Taxation and the Allocation of Talent*

Benjamin B. Lockwood† Charles G. Nathanson‡ E. Glen Weyl§

November 2015

Abstract

Taxation affects the allocation of talented individuals across professions by blunting material incentives and thus relatively magnifying the non-pecuniary benefits of pursuing a "calling." Estimates from the literature suggest high-paying professions have negative externalities, whereas low-paying professions have positive externalities. A calibrated model therefore prescribes negative rates on middle-class incomes and positive rates on the rich. However, the welfare gains from implementing such a policy are small and are dwarfed by the gains from profession-specific taxes and subsidies. These results depend crucially on externality estimates and labor-substitution patterns across professions. Both merit greater empirical study.

*This paper builds on and replaces a 2007 working paper solely authored by Weyl, "Psychic Income, Taxes and the Allocation of Talent." We gratefully acknowledge the financial support of the Alfred P. Sloan and Marion Ewing Kauffman Foundations, which financed the research assistance of Joshua Bosshardt, Joe Mihm, Matt Solomon, and Daichi Ueda. We owe a debt to Claudia Goldin for providing access to data. We also appreciate the helpful comments of Tony Atkinson, Raj Chetty, Neale Mahoney, Andrei Shleifer, Joel Slemrod, Matthew Weinzierl, Vanessa Williamson, workshop attendees at Harvard University, Princeton University, the Hebrew University of Jerusalem, KU Leuven the 2013 Allied Social Sciences Association meetings, the NBER Public Economics meeting the 2013 Society for Economic Dynamics meeting, and especially Charlotte Cavaille, as well as the insightful discussion given by Florian Scheuer. We also thank the editor, Jesse Shapiro, and six anonymous referees for detailed comments. All errors are our own.

†Department of Economics, Harvard University: Littauer Center, 1805 Cambridge Street; Cambridge, MA 02138: lockwood@fas.harvard.edu, http://www.benlockwood.com.

‡Kellogg School of Management, Northwestern University: 2001 Sheridan Road; Evanston, IL 60208: nathanson@kellogg.northwestern.edu, http://www.charlesnathanson.com.

§Microsoft Research New England, One Memorial Drive, Cambridge, MA 02142 USA and Department of Economics, University of Chicago: glenweyl@microsoft.com.

If we don't have an economy built on bubbles and financial speculation, our best and brightest won't all gravitate towards careers in banking and finance. Because if we want an economy that's built to last, we need more of those young people in science and engineering. This country should not be known for bad debt and phony profits. We should be known for creating and selling products all around the world.

- President Barack Obama, Speech at Osawatomie High School, December 6, 2011

1 Introduction

The allocation of talented individuals across professions varies widely over time and space. According to Goldin et al. (2013), more than twice as many male Harvard alumni from the 1969-1972 cohorts pursued careers in academia and in non-financial management as pursued careers in finance. Twenty years later, careers in finance were 50% more common than in academia and were comparable to those in non-financial management. If private product approximates social product, talented individuals constitute a large fraction of many societies' human capital: in the United States, for example, the top 10% of income earners generate just under half of all income, and nearly the top 1% generate one fifth (Atkinson et al., 2011). Furthermore, if, as Baumol (1990) and Murphy et al. (1991) argue, different professions have different ratios of social to private product, these differences in talent allocation across societies have important implications for aggregate welfare. Recent evidence strongly suggests such externalities not only exist but are large (Murphy and Topel, 2006; French, 2008). In this paper, we quantitatively evaluate the impact of non-linear income taxation on the allocation of talent, and we compute the tax schedule that maximizes aggregate (Pigouvian) welfare.

Our analysis adds to a growing literature (Philippon, 2010; Piketty et al., 2014; Rothschild and Scheuer, 2014*a*,*b*) that emphasizes the role of income taxation in responding to externalities of some activities. We extend this literature—and the perturbation approach used more generally to derive optimal taxes (Saez, 2001)—by incorporating a discrete, long-run "allocative" elasticity that governs talented workers' choice of profession. This margin of labor supply is distinct from both the standard short-run intensive margin of effort emphasized in the literature above and the extensive margin of exiting the labor force studied by Saez (2002).

In this allocative framework, workers make a long-term choice between well-paying professions and lower-paying "callings" that offer higher non-pecuniary benefits.[1] Higher marginal tax rates incent workers to "follow their passion" by blunting the relative after-tax pecuniary compensation of the more lucrative professions. To the extent that better-paying professions generate negative

[1]By contrast, Rothschild and Scheuer assume a concave utility function of continuous effort in different activities, which implies individuals will never make discrete moves across income and externality levels.

(or less positive) externalities, raising marginal tax rates can generate social welfare gains from the movement of workers into socially productive professions. Because individuals might switch into a number of professions—each generating different externalities and tax revenues—when taxes rise, the full set of substitution patterns of individuals across professions becomes critical to determining optimal taxes. As we highlight theoretically in Section 2 and with a simple example in Section 3, because they involve discrete jumps in income, these substitution patterns make applying the first-order approach of Mirrlees (1971) invalid and thus complicate the elegant characterizations of optimal tax rates provided by Rothschild and Scheuer in a model where only the intensive margin is operative.

