per capita tax revenue to neutralize the income effect does not work; average tax revenue is a vanishing fraction of the highest incomes. But a variation on Prescott's solution does succeed: we assume that the rebates are heterogeneous according to talent. In particular, the rebates are such that equilibrium consumption of a researcher with talent θ is given by

$$c_\theta = w_s \theta e (1 - \tau)^{1-\alpha} \quad (24)$$

where α parameterizes the extent to which the rebate depends on the tax rate.¹⁴

The FOC for effort then reduces to

$$e^* = [\varphi(1 - \tau)^\alpha]^{\frac{\varepsilon}{1+\varepsilon}} \equiv [\varphi^{1/\alpha}(1 - \tau)]^{\zeta_u} \quad (25)$$

where the elasticity of effort with respect to the keep rate is $\zeta_u \equiv \alpha \frac{\varepsilon}{1+\varepsilon}$. Importantly, therefore $\eta_{x,1-\tau} = (\gamma + \psi)\zeta_u$ as in the earlier sections of the paper. That is, the elasticity of effort with respect to the keep rate depends on the uncompensated elasticity ζ_u , which is smaller than the compensated Frisch elasticity ε , in a way that depends on α . Also, notice that entrepreneurs of differing talent put in the same effort in equilibrium.

¹³Setting $R = 0$ and combining the FOC with the entrepreneur's budget constraint gives

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

and an increase in τ will reduce e . But as $\theta \rightarrow \infty$, e becomes independent of τ .

¹⁴That is, the researcher solves the problem in equation (22) taking the following rebates as exogenous:

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

where e^* does not depend on θ and is given in equation (25). Note that these rebates differ according to the entrepreneur's skill, θ . This raises the question of how this could be microfounded in a Mirrleesian framework in which it is typically assumed that the government cannot observe individuals' skills. Given that we calibrate the elasticity of entrepreneurial effort based on empirical evidence, the issue of precisely why top earners respond to taxes is not central to the present paper. But clearly this is an interesting question for future research.

Consumption, however, does depend on talent, θ :

$$c_\theta^* = \varphi^{\frac{\varepsilon}{1+\varepsilon}} w_s \theta (1 - \tau)^{1-\alpha+\zeta_u}. \quad (26)$$

If $\alpha = 1$, both consumption and effort are proportional to $(1 - \tau)^{\zeta_u}$. Alternatively, a smaller α moves us closer to the situation where the substitution and income effects cancel (for a given ε), 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:

$$SWF \equiv Lu(c^w) + S_b u(c^b) + S_a \int u(c_\theta^s, e_\theta^s) dF(\theta) + M \int u(c_\theta^m, e_\theta^m) dF(\theta).$$

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

$$\begin{aligned} \overline{SWF} &\equiv \frac{SWF - \omega}{\varphi} \cdot \frac{1}{L + S_b + S_a + M} \\ &= \log Y + \ell \log(1 - \tau_0) + s(1 - \alpha) \log(1 - \tau) - \frac{\zeta_u}{\alpha} s(1 - \tau)^\alpha, \end{aligned} \quad (27)$$

where $s \equiv \frac{S_a + M}{L + S_b + S_a + M}$, $\ell \equiv 1 - s$, and ω is a term that does not depend on τ or τ_0 .¹⁵

As shown in the [appendix](#), we then have the following result:

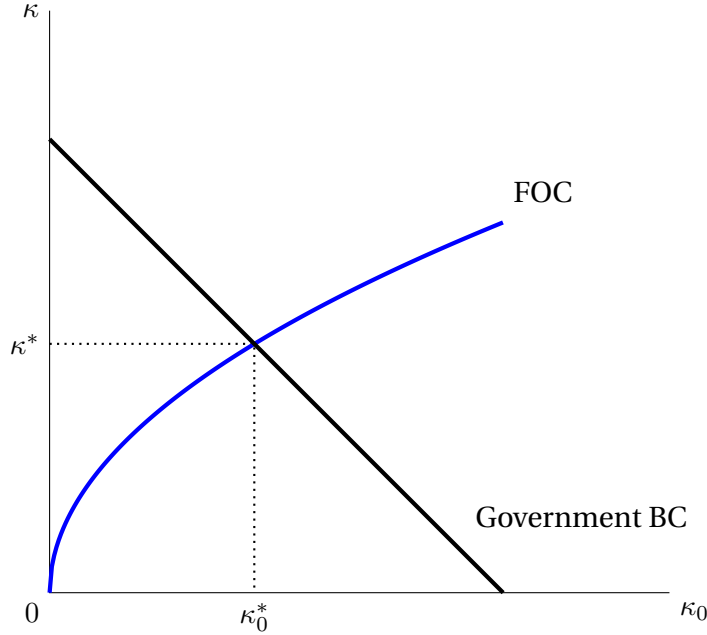
Proposition 2 (Maximizing Social Welfare): *The tax rates that maximize the social welfare function in (27) subject to the government budget constraint satisfy the following two equations in two unknowns, written in terms of the keep rates κ and κ_0 (i.e. $\kappa \equiv 1 - \tau$):*

$$s\zeta_u \kappa^\alpha + \frac{\kappa}{\kappa_0} \cdot \frac{\ell}{1 - \Delta\rho} (\Delta\rho + \bar{\rho}\eta_{Y,1-\tau}) = \eta_{Y,1-\tau} \left(1 + \frac{\bar{\rho}\ell}{1 - \Delta\rho} \right) + s(1 - \alpha) \quad (28)$$