The core of this paper is therefore a structural model of profession choice that imposes strong restrictions on substitution patterns in order to estimate how the allocation of talent would change under different income-tax regimes. The key inputs into our estimation are the distributions of income within different professions, the elasticities of labor supply on both the intensive and allocative margins, and the aggregate externalities on society from each profession. We take the externality estimates from the economics literature, which suggests these externalities are, although highly uncertain, likely to be huge and quite heterogeneous. Murphy and Topel (2006) estimate that medical research generates a positive externality of more than 15% of GDP, whereas French (2008) calculates the financial profession's income includes 1.4% of GDP in rent-seeking. Our main findings for optimal policy are the following:

1. The optimal income tax features top rates of about 35%, which are close to the existing top rates in the year from which we draw data (2005). This positive top rate induces long-term migration of talented workers to professions in which they earn less income but produce more externalities. However, because both positive and negative externality professions exist at high income, taxes have a theoretically ambiguous effect through the intensive margin and empirically are optimally slightly negative absent the allocative response. Section 3 shows these results analytically for a simplified case with three professions.

2. Although the optimal non-linear, profession-general tax rates differ significantly from 0, they achieve only small welfare gains (1.2%) relative to *laissez-faire*. The optimal tax fails to reallocate much talent to the professions in which it is undersupplied relative to the social optimum. By contrast, profession-targeted policies can achieve much more. We show an optimal linear subsidy to research professions achieves more than 40 times the welfare gains of our baseline optimal tax.

3. In Section 5.3, we show the key features of the optimal nonlinear income tax are robust to the details of how externalities accrue—which professions affect output in which others, and whether the externalities are linear or have diminishing returns to scale. Our results *are* sensitive to the magnitude of externalities we assume, especially in the research and

management professions, and to the nature of allocative substitution across professions. This sensitivity suggests these understudied patterns are crucial to determining optimal tax policy.

Throughout this paper, we analyze only efficiency, rather than redistributive, gains from taxation. Adding redistributive motives (as in a previous draft of this paper) would make our analysis normatively more complete without significant computational complexity. We focus on pure efficiency maximization because it highlights as sharply as possible the role of substitution patterns. In particular, efficiency maximization implies (see Section 2.2) the *total* elasticity of taxable income, which is so crucial in the canonical Vickrey (1945) redistributive framework, is irrelevant for deriving the optimal tax schedule. Only the relative importance of the allocative and intensive margins impacts optimal tax rates.

Another reason to restrict attention to efficiency is to probe the explanatory power of the "Just Desserts Theory" of Mankiw (2010) that taxation should ensure individuals receive their social contribution. Perhaps surprisingly, we show such a theory is able to account for the broad outlines of existing US income taxation. Furthermore, wealth maximization strikes us as more consistent with patterns of opinion on taxation outside of mainstream neoclassical economics, particularly among the public (Brown, 2011; CNN/ORC, 2011; Parker, 2012) and leading ideologues of the left and right such as Marx (1867) and Rand (1957), respectively. We believe this paper's framework is a useful tool for organizing, comparing, informing and potentially reconciling views of optimal taxation outside of economics, which many have argued is an important goal of applied welfare economics (Gul and Pesendorfer, 2008; Weinzierl, 2014; Saez and Stantcheva, Forthcoming).

Given the centrality of income taxes in the public debate, our analysis focuses on non-linear income taxes applied uniformly across professions. However, our conclusions strongly suggest targeted policies, such as subsidies for research, would more effectively correct any misallocation of talent. The large heterogeneity of professions at a given income level blunts the ability of an income tax to reallocate workers from and to specific professions. The magnitude of profession-specific subsidies and taxes suggested by our quantitative analysis—as much as several hundred percent—should be approached with caution given they could be subject to gaming as suggested by Rothschild and Scheuer and that quantifying the general equilibrium effects they would cause is difficult. Nonetheless, large subsidies and taxes have the potential to enhance efficiency much more dramatically than a standard, non-linear income tax.

In addition to our focus on allocation and efficiency, our analysis departs from the literature in several other ways. First, we allow for positive externalities, not just rent-seeking as in Rothschild and Scheuer, and these positive externalities turn out to be the largest quantitative contributors to our results. Second, our analysis is primarily quantitative. The present model incorporates a large number of professions and is estimated using a variety of data sources; previous literature on how income taxation should respond to externalities has involved primarily qualitative, illustrative models. As an example of the marginal contribution of this quantitative approach, we find that the targeting of externalities between professions emphasized by Rothschild and Scheuer in a model with two activities is quantitatively less important than labor-substitution patterns and levels of externalities once one accounts for the the many professions that skilled workers may choose.[2] Finally, in contrast to Piketty et al. (2014) and Rothschild and Scheuer, we abstract from any role taxes may have on the allocation of time within a profession across activities of different merit, assuming a homogeneous externality created by all output of a given profession.

2 A Model of Optimal Income Taxation with Externalities

All formal proofs of results and omitted derivations in this section appear in Appendix A.1.

2.1 Statement of the problem

A mass 1 of individuals work in n professions. Each worker is characterized by a $2n$-dimensional type $\theta = (a, \psi)$, where $a = (a_1, ..., a_n)$ is a vector of profession-specific productivities and $\psi = (\psi_1, ..., \psi_n)$ is a vector of the non-pecuniary utility the worker receives in each profession. The distribution of types θ among workers is given by a non-atomic and differentiable distribution function f with full support on a convex and open $\Theta \subseteq \mathbb{R}^{2n}$.

Labor supply consists of allocative and intensive margins. On the allocative margin, each worker chooses exactly one of the n professions to enter; we denote the profession choice of a worker of type θ by $i(\theta)$. The intensive margin consists of a choice of hours $h_i(\theta)$ to work in profession i, where $h_i(\theta) \geq 0$ for all i and θ. Because each individual works in only one profession, $h_i(\theta) = 0$ for $i \neq i(\theta)$.

The allocative margin differs from the intensive margin that is the focus of most of the optimal taxation literature since Mirrlees (1971), including the recent and closely related work of Rothschild and Scheuer (2014a,b, henceforth RS). In particular, the allocative margin involves individuals making non-local changes across discretely different income levels and across professions that involve different externalities. Analysis following Mirrlees, such as that of RS, relies on a "first-order approach" that assumes all changes in income are local.

Following RS, we assume externalities in this economy operate through production. This assumption allows us to consider a richer range of ways in which externalities are targeted across professions, while minimizing the notational burden. In particular, externalities falling uniformly on consumption may be represented in this approach by ensuring externalities fall in a proportionate

[2]Rothschild and Scheuer include a quantitative calibration of the size of the effect they analyze. These results are consistent with our quantitative results, because Rothschild and Scheuer use our numbers on externalities as well as the same set of structural specifications on the ability distribution and the shape of the income distribution. However, their model has only two sectors and no allocative margin, implying much larger changes are possible in our model, depending on the size of these various factors, than in their analysis.

and appropriately (endogeneously) scaled way on all professions.

For all profession pairs i, j, output in profession j can affect the productivity of workers in profession i. These relationships are summarized through nonnegative functions $E_i(Y_1, ..., Y_n)$ as in RS.[3]

The private product of a worker in i is linear in hours worked h_i. Hence, the private product of worker θ in profession i coincides with that worker's income and is given by

$$y_i(\theta) = a_i(\theta)h_i(\theta)E_i(Y_1, ..., Y_n), \tag{1}$$

where $Y_j = \int_\Theta y_j(\theta)f(\theta)d\theta$ is the total output in profession j. When E_i does not depend on Y_j, profession j exerts no externality on profession i. An economy without externalities corresponds to the case in which all E_i are constant.

Worker utility is linear in after-tax income, non-pecuniary utility ψ, and an hours cost function $\phi(\cdot)$ for which $\phi'(\cdot), \phi''(\cdot) > 0$:

$$U(\theta) = y_{i(\theta)}(\theta) - T(y_{i(\theta)}(\theta)) - \phi(h_{i(\theta)}) + \psi_{i(\theta)}(\theta), \tag{2}$$

where $T(\cdot)$ is the tax schedule set by the government. This specification abstracts from income effects, as does much of the recent literature on optimal taxation (Diamond, 1998). This set-up is particularly convenient in our setting, because introducing income effects without adding a redistributive motive would require departing from the simple utilitarian welfare criterion we employ. Although quasi-linear utility is a strong restriction, relaxing it seems likely to reinforce our primary conclusion that positive externalities at middle incomes and negative ones at high incomes call for subsidizing the middle class and taxing the wealthy.[4] In our specification, the cost of effort and the non-pecuniary benefit or cost of a profession are additively separable, thereby ruling out richer interactions between intensive and allocative labor-supply decisions.[5]

We assume the functional form $\phi(h) = h^{1/1+\sigma}/(1/1+\sigma)$, which leads all workers to have the same, constant intensive elasticity of labor supply σ. Each worker takes as given this tax schedule $T(\cdot)$ and the profession outputs $Y_1, ..., Y_n$, and then chooses a profession $i^*(\theta)$ and hours $h^*_{i^*(\theta)}(\theta)$ to maximize

[3]A slight difference exists between our setup and that considered by RS. Whereas our externalities depend on the total *output* of each profession, theirs depend on the hours worked in each profession. Thus, in our model, externalities from profession i to profession j directly amplify the externalities of j, whereas in RS, the i externalities do so only indirectly by drawing labor into j. In most of our calibrations, however, the two specifications do not differ greatly.

[4]In particular, in the presence of income effects, such a pattern would encourage lower earnings and higher non-pecuniary income among all but the highest and lowest earners. The subsidies for entering the middle class would, through both income and substitution effects, encourage the poor to enter the middle class and discourage aspirations above this level by raising the income of the middle class and reducing substitution benefits of aspiring above this level.

[5]However, in our empirical application, we scale the size of ψ to the hours that an individual works under the *laissez-faire* allocation to maintain a balance between the magnitude of these two effects.

utility. To capture the case in which a worker is indifferent between multiple professions, we let $I^*(\theta)$ denote the *set* of professions that maximize the utility of a type-θ worker. When $|I^*(\theta)| > 1$, the worker chooses $i^*(\theta) \in I^*(\theta)$ randomly. We denote the total utility, income, and non-pecuniary utility at the optimal profession and hours choices by $U^*(\theta)$, $y^*(\theta)$, and $\psi^*(\theta)$, respectively. We simplify notation by defining $h^*(\theta) = h^*_{i^*(\theta)}(\theta)$, and also let $U_i^*(\theta)$ and $y_i^*(\theta)$ denote the utility and income resulting from maximizing utility conditional on $i^*(\theta) = i$.

The government must finance a net expenditure of R, and chooses a tax schedule $T(\cdot)$ that maximizes total worker utility while raising this revenue:

$$T = \arg\max_{\overline{T}} \int_\Theta U^*(\theta) f(\theta) d\theta \;\Big|\; \int_\Theta \overline{T}(y^*(\theta)) f(\theta) d\theta \geq R.$$

In our estimation of the optimal income tax in Section 4, we focus on bracketed tax systems that are messy to characterize analytically because they lead to "bunching" of workers with different productivity at the same income. For expositional clarity and comparability with existing literature, in this section, we follow Mirrlees (1971) and Saez (2001) in restricting attention to tax schedules for which an interior solution for hours always exists and is smooth. Analogs to our results here apply to bracketed schedules.[6]

Assumption 1. *The government considers only tax schedules T whose second derivative exists, and such that for all incomes y, $T'(y) < 1$ and*

$$\frac{yT''(y)}{1 - T'(y)} > -\frac{1}{\sigma},$$

where σ is the elasticity of labor supply.