¹⁵In particular,

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

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

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

$$\kappa_0(1 - \Delta\rho) + \kappa\Delta\rho = 1 - \Omega. \quad (29)$$

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 (28) can be rewritten as

$$\kappa = \frac{\eta_{Y,1-\tau} \left(1 + \frac{\bar{\rho}\ell}{1-\Delta\rho}\right)}{s\zeta_u + \frac{1}{\kappa_0} \frac{\ell}{1-\Delta\rho} (\Delta\rho + \bar{\rho}\eta_{Y,1-\tau})} \quad (30)$$

Our two equations — (29) and (30) — can then be analyzed graphically, as in Figure 1. In fact, inspecting equation (28) 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.
- $\uparrow\zeta_u \Rightarrow \uparrow\kappa$ 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 $\alpha = 1/2$.

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

Degree of IRS, γ	— $\zeta_u = 0.2$ —			$\zeta_u = 0.1$ τ^*	$\zeta_u = 0.3$ τ^*
	τ^*	τ_0^*	GDP loss if $\tau = 0.75$		
0	0.77	0.14	-0.3%	0.87	0.68
1/8	0.59	0.16	2.6%	0.77	0.44
1/4	0.42	0.18	6.4%	0.68	0.22
1/2	0.12	0.21	15.1%	0.49	-0.17
1	-0.40	0.27	32.7%	0.16	-0.81

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 $Y \propto (1 - \tau)^{(\gamma + \psi)\zeta_u}$ 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: $\alpha = 1/2$

Degree of IRS, γ	— $\zeta_u = 0.2$ —			$\zeta_u = 0.1$ τ^*	$\zeta_u = 0.3$ τ^*
	τ^*	τ_0^*	GDP loss if $\tau = 0.75$		
0	0.46	0.17	2.3%	0.51	0.40
1/8	0.28	0.19	5.6%	0.42	0.16
1/4	0.12	0.21	9.6%	0.33	-0.06
1/2	-0.17	0.24	18.2%	0.16	-0.45
1	-0.67	0.30	35.4%	-0.15	-1.07

Note: See notes to Table 3.

Two properties of these examples stand out. First, for $\gamma \geq 1/2$ and $\alpha = 1$, the top marginal tax rate is quite low, at 12% when $\zeta_u = 0.2$ and 49% when $\zeta_u = 0.1$. The numbers are even lower when $\alpha = 1/2$. Despite the utilitarian impetus to redistribute from 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, 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 $Y \propto (1 - \tau)^{(\gamma+\psi)\zeta_u}$.¹⁶ For example, when $\gamma = 1/2$, $\alpha = 1$, and $\zeta_u = 0.2$, the welfare-maximizing top tax rate is 12%, and GDP declines by 15.1% if instead the top tax rate is 75%.

5. Discussion

In this section, we summarize our results, discuss several issues related to economic growth and taxation, and provide additional intuition.

5.1 Summary

We summarize our calibration exercises in Table 5. As shown earlier, our basic setup delivers extremely high revenue-maximizing top tax rates in the absence of the idea channel ($\gamma = 0$). But when the idea channel is present, top tax rates fall considerably. To begin, consider the case in which the uncompensated labor supply elasticity is just $\zeta_u = 0.1$. The revenue-maximizing rate falls from 95% to 81% when the idea channel is present and $\gamma = 1/2$. However, in this case, 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 worker welfare is considerably lower, falling to 50%; notice that relative to the revenue-maximizing top tax rate of 95%, the keep rate $1 - \tau$ has increased by a factor of ten, from 5% to 50%. Taking into account the welfare of the entrepreneurs lowers the top rate further to 49% or 16%, depending on the income effect parameter α . The large change in tax rates is similar when $\zeta_u = 0.2$, declining from 91% to 17% across the exercises and by even more if $\zeta_u = 0.3$. Taking ideas into account has first-order effects

¹⁶This can be seen by combining equation (5) with (25).