As shown in Appendix A.1, any solution to the worker's first-order condition for hours is a strict local maximum (due to a negative second-order condition) when this inequality holds. As a result, the hours choice admits a unique maximum.

Given the quantitative focus of this paper, we follow RS in assuming the existence of a unique Hicksian stable competitive equilibrium of the economy; our necessary conditions for optimization are valid only for tax schedules that induce such an equilibrium.[7]

[6] For example, an analog of Proposition 1 with averages taken over over the range of incomes affected by each bracket, no intensive margin response exists at upward kinks in the tax schedule, and non-local jumps occur even along the intensive margin around downward kinks in the tax schedule. See Slemrod et al. (1994) for a detailed discussion of related issues in the context of the standard Vickrey model.

[7] Equilibrium existence in economies with externalities in production is a notoriously difficult and unsettled problem in general equilibrium theory outside of very special parametric cases (Chipman, 1970). Intuitively, greater production in one profession might raise wages sufficiently in other professions to induce, through feedback loops operating both directly and through labor supply, an explosive path that can undermine equilibrium uniqueness, stability, and even existence. We know of no primitive conditions on our economy or on those studied by RS guaranteeing uniqueness, existence, or stability.

2.2 The government's first-order condition

The tax schedule T consists of a lump-sum tax T_0 paid by all workers, and a marginal tax schedule $T'(\cdot)$. These two aspects of the tax schedule uniquely determine T by the formula

$$T(y) = T_0 + \int_0^y T'(\overline{y})d\overline{y}. \tag{3}$$

The government chooses T_0 and $T'(\cdot)$ to maximize worker utility while raising revenue R.

The equilibrium allocation of output $Y_1^*, ..., Y_n^*$ depends on $T'(\cdot)$ and not on T_0. Indeed, workers' intensive labor-supply choices depend on $T(\cdot)$ only through $T'(\cdot)$. And their profession choices depend on level differences in utility across professions, which remain constant—due to quasi-linear utility—as the common lump sum grant T_0 changes. This invariance condition means the optimal marginal tax schedule $T'(\cdot)$ cannot depend on R.

Lemma 1. *The optimal* marginal *tax schedule $T'(\cdot)$ is independent of the revenue requirement R.*

Due to Lemma 1, we ignore the revenue requirement in this paper and focus on the choice of the optimal marginal tax schedule $T'(\cdot)$.

To derive the optimal $T'(\cdot)$, we follow the intuitive perturbation approach to calculus of variations pioneered in economics by Wilson (1993) and in optimal income taxation by Saez (2001). Suppose the government slightly raises the marginal tax rate $T'(y)$ by dT' for incomes between y and $y + dy$, and rebates the additional revenue to workers through lowering T_0. This perturbation leaves the total revenue raised by the tax unchanged, but could raise or lower utility by leading workers to adjust their labor supply. At the optimum $T'(\cdot)$, the resulting change to utility is 0.

Raising $T'(y)$ leads to both intensive and allocative labor-supply changes. On the intensive margin, workers for whom $y^*(\theta) = y$ lower their hours $h^*(\theta)$. We denote the set of these workers by $\Theta(y) = \{\theta \mid y^*(\theta) = y\}$, and the set of such workers in profession i by $\Theta_i(y) = \{\theta \mid y^*(\theta)$ and $i^*(\theta) = i\}$. The tax change also lowers the level of after-tax income by $dT'dy$ of all workers earning above y. Therefore, the tax change induces profession switching for workers who are indifferent between a profession in which they earn more than y, and a profession in which they earn less. We denote the set of such workers by

$$\Theta_S(y) \equiv \left\{\theta \;\middle|\; \text{there exist } i_l, i_h \in I^*(\theta) \text{ such that } y^*_{i_l}(\theta) < y < y^*_{i_h}(\theta)\right\}.$$

The perturbation to $T'(\cdot)$ causes additional, secondary labor-supply changes. Due to externalities operating through the E_i, the intensive and allocative margin adjustments just described

In our empirical applications, we always verify local equilibrium stability. We have generally found that in specifications with externalities that are too large, we run into problems. Luckily, this issue does not arise in any of our baseline scenarios, and we have deliberately chosen robustness checks that do not create problems here. We discuss these issues and some conjectures derived from these empirical results further in Appendix A.1.

change the productivity in all professions, leading all workers to modify their labor supply. Sufficient statistics that we term *externality ratios* capture the resulting changes to aggregate utility. The externality ratio e_i of profession i equals

$$e_i \equiv \frac{\partial}{\partial Y_i} \int_\Theta U^*(\theta) f(\theta) d\theta,$$

where the partial derivative denotes the cumulative effect on welfare through changes in the E_j that result from a change in Y_i. Thus, the externality ratio of a profession gives the marginal externality of a dollar earned in that profession. It can be positive or negative. When a profession causes no externalities, $\partial E_j / \partial Y_i \equiv 0$ for all j, so the externality ratio equals 0.

This ratio is a central yet subtle object in our analysis, so we describe its meaning and derivation in some detail. A change in Y_i induces a series of subsequent changes. The direct effect of a change in Y_i is to alter the productivity in all professions. These productivity changes lead to adjustments in labor supply on both the intensive and allocative margins: workers choose to work more or less and also may choose different professions entirely. These labor-supply responses change the output Y_j in each profession, inducing another round of adjustments in labor supply, which beget yet another round of adjustments, and so on. Externality ratios solve the fixed-point problem that captures this infinite series of labor-supply adjustments. The solution is local to the equilibrium under consideration. In Appendix A.1, we explicitly solve this problem to express the e_i in terms of the Jacobian of the externality function E at the equilibrium $(Y_1^*, ..., Y_n^*)$ and the full set of labor-supply responses.

The average externality ratio of workers earning y is

$$e(y) = \frac{\sum_{i=1}^n e_i \int_{\Theta_i(y)} f(\theta) d\theta}{\int_{\Theta(y)} f(\theta) d\theta}.$$

Using the definition of worker utility (2) and the revenue requirement $\int_\Theta T(y^*(\theta)) f(\theta) d\theta = R$, we write the government's objective function as

$$\int_\Theta U^*(\theta) f(\theta) d\theta = -R + \int_\Theta \left(y^*(\theta) - \phi(h^*(\theta)) + \psi^*(\theta)\right) f(\theta) d\theta. \tag{4}$$

The government maximizes the integral on the right: total income less disutility from labor plus non-pecuniary utility from work. We calculate how the perturbation to $T'(\cdot)$ at y changes this integral. We first consider the change from intensive margin labor-supply adjustments. Then we separately consider how allocative margin adjustments change utility, and finally we present the first-order condition that combines these effects.

2.2.1 Intensive margin

Consider a worker in profession i for whom $y^*(\theta) = y$. Denote the *wage* of this worker by $w_i(\theta) = a_i(\theta)E_i(Y_1^*, ..., Y_n^*)$. The hours for this worker are determined by $h^*(\theta)^{1/\sigma} = w_i(\theta) (1 - T'(w_i(\theta)h^*(\theta)))$. For h near $h^*(\theta)$, the relationship $T'(w_i(\theta)h) = T'(y) + w_i(\theta)(h - h^*(\theta))T''(y)$ holds to the first order. Using this first-order expansion, we totally differentiate the hours equation with respect to $T'(y)$ to find that

$$dh^*(\theta) = \frac{\sigma h^*(\theta)}{1 - T'(y) + \sigma y T''(y)} dT'.$$

This intensive-margin response directly changes the type-θ worker's contribution to (4), and also alters the income of other workers through an externality. The direct effect is $(w_i(\theta) - \phi'(h^*(\theta)))dh^*(\theta)$. Because $\phi'(h^*(\theta)) = w_i(\theta)(1-T'(y))$, the direct effect reduces to $T'(y)w_i(\theta)dh^*(\theta)$. To uncover the externality, note $dY_i^* = w_i(\theta)dh^*(\theta)$, so the externality equals $e_i w_i(\theta)dh^*(\theta)$. We sum the direct and externality effects on utility across all workers earning y to obtain the complete change in the government's objective from intensive margin adjustments. Let $f(y) = \int_{\Theta(y)} f(\theta)d\theta$; the mass of workers earning between y and $y + dy$ is $f(y)dy$. The complete intensive-margin change in the government objective from the perturbation to the tax schedule is

$$\partial^{int} \int_\Theta U^*(\theta)f(\theta)d\theta = \frac{\sigma y f(y)}{1 - T'(y) + \sigma y T''(y)} \left(T'(y) + e(y)\right) dT'dy. \tag{5}$$

2.2.2 Allocative margin

Because it involves discrete changes in income, the allocative margin is similar to the extensive margin of exit out of the labor force studied by Saez (2002). However, the choice to exit the labor force is (mathematically) simpler than the choice of a profession. Higher marginal taxes always make exiting the labor force more attractive, and the point in the labor market to which individuals transition upon exit is always the same. The allocative margin involves richer effects, as different individuals substitute to different professions with different externality ratios depending on where in the tax schedule marginal rates are changed.

To see these dynamics more precisely, consider a worker for whom $\theta \in \Theta_S(y)$. This worker is indifferent between a profession i_h in which she earns $y_{i_h}^*(\theta)$, and a profession i_l in which she earns $y_{i_l}^*(\theta)$, with $y_{i_l}^*(\theta) < y < y_{i_h}^*(\theta)$. The tax perturbation decreases the after-tax income, and hence utility, in i_h by $dT'dy$ while leaving utility in i_l unchanged. As a result, the worker switches from i_h to i_l.

This switch directly changes the value of the government's objective function (4) by $y_{i_l}^*(\theta) - \phi(h_{i_l}^*(\theta)) + \psi_{i_l}^*(\theta) - \left(y_{i_h}^*(\theta) - \phi(h_{i_h}^*(\theta)) + \psi_{i_h}^*(\theta)\right)$. By the envelope theorem (because the worker receives the same utility in i_l and i_h), this difference equals the fiscal externality $T(y_{i,l}^*(\theta)) - T(y_{i_h}^*(\theta))$.

We define the average proportional tax change from switching workers by

$$\Delta_T(y) \equiv \int_{\Theta_S(y)} \frac{T(y^*_{i_h}(\theta)) - T(y^*_{i_l}(\theta))}{y} \frac{f(\theta)}{f_S(y)} d\theta,$$

where $f_S(y) = \int_{\Theta_S(y)} f(\theta) d\theta$ is the density of switching workers.