Table 5: Summary of Calibration Exercises: Top rate, τ

Exercise	$\zeta_u = .1$	$\zeta_u = .2$	$\zeta_u = .3$
<i>No ideas, $\gamma = 0$</i>			
Revenue-maximization, $\tau_0 = 0$	0.98	0.96	0.94
Revenue-maximization, $\tau_0 = 0.20$	0.95	0.91	0.87
<i>With ideas, $\gamma = 1/2$</i>			
Revenue-maximization	0.81	0.65	0.52
Maximize worker welfare	0.50	0.09	-0.26
Maximize utilitarian welfare ($\alpha = 1$)	0.49	0.12	-0.17
Maximize utilitarian welfare ($\alpha = 1/2$)	0.16	-0.17	-0.45

Note: This table collects the results from the various calibration exercises for the default parameter values of $\Delta\rho = 0.10$ and $s = 0.10$. In the second and third rows, we add the consideration that $\tau_0 = 0.2$. In the final three rows, we instead require the tax system to raise 20% of GDP in revenue ($\Omega = 0.2$).

on optimal top income taxation in this setting.

5.2 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 and it is an important topic for further research.

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.

I see two general 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. Growth has slowed in recent decades; perhaps it would have slowed even more had top tax rates not declined.

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.

5.3 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 reduce 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 by¹⁷

$$\tilde{r} = g_Y + \lambda g_y \left(\frac{1}{\rho_s(1-\tau)} - \frac{1}{\gamma} \right) \quad (31)$$

where $\rho_s \equiv w_s \tilde{S}_a / Y$ is the share of income paid to entrepreneurs and λ is the exponent on research effort in the idea production function. At the margin, after taxes, entrepreneurs earn $\rho_s(1-\tau)$. Speaking loosely, this equation suggests that the optimal subsidy is such that $\rho_s(1-\tau) \approx \gamma$. That is, you want entrepreneurs to earn a share of income roughly equal to the importance of applied ideas in the economy.¹⁸

To put some numbers on this, suppose $g_y = 2\%$, $g_Y = 3\%$, $\gamma = 1/2$, $\lambda = 1$, and $\rho_s = 10\%$. For $\tau = 1/2$, equation (31) 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; see Hall, Mairesse and Mohnen (2010) for a survey. 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.¹⁹ Of course it is not at all clear that social returns to formal R&D should be compared to the social return to effort in our model, particularly since one of our key points is that applied research is not easily subsidized through formal R&D subsidies, while the social return estimate in Bloom et al. is identified 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.

¹⁷A full derivation is presented in Appendix C.

¹⁸Interestingly, 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 that involve equity as well as efficiency. Also, we say “roughly” because the planner’s interest rate is not g_Y , but this is close.

¹⁹Two 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%.

5.4 Intuition and First-Best Entrepreneurial Effort

This section justifies some key intuition provided back in Section 2.2. We derive the first-best effort that maximizes social welfare, i.e. if the social planner were free to pick everyone's consumption and effort. We simplify the problem by ignoring heterogeneity in talent, i.e. setting $\theta = 1$ for everyone.

Consider the utilitarian social welfare function studied earlier:

$$SWF \equiv Lu(c^w) + S_b u(c^b) + S_a \int u(c_\theta^s, e_\theta^s) dF(\theta) + M \int u(c_\theta^m, e_\theta^m) dF(\theta).$$

Let r_i denote the consumption share of GDP for group $i \in \{w, b, a, m\}$, paralleling the income shares ρ_i .

Social welfare can then be expressed as

$$\frac{SWF}{Pop} = \hat{\varphi} + \varphi \log Y - s_a \frac{\varepsilon}{1 + \varepsilon} e_a^{\frac{1+\varepsilon}{\varepsilon}} - m \frac{\varepsilon}{1 + \varepsilon} e_m^{\frac{1+\varepsilon}{\varepsilon}}, \quad (32)$$

where $Pop = L + S_b + S_a + M$ denotes the population, $s_a \equiv \frac{S_a}{Pop}$, $m \equiv \frac{M}{Pop}$, and $\hat{\varphi}$ is a term that captures the distributional shares (r_i) but does not depend on effort.²⁰ If the planner were free to choose every allocation — the r_i consumption shares for each group as well as effort for the entrepreneurs and managers — the first best allocation would be to give everyone equal consumption and to pick efforts to maximize social welfare in equation (32). The solution implies

$$e_a^{fb} = \left(\frac{\varphi \gamma}{s_a} \right)^{\frac{\varepsilon}{1+\varepsilon}}. \quad (33)$$

This can be compared to the solution from the optimal tax problem, given above in equation (25) and repeated here:

$$e^* = (\varphi(1 - \tau)^\alpha)^{\frac{\varepsilon}{1+\varepsilon}}.$$

²⁰In particular,

$$\hat{\varphi} \equiv \sum_i \varphi \ell_i \log(r_i / \ell_i) - \varphi \log Pop$$

where ℓ_i denotes the population share of each group, e.g. $\ell_a = s_a = S_a / Pop$.