The worker's switch from i_h to i_l also changes the government's objective function through externalities. The worker's presence in profession i increases Y^*_i by $dY^*_i = y^*_i(\theta)$, so the total externality of a worker's presence in i is $e_i y^*_i(\theta)$. The change in externalities from switching from i_h to i_l is therefore $e_{i_l} y^*_{i_l}(\theta) - e_{i_h} y^*_{i_h}(\theta)$. We define the average proportional externality change from switching workers by

$$\Delta_e(y) \equiv \int_{\Theta_S(y)} \frac{e_{i_h} y^*_{i_h}(\theta) - e_{i_l} y^*_{i_l}(\theta)}{y} \frac{f(\theta)}{f_S(y)} d\theta.$$

Recall these externalities incorporate *all* of the indirect, general equilibrium effects of production in a profession.

We sum the direct and externality effects on utility across all switching workers to obtain the complete change in the government's objective from allocative margin adjustments. Because the change in the relative income of i_h and i_l is $dT'dy$, the completely allocative margin change in the government objective from the perturbation to the tax schedule is

$$\partial^{all} \int_{\Theta} U^*(\theta) f(\theta) d\theta = y f_S(y) \left(\Delta_T(y) + \Delta_e(y)\right) dT'dy. \tag{6}$$

Note the distinction between this formula and that arising in Saez (2002)'s analysis, where this term consists only of aggregate taxes paid by each individual and the only relevant densities are those of exiting the labor force.

2.2.3 Total first-order condition

The government's first-order condition holds when the intensive margin change (5) and allocative margin change (6) to the government's objective resulting from the tax perturbation sum to 0. Because we arbitrarily chose the income y at which $T'(\cdot)$ was perturbed, the first-order condition holds for all y. Proposition 1 produces the first-order condition by adding (5) and (6) and then dividing by $yf(y)dT'dy$.

Proposition 1. *The optimal tax schedule T for the government satisfies the equation*

$$0 = \underbrace{\frac{\sigma f(y)}{1 - T'(y) + \sigma y T''(y)} \left(T'(y) + e(y)\right)}_{intensive} + \underbrace{f_S(y) \left(\Delta_T(y) + \Delta_e(y)\right)}_{allocative} \tag{7}$$

for all incomes y. Here σ is the elasticity of labor supply, $e(y)$ is the average externality ratio of *output for workers earning y, $f(y)$ is the measure of workers earning y, $f_S(y)$ is the measure of workers indifferent between earning above y in one profession and below y in another, $\Delta_T(y)$ is the average proportional difference in taxes between the two professions for such workers, and $\Delta_e(y)$ is the average proportional difference in externalities between the two professions for such workers.*

The optimal tax T is Pigouvian, because it offsets externalities on both the intensive and allocative margins. Without externalities, $e(y)$ and $\Delta_e(y)$ globally equal 0, in which case the optimal tax given by Proposition 1 is lump sum ($T' \equiv 0$). We build further intuition by considering the intensive and allocative margins separately.

When only the intensive margin is present, the optimal tax satisfies $T'(y) = -e(y)$. In this case, the marginal tax rate exactly equals the average negative externality ratio at each income level. RS refer to this tax as the "Pigouvian" correction because it appears in a model with only an intensive margin. In particular, the weight of this effect in the total first-order condition scales with $\sigma f(y)$, the product of the intensive labor-supply elasticity and the number of individuals subject to this elasticity. The greater this product is, the more closely the optimal tax satisfies $T'(y) = -e(y)$.

Conversely, the optimal tax in the presence of just the allocative margin satisfies $\Delta_T(y) = -\Delta_e(y)$ for all y. In this case, taxes offset gross changes in negative externalities from workers switching professions. The weight of this effect scales with $f_S(y)$, the measure of the workers who switch profession around y. The more sensitive profession choices are to income differences, the greater $f_S(y)$ becomes and the more closely the optimal tax satisfies $\Delta_T(y) = -\Delta_e(y)$.

Note the optimal tax is related only to the *relative size* of the allocative and intensive responses,

$$\frac{f_S(y)\left[1 - T'(y) + \sigma y T''(y)\right]}{\sigma f(y)},$$

and not to the level of these responses. For example (assuming a linear tax for the moment), suppose σ and f_S doubled so that the size of both the intensive and allocative responses were twice as large. This doubling would have no impact on optimal taxes, in sharp contrast to the standard Vickrey model whereby a redistributive state is constrained in its ability to extract revenue by the overall elasticity of taxable income.

3 An Example with Three Professions

This section builds quantitative intuition in closed form for the full calibration in a simple example that captures the key features of the data and our estimation. In particular, we use Proposition 1 to calculate the optimal top tax rate, $\lim_{y\to\infty} T'(y)$, for the optimal T. This rate measures the marginal tax rate the top earners face (although possibly only at extremely high incomes) and has been explored by other papers that derive optimal income-tax schedules (Saez, 2001; Saez et al., 2012).[8] Given our focus on the allocation of talented individuals, many of whom earn very high incomes, this limiting rate seems particularly relevant in our context.