To generate first-best effort, the planner would therefore set the top tax rate to satisfy

$$(1 - \tau)^\alpha = \frac{\gamma}{s_a}. \quad (34)$$

As discussed in Section 2.2, if $s_a < \gamma$ (e.g. because a small number of researchers generates growth through the nonrivalry of ideas γ), then first-best effort would involve a negative top tax rate. The expression given in Section 2.2 is exactly this one, with $\alpha = 1$ for simplicity.

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	$Y_t = \int_0^{A_t} x_{it}^{1-\psi} di (\mathbb{E}[\theta e] M_t)^\psi$
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	$\dot{A}_t = \bar{a} (\mathbb{E}[\theta e] S_{at})^\lambda A_t^{\phi_a} B_t^{\beta_a}$
Basic ideas	$\dot{B}_t = \bar{b} S_{bt}^\lambda B_t^{\phi_b}$
Talent heterogeneity	$\theta_i \sim F(\theta)$
Utility (S_a, M)	$u(c, e) = \varphi \log c - \frac{\varepsilon}{1+\varepsilon} e^{\frac{1+\varepsilon}{\varepsilon}}$
Utility (L, S_b)	$u(c) = \varphi \log c$

We begin by stating the key decision problems that have to be solved by the various agents in the economy and then we combine them 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 work inelastically and consume their income after taxes:

$$c^w = w(1 - \tau_0) \quad (35)$$

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

$$\max_{\{c_t, e_t\}} \int_0^\infty u(c_t, e_t) \exp(-\delta t) dt, \quad (36)$$

subject to

$$\dot{v}_t = (r_t - n)v_t + \bar{w}(\tau - \tau_0) + w_s \theta e(1 - \tau) + R_t - c_t, \quad v_0 \text{ given}, \quad (37)$$

$$u(c, e) = \varphi \log c - \frac{\varepsilon}{1 + \varepsilon} e^{\frac{1+\varepsilon}{\varepsilon}} \quad (38)$$

$$\lim_{t \rightarrow \infty} v_t \exp\left(-\int_0^t (r_s - n) ds\right) \geq 0. \quad (39)$$

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 θ , which they take as given. We are suppressing the dependence of v_t and R_t on type θ 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

$$Y_t = \tilde{M}_t^\psi \int_0^{A_t} x_{it}^{1-\psi} di. \quad (40)$$

That is, at each point in time t , taking as given the manager's wage rate w_{mt} , the measure

of intermediate goods A_t , and the prices of the intermediate goods p_{it} , the representative firm solves

$$\max_{\{x_{it}\}, \tilde{M}_t} \tilde{M}_t^\psi \int_0^{A_t} x_{it}^{1-\psi} di - w_{mt} \tilde{M}_t - \int_0^{A_t} p_{it} x_{it} di. \quad (41)$$

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 when 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

$$\max_{p_{it}} \pi_{it} \equiv (p_{it} - w_t)x(p_{it}) \quad (42)$$

where $x(p_{it})$ is the demand from the final goods sector for intermediate good i at price p_{it} . This demand curve comes from a first-order condition in Problem (FG).²¹

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 :

$$\dot{A}_t = \bar{a} \tilde{S}_{at}^\lambda A_t^{\phi_a} B_t^{\beta_a} \quad (43)$$

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

$$\dot{A}_t = \bar{v}_t \tilde{S}_{at}. \quad (44)$$

That is, the duplication effects associated with λ and the knowledge spillovers associated with ϕ_a and β_a in equation (43) are assumed to be external to the individual

²¹Specifically, 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

$$\max_{\tilde{S}_{at}} P_{At} \bar{\nu}_t \tilde{S}_{at} - w_{st} \tilde{S}_{at}, \quad (45)$$

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} hiring basic researchers at 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

$$\dot{B}_t = \bar{b} S_{bt}^\lambda B_t^{\phi_b}. \quad (46)$$

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

$$\begin{aligned} \tau_0(wL + w_b S_b + w_s \mathbb{E}[\theta e] S_a + w_m \mathbb{E}[\theta e] M) + (\tau - \tau_0)[(w_s \mathbb{E}[\theta e] - \bar{w}) S_a + (w_m \mathbb{E}[\theta e] - \bar{w}) M] \\ = G_{bt} + \int R(\theta) dF(\theta) + G_{ot} \equiv \Omega Y_t. \end{aligned} \quad (47)$$

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 $\{c_{\theta t}^s, c_{\theta t}^m, c_t^w, e_{\theta t}^s, e_{\theta t}^m, v_{\theta t}^s, v_{\theta t}^m, \{x_{it}\}, Y_t, \tilde{M}_t, \tilde{S}_{at}, S_{bt}, L_t, A_t, B_t, \bar{\nu}_t, G_{bt}, G_{ot}\}_{t=0}^\infty$ and prices $\{w_t, w_{st}, w_{mt}, r_t, \{p_{it}\}, \{\pi_{it}\}, R_t(\theta), P_{At}\}_{t=0}^\infty$ such that for all t :