3.1 Specification and optimal top tax rate

Three professions exist: U, H, and L. Some fraction of the workers are "unskilled" and are restricted to U. The remaining workers are "skilled" and choose between H and L.[9] For each skilled worker, productivity $a_h(\theta)$ in H exceeds productivity $a_l(\theta)$ in L by a constant multiple $r^{1/(1+\sigma)}$, where $r > 1$, which leads in equilibrium to income that is higher in H than in L by a factor of r. Past some point $\overline{a}$, the distribution of a_i in each profession is Pareto, with conditional probability distribution $\Pr(a_i(\theta) \geq a \mid a_i(\theta) \geq \overline{a}) = (\overline{a}/a)^{\alpha(1+\sigma)}$ for some $\alpha > 0$; in equilibrium, the Pareto exponent for the *income* distribution will equal α. For skilled workers, non-pecuniary utility ψ_i of working in $i = H$ or L is distributed as $\psi_i \mid a \sim \beta^{-1}[(a_l^{1+\sigma} + a_h^{1+\sigma})/2](\overline{\psi}_i + F_\psi)$, where the $\overline{\psi}_i$ are constants and F_ψ is a standard Gumbel distribution given by $F_\psi = e^{-e^{-\psi}}$. The fraction is a normalization to ensure non-pecuniary utility is of the same order of magnitude as income, and $\beta > 0$ is a parameter we call the *allocative sensitivity*. Output in U causes no externality, whereas H and L output both affect productivity in U. Thus, E_l and E_h are equal to 1, whereas $E_u(Y_l, Y_h)$ depends on the output of the skilled professions. E_u increases in Y_l and decreases in Y_h, so $e_h < 0 < e_l$.

This specification broadly matches the data we present in Section 4. In our baseline analysis, engineering, teaching, and research professions cause positive externalities, whereas law and finance lead to negative externalities. We find the incomes in the first set of professions are lower than those in the second in the upper tail of the income distribution, which we assume is Pareto in line with a large empirical literature. The conditional ψ_i distributions follow the Gumbel distributions given above with the same normalizations.

The present specification allows us to explicitly calculate the optimal top tax rate in the special cases in which only the intensive or allocative labor-supply margin operates. We first analyze the intensive optimal top tax rate. From Proposition 1, this rate satisfies $\tau_{int} = -\lim_{y\to\infty} e(y)$. Hence,

$$\tau_{int} = -(s_h e_h + s_l e_l), \tag{8}$$

where e_h and e_l are the externality ratios and s_i is the share of workers at top incomes in profession

[8] Technically, Saez (2001) calculates the optimal constant rate above a given high but finite threshold of earnings. He does so by imposing by assumption the limiting property that the distribution has Pareto tails beyond a finite threshold and that beyond this threshold, the social planner places no weight on the utility of sufficiently rich individuals. Both assumptions are, in standard models of the income distribution and of social welfare, limiting properties for high-enough income. Thus, effectively, Saez's analysis focuses on the limiting optimal tax. Our focus on the limit is from slightly different sources. Workers earning income near the bottom threshold may decide to switch to a profession earning income lower than that threshold (and hence outside the interval) were the top rate to change. Therefore, we cannot precisely characterize the optimal top rate on such an interval and must resort to a limiting argument.

[9] Formally, $\psi_h(\theta) = \psi_l(\theta) = -\infty$ for the unskilled workers and $\psi_u(\theta) = -\infty$ for the skilled workers.

i.[10] This tax is more positive when the share s_h of top earners in H is higher and when the negative externality e_h is larger in magnitude. Conversely, the intensive optimal top tax rate is less positive when s_l is larger and when e_l is greater. The rate τ_{int}, dubbed the "Pigouvian correction" by RS, is optimal when profession choices are fixed.

The allocative optimal top rate looks quite different from τ_{int}. From Proposition 1, $\Delta_T(y) + \Delta_e(y) = 0$ for high incomes at this rate. These difference terms are determined by the relative income for the same skilled worker in H and L, rather than by the distribution of workers earning any given income. Because $y_i^*(\theta) = a_i^{1+\sigma}(\theta)(1 - T'(y_i^*(\theta)))^\sigma$, $y_h^*(\theta) = ry_l^*(\theta)$ at high incomes.[11] The parameter r equals the ratio of income in H to income in L for a skilled worker. Therefore, each switching worker's contribution to $\Delta_T(y)$ is $\tau(r-1)y_l^*(\theta)$ and to $\Delta_e(y)$ is $(re_h - e_l)y_l^*(\theta)$, where τ is the top tax rate. The optimum sums these to 0, and is

$$\tau_{all} = -\frac{re_h - e_l}{r - 1}. \tag{9}$$

Intuitively, this rate equals the change in negative externalities from a switching worker divided by the change in that worker's income. Although it is the allocative margin analogue of RS's Pigouvian correction, it often behaves very differently quantitatively. In particular, τ_{all} is unambiguously positive because L produces positive externalities and H causes negative externalities ($e_l > 0 > e_h$). This result stands in contrast to τ_{int}, which could be positive or negative.

The size of τ_{all} is greater when e_h or e_l is greater in magnitude. Unlike τ_{int}, τ_{all} depends not on share of the population in H and L but on r, the ratio of income in H to L for a given worker. Simple differentiation shows it to be strictly decreasing in r so long as $e_h < e_l$. To see this relationship between τ_{all} and r dramatically, note that as $r \to 1$, a switching worker is indifferent between working in H and L and both yield the same income and therefore tax revenue. However, a switch to L increases social welfare by $e_l - e_h$ times the worker's income, so the τ_{all} becomes arbitrarily large to compensate this discrete change in externalities with a discrete change in tax revenue accrued over a very small difference in incomes.

The true optimal tax τ^* combines the logic of both τ_{int} and τ_{all} and is always strictly between these two rates, as we show in Appendix A.2.

3.2 Calibration

To calculate τ^*, we need values of α, σ, β, r, e_h, e_l, $\overline{\psi}_h - \overline{\psi}_l$, and the share of workers that are skilled. We take these values from the data used in the estimation in the next section and thus discuss our calibration choices only briefly here and expand on this discussion in Appendix B. We also explore the different values of τ^* generated by a reasonable range of the parameters.

[10] Formally, $s_i = \lim_{y\to\infty} \int_{\Theta_i(y)} f(\theta)d\theta / \int_{\Theta(y)} f(\theta)d\theta$.

[11] This statement requires $\lim_{y\to\infty} T'(y) < 1$ or $\sigma = 0$.

We first set α, the Pareto parameter for the tail of the US income distribution, to 1.5 based on the ratio of the total income earned by the top 1% of the US income distribution to the 99^{th} percentile of the income distribution. We set $\sigma = 0.24$ and $\beta = 1.5$ based on our estimation in the next section, where we try to match the elasticity of income with respect to the tax rate (Chetty, 2012) and the concurrent growth in relative finance wages and employment from 1980 to 2005 (Philippon and Reshef, 2012). To calculate r, we compare the incomes in H and L at the same percentiles of the profession-specific distributions. Because productivity in H and L are perfectly correlated, a worker in the 99^{th} percentile of H incomes will also be in the 99^{th} percentile of L incomes. We use a 99^{th} percentile income in H (finance and law) of \$1,900,000 and in L (engineering, research and teaching) of \$400,000 based on a weighted-average of our profession-specific income-distribution estimations across the professions that make up H and L.

We next choose the externality ratios e_h and e_l. These ratios are not parameters but endogenous statistics describing an equilibrium allocation of labor. For the purposes of deriving the top tax rate, we may take e_h and e_l as given—the top tax rate determines labor allocations only at the very top, and hence does not materially change the economy-wide labor allocation that determines e_h and e_l.

To calculate the externality ratios, we take a weighted average of approximate externality ratios of the professions constituting each of H and L. As we discuss in Appendix B, dividing a profession's aggregate spillover by its aggregate income provides an accurate approximation of its externality ratio. In Section 4, we estimate these aggregate spillovers by drawing on the economics literature, and we calculate the aggregate incomes using data on profession-specific income distributions and worker counts from the Bureau of Labor Statistics (BLS) and the Internal Revenue Service (IRS). These figures yield approximate externality ratios of -0.33 for finance, -0.10 for law, 0.15 for engineering, 9.28 for research, and 2.28 for teaching. An average using weights proportional to the representation of these professions at high incomes in the data then yields $e_h = -0.24$ and $e_l = 2.67$.

Finally, we set $\overline{\psi}_h - \overline{\psi}_l$ to match the share of workers in H and L, given the data and the tax rate in 2005. According to Bakija et al. (2012), 7.2% of the top 1% of earners in 2005 were in L and 22.3% were in H.

Using these parameters and externality ratios, we calculate the optimal top tax rate to be $\tau^* = 0.24$. Relative to a *laissez-faire* tax rate of 0, τ^* induces 11% more of skilled workers subject to the tax rate to choose the lower-paying but higher-externality profession L. To break down the top tax rate, we calculate τ_{int} and τ_{all} at the optimum. When $\tau = \tau^*$, $s_h = 0.18$ and $s_l = 0.08$, leading to an intensive optimal tax rate of $\tau_{int} = -0.17$. Thus, the intensive optimal rate is negative, even though the total optimal rate is positive. The negative τ_{int} results because the order-of-magnitude higher externalities from L overwhelm the negative externalities from H, because H has only three times greater representation at high incomes. By contrast, $\tau_{all} = 1.03$, confiscating more than all of the marginal income of top earners. The total optimum τ^* balances the intensive and allocative

TABLE 1
Optimal Top Tax Rate for Different Parameter Values

	r	σ	β	e_h	e_l
Half of Baseline	26%	34%	15%	20%	20%
Baseline	24%	24%	24%	24%	24%
Double Baseline	19%	15%	30%	31%	28%

Notes: This table reports the optimal top tax rate τ^* in this section's example. In each column, we hold all but the header parameter constant, varying that parameter to 50%, 100%, and 200% of its baseline value and reporting the optimal top tax rates. The baseline parameters are $r = 4.7$, $\sigma = 0.24$, $\beta = 1.5$, $e_h = -0.24$, and $e_l = 2.67$.

optima at a rate of 0.24. This rate is reasonably close to the top tax rate in our full estimation of 0.37.

Table 1 reports the sensitivity of τ^* to changes in the parameters. For each parameter, we recalculate τ^* using values at half and double the baseline, while holding the other parameters constant.[12] The results confirm the intuition discussed above. Higher values of r lower optimal rates, as profession switching generates smaller positive externalities relative to lost tax revenue when r is greater. Higher values of σ lower optimal rates, as a greater σ makes the intensive margin more important, and the intensive optimal tax rate is negative. Similarly, a greater β increases the optimal top rate as it makes the allocative margin more important. Finally, higher absolute values of the externalities increase the optimal top tax rate by increasing the efficiency gains from switches. Raising the negative externality in H has a greater impact than raising it in L, despite the much greater magnitude of the externality in L. Intuitively, a greater externality in H raises both τ_{all} and τ_{int}, but increasing the positive externality of L lowers τ_{int} while raising τ_{all} resulting in a more ambiguous effect on $\tau^\star$.

4 Empirical Strategy

In this section, we specify the richest version of the full model that we can meaningfully estimate and we fit it to data from the United States in 2005. Appendix A.3 includes all derivations and proofs, and Appendix C summarizes additional empirical details.

[12] Thus, as σ moves, the income ratio r and the Pareto parameter α for income stay constant. These quantities are calibrated to match observed data, so we do not want them to change as σ moves. Our specification allows r and α to stay constant as σ moves, by involving σ in the distribution of skilled productivity.

4.1 Specification

Our specification of f is as follows. We separate the professions into n skilled professions $i = 1, ..., n$ and 1 low-skilled profession, which we index by $i = 0$. An exogenous