1. $c_{\theta t}^s, c_{\theta t}^m, c_t^w, e_{\theta t}^s, e_{\theta t}^m, v_{\theta t}^s, v_{\theta t}^m$, and $v_{\theta t}^m$ 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: $\int (S_{at}v_{\theta t}^s + M_tv_{\theta t}^m)dF(\theta) = P_{At}A_t$.
7. (w_t) The labor market clears: $L_t + S_{bt} = N_t$.
8. (w_{st}) The market for entrepreneurs clears: $\mathbb{E}[\theta e]S_{at} = \tilde{S}_{at}$.
9. (w_{mt}) The manager market clears: $\mathbb{E}[\theta e]M_t = \tilde{M}_t$.
10. (\bar{v}_t) The idea production function is satisfied: $\bar{v}_t = \nu\tilde{S}_{at}^{\lambda-1}A_t^{\phi_a}B_t^{\beta_a}$.
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 $R_t(\theta) = w_s e^* \theta (1-\tau)^{1-\alpha} - [\bar{w}_t(\tau - \tau_0) + w_s e^* \theta (1-\tau) + (r - g_Y)v_t]$.
13. Y_t is given by the production function in (40).
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 (43).
16. B_t is given by the production function in (46).
17. G_{bt} is exogenously given.
18. (G_{ot}) The government budget constraint in (47) 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

$$Y = A^\psi \tilde{M}^\psi L^{1-\psi}. \quad (48)$$

The stock of ideas along the BGP is in turn given by²²

$$A_t^* = \left(\nu_a \tilde{S}_{at} S_{bt}^{\frac{\beta_a}{1-\phi_b}} \right)^{\frac{\lambda}{1-\phi_a}}. \quad (49)$$

Putting these two equations together,

$$Y_t^* = \left(\nu_a \mathbb{E}[\theta e] S_a S_b^\beta \right)^\gamma \tilde{M}_t^\psi L_t^{1-\psi} \quad (50)$$

where $\gamma \equiv \frac{\psi\lambda}{1-\phi_a}$ and $\beta \equiv \frac{\beta_a}{1-\phi_b}$. This is exactly the specification of the production function for our simple model in (5).

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

$$y_t^* \propto \left(\tilde{S}_{at} S_{bt}^\beta \right)^\gamma. \quad (51)$$

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

$$g_y = \gamma(1 + \beta)g_S = \gamma(1 + \beta)n. \quad (52)$$

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

$$\frac{\dot{c}_t(\theta)}{c_t(\theta)} = \frac{\dot{c}_t}{c_t} = (r_t^{eq} - n - \delta) \quad (53)$$

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

$$c_t^*(\theta) = \varphi^{\frac{\varepsilon}{1+\varepsilon}} w_s \theta (1 - \tau)^{1-\alpha(1-\frac{\varepsilon}{1+\varepsilon})}. \quad (54)$$

Equilibrium effort is then independent of θ :

$$e^* = [\varphi(1 - \tau)^\alpha]^{\frac{\varepsilon}{1+\varepsilon}} \quad (55)$$

²²Define

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

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

$$\frac{w_m \tilde{M}}{Y} = \psi \equiv \rho_m \quad \text{and} \quad \frac{Apx}{Y} = 1 - \psi \quad (56)$$

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

$$p_{it}^{eq} = p_t^{eq} = \frac{w_t}{1 - \psi}. \quad (57)$$

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

$$\frac{A\pi}{Y} = \psi(1 - \psi). \quad (58)$$

Also, recall from the resource constraint for labor that $Ax = L$. Substituting this into (56) together with $p = w/(1 - \psi)$ yields

$$\frac{wL}{Y} = (1 - \psi)^2 \equiv \rho_\ell. \quad (59)$$

(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

$$\frac{w_s \tilde{S}_a}{Y} = g_A \frac{P_A A}{Y}. \quad (60)$$

The arbitrage equation pins down the price of an idea as

$$P_A = \frac{\pi}{r - g_\pi} = \frac{\pi}{r - (g_Y - g_A)}. \quad (61)$$

Therefore

$$\frac{P_A A}{Y} = \frac{1}{r - (g_Y - g_A)} \cdot \frac{A\pi}{Y} = \frac{1}{r - (g_Y - g_A)} \cdot \psi(1 - \psi). \quad (62)$$

And substituting this back into (60) gives

$$\frac{w_s \tilde{S}_a}{Y} = \frac{g_A}{r - (g_Y - g_A)} \cdot \psi(1 - \psi) \equiv \rho_s \quad (63)$$

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.²³

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. When the creation of ideas is the ultimate source of economic growth, this force can sharply constrain welfare-maximizing top tax rates.

The point of this paper is not to calculate precisely the value of the “correct” optimal top tax rate. As an extensive literature points out, there are many considerations that

²³One might wonder about where the payments to basic researchers $w_b S_b$ 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.

need to be taken into account to provide such a number. Instead, the point is that the discovery of nonrival ideas by entrepreneurs appears to have a first-order effect on the calculation and should be included in future work.

Many unanswered questions remain and deserve further study. Why is the evidence on economic growth and taxes so murky? How sensitive is innovation to taxation? Akcigit, Grigsby, Nicholas and Stantcheva (2021) suggest that the effects are larger than one might have expected, but clearly this is an important avenue for future research. It would also be valuable to study transition dynamics. The gains from redistribution kick in immediately while the losses from accumulating fewer ideas accrue slowly. This change would likely lead to higher top rates.

A. Proofs

The Revenue-Maximizing Top Tax Rate in Equation (9). Starting from equation (8), note that $\frac{dY}{d\tau} = -\frac{dY}{d(1-\tau)}$ so that

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

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

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

The first order condition with respect to τ_0 is

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

The first order condition with respect to τ is

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

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

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

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

$$\overline{SWF} = \log Y + \ell \log(1 - \tau_0) + s(1 - \alpha) \log(1 - \tau) - s \frac{\zeta_u}{\alpha} (1 - \tau)^\alpha \quad (\text{A5})$$

subject to the government budget constraint (GBC):

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

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

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

The first order condition with respect to τ is

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

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

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

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

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

which can be simplified further to

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

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

$$\eta + s(1 - \alpha) - \frac{\eta^\ell(1 - \rho)}{1 - \Delta\rho} = \zeta_u s \kappa^\alpha + \ell \frac{\Delta\rho}{1 - \Delta\rho} \frac{\kappa}{\kappa_0} \left(1 + \frac{\eta\rho}{\Delta\rho} \right) - \frac{\eta^\ell}{1 - \Delta\rho} \frac{1 - \Omega}{\kappa_0} \\ = \zeta_u s \kappa^\alpha + \frac{1}{\kappa_0} \frac{\ell}{1 - \Delta\rho} [\kappa(\Delta\rho + \eta\rho) - \eta(1 - \Omega)] \\ = \zeta_u s \kappa^\alpha + \frac{1}{\kappa_0} \frac{\ell}{1 - \Delta\rho} [\kappa\Delta\rho + \eta[\rho\kappa - (1 - \Omega)]] \quad (\text{A8})$$

Now recall that the government budget constraint can be manipulated:

$$(1 - \Delta\rho)\kappa_0 + \Delta\rho\kappa = 1 - \Omega \quad (\text{A9})$$

$$\Rightarrow (\rho - \bar{\rho})\kappa - (1 - \Omega) = -(1 - \Delta\rho)\kappa_0$$

$$\Rightarrow \rho\kappa - (1 - \Omega) = \bar{\rho}\kappa - (1 - \Delta\rho)\kappa_0. \quad (\text{A10})$$

Substituting this expression into (A8) gives

$$\eta + s(1 - \alpha) - \frac{\eta^\ell(1 - \rho)}{1 - \Delta\rho} = \zeta_u s \kappa^\alpha + \frac{1}{\kappa_0} \frac{\ell}{1 - \Delta\rho} [\kappa\Delta\rho + \eta[\bar{\rho}\kappa - (1 - \Delta\rho)\kappa_0]] \\ = \zeta_u s \kappa^\alpha + \frac{\kappa}{\kappa_0} \frac{\ell}{1 - \Delta\rho} [\Delta\rho + \eta\bar{\rho}] - \eta^\ell$$

Rearranging and collecting terms:

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

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

B. Star Inventors and Regular Inventors — An Extended Model

Suppose there are two kinds of inventors: star inventors and regular inventors. Star inventors produce big ideas and always face the top tax rate, while regular inventors only rarely invent a big idea and are therefore very unlikely to face the top tax rate. With (low) probability p , regular inventors produce a big idea (θ_h) and face the top tax rate, while with high probability $1 - p$ they produce a small idea (θ_l) and face τ_0 . Star inventors are a fraction q of the population while regular inventors make up the rest.

As in the detailed model in Section 4, star inventors have an elasticity of effort with respect to the keep rate equal to a preference parameter ζ_{star} . In contrast, regular inventors have a much lower elasticity, because they are unlikely to ever face the top rate. This low elasticity could be microfounded, but it is simpler to just assert that they have a very low value ζ , possibly equal to zero.

If S_a is the total number of researchers, then from equation (2) the number of ideas produced by the inventors is (ignoring basic research for now and setting $\beta = 0$)

$$\begin{aligned} A &= \nu_a S_a (q \theta_h e_s + (1 - q) e_r [p \theta_h + (1 - p) \theta_l]) \\ &= \nu_a S \cdot \sum_{i \in \{s, r\}} \omega_i e_i \end{aligned}$$

where the second equation allows for the possible generalization to more types: e_i is the effort of type i and ω_i is the total number of ideas produced by a unit of effort by type i . For example, $\omega_s = q \theta_h$ for the star inventors.

The key thing we need to know for computing the optimal top tax rate — the suffi-

cient statistic — is $\frac{d \log A}{d \log(1-\tau)}$; in the main text, this simplifies to $\frac{d \log e}{d \log(1-\tau)}$ since we only have one type. It is easy to take derivatives of the equation above to see that

$$\frac{d \log A}{d \log(1-\tau)} = \sum_i \frac{\omega_i e_i}{A} \cdot \frac{d \log e_i}{d \log(1-\tau)} \quad (\text{A11})$$

idea share tax elasticity

In other words, the key elasticity for the paper is the idea-share weighted average of the underlying elasticities of the different types of inventors.

One can then use the evidence in Appendix Table C.37 of Akcigit, Grigsby, Nicholas and Stantcheva (2021) to estimate $\frac{d \log A}{d \log(1-\tau)}$. For example, assigning a zero elasticity to the regular types, the estimate would just equal the elasticity for star inventors multiplied by their share of idea creation. Their elasticity in Table C.37 is 0.8 or 0.9. Like many things, idea production is highly skewed toward top inventors. For example, the top 5% of inventors produce 22% of patents and 36% of citations according to private communication from Ufuk Akcigit. Table C.37 treats star inventors as the top 10% of inventors, so those shares are presumably more like 30% and 45%. Multiplying these two numbers together leads to estimates of $\frac{d \log A}{d \log(1-\tau)}$ of between $0.30 \times 0.80 = 0.24$ and $0.45 \times 0.9 = 0.405$. These numbers are consistent with the baseline numbers used in the paper, and even a little larger.

C. 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

$$\Delta A_{t+1} = G(\tilde{S}_t, A_t, B_t) = \tilde{S}_t^\lambda A_t^{\phi_a} B_t^{\beta_a} \quad (\text{A12})$$

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 to be gained by this variation as \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 $\tilde{S}_t = \bar{\epsilon} \theta S_t$, and the key distortion in the economy is on research effort. In particular, the researcher's first order condition from equation (22) is

$$-\frac{1}{\theta} \cdot \frac{u_e}{u_c} = w_s(1 - \tau) \equiv \tilde{w}. \quad (\text{A13})$$

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

$$\tilde{r} \approx \lambda \sigma g_A \cdot \frac{1}{\rho_s(1 - \tau)} + g_Y - (1 - \phi_a)g_A. \quad (\text{A14})$$

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 (31).

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," *Review of Economic Studies*, forthcoming 2020.

- , John Grigsby, Tom Nicholas, and Stefanie Stantcheva, “Taxation and Innovation in the 20th Century,” *The Quarterly Journal of Economics*, forthcoming 2021.
- , 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.
- Arkolakis, Costas, Sun Kyoung Lee, and Michael Peters, “European Immigrants and the United States’ Rise to the Technological Frontier,” Technical Report 2020. Yale University manuscript.
- Atkeson, Andrew and Ariel Burstein, “Aggregate Implications of Innovation Policy,” *Journal of Political Economy*, December 2019, 127 (6), 2625–2683.
- Badel, Alejandro and Mark Huggett, “The Sufficient Statistic Approach: Predicting the Top of the Laffer Curve,” *Journal of Monetary Economics*, 2017, 87 (C), 1–12.
- , —, and Wenlan Luo, “Taxing Top Earners: A Human Capital Perspective,” *The Economic Journal*, 2020, 130 (629), 1200–1225.
- Bassetto, Marco and Jess Benhabib, “Redistribution, Taxes, and the Median Voter,” *Review of Economic Dynamics*, 2006, 9 (2), 211–223.
- Bell, Alexander M., Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, “Do Tax Cuts Produce More Einsteins? The Impacts of Financial Incentives vs. Exposure to Innovation on the Supply of Inventors,” *Journal of the European Economic Association*, April 2019, 17 (3), 651–677.
- , —, —, —, and —, “Who Becomes an Inventor in America? The Importance of Exposure to Innovation,” *The Quarterly Journal of Economics*, 2019, 134 (2), 647–713.
- Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, “Are Ideas Getting Harder to Find?,” *American Economic Review*, April 2020, 110 (4), 1104–44.
- , 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.

- Buera, Francisco J and Ezra Oberfield, “The Global Diffusion of Ideas,” *Econometrica*, 2020, 88 (1), 83–114.
- Chamley, Christophe, “Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives,” *Econometrica*, May 1986, 54 (3), 607–622.
- Chetty, Raj, “Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, 2012, 80 (3), 969–1018.
- 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.
- Garicano, Luis, Claire Lelarge, and John Van Reenen, “Firm Size Distortions and the Productivity Distribution: Evidence from France,” *American Economic Review*, 2016, 106 (11), 3439–79.
- 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, “Measuring the Returns to R&D,” in Bronwyn H. Hall and Nathan Rosenberg, eds., *Handbook of the Economics of Innovation*, Vol. 2, 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., “R&D-Based Models of Economic Growth,” *Journal of Political Economy*, August 1995, 103 (4), 759–784.

- , “Sources of U.S. Economic Growth in a World of Ideas,” *American Economic Review*, March 2002, 92 (1), 220–239.
- , “Growth and Ideas,” in Philippe Aghion and Steven A. Durlauf, eds., *Handbook of Economic Growth*, New York: North Holland, 2005, pp. 1063–1111.
- , “The Past and Future of Economic Growth: A Semi-Endogenous Perspective,” *Annual Review of Economics*, August 2022, 14, forthcoming.
- and John C. Williams, “Measuring the Social Return to R&D,” *Quarterly Journal of Economics*, November 1998, 113 (4), 1119–1135.
- Judd, Kenneth L., “Redistributive Taxation in a Simple Perfect Foresight Model,” *Journal of Public Economics*, October 1985, 28 (1), 59–83.
- 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.
- Kortum, Samuel S., “Research, Patenting, and Technological Change,” *Econometrica*, 1997, 65 (6), 1389–1419.
- Lockwood, Benjamin, Charles Nathanson, and E. Glen Weyl, “Taxation and the Allocation of Talent,” *Journal of Political Economy*, 2017, 125 (5), 1635–1682.
- Lucas, Robert E., Jr., “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, 1978, 9, 508–523.
- 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.
- Ngai, Rachel and Roberto Samaniego, “Accounting for Research and Productivity Growth Across Industries,” *Review of Economic Dynamics*, July 2011, 14 (3), 475–495.
- Peters, Michael, “Market Size and Spatial Growth—Evidence from Germanys Post-War Population Expulsions,” 2019. Yale University manuscript.

- and Connor Walsh, “Population Growth and Firm Dynamics,” 2021. Yale University manuscript.
- 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.
- Prescott, Edward C., “Why Do Americans Work So Much More than Europeans?,” *Quarterly Review*, July 2004, *Federal Reserve Bank of Minneapolis*, 2–13.
- Romer, Paul M., “Endogenous Technological Change,” *Journal of Political Economy*, October 1990, 98 (5), S71–S102.
- Rothschild, Casey and Florian Scheuer, “Redistributive Taxation in the Roy Model,” *The Quarterly Journal of Economics*, 2013, 128 (2), 623–668.
- and — , “A Theory of Income Taxation under Multidimensional Skill Heterogeneity,” Working Paper 19822, National Bureau of Economic Research January 2014.
- and — , “Optimal Taxation with Rent-Seeking,” *Review of Economic Studies*, 2016, 83 (3), 1225–1262.
- Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin, “Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium,” *Econometrica*, 2020, 88 (2), 469–493.
- Saez, Emmanuel, “Using Elasticities to Derive Optimal Tax Rates,” *Review of Economic Studies*, 2001, 68, 205–229.
- , Joel Slemrod, and Seth H. Giertz, “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, March 2012, 50 (1), 3–50.
- Scheuer, Florian and Iván Werning, “The Taxation of Superstars,” *Quarterly Journal of Economics*, 2017, 132 (1), 211–270.
- Seegerstrom, Paul, “Endogenous Growth Without Scale Effects,” *American Economic Review*, December 1998, 88 (5), 1290–1310.
- Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick, “Capitalists in the Twenty-First Century*,” *The Quarterly Journal of Economics*, 07 2019, 134 (4), 1675–1745.
- Stiglitz, Joseph E., “Self-Selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 1982, 17 (2), 213–240.

- Stokey, Nancy L. and Sergio Rebelo, "Growth Effects of Flat-Rate Taxes," *Journal of Political Economy*, June 1995, 103, 519–550.
- Straub, Ludwig and Iván Werning, "Positive Long Run Capital Taxation: Chamley-Judd Revisited," *American Economic Review*, January 2020, 110 (1), 86–119.
- Tervio, Marko, "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 2008, 98 (3), 642–668.
- Trabandt, Mathias and Harald Uhlig, "The Laffer Curve Revisited," *Journal of Monetary Economics*, 2011, 58 (4), 305–327.
- Wolfe, Raymond M., "Business R&D Performance in the United States Tops \$300 Billion in 2012," *NCSES Info Brief, NSF*, October 2014, 15-303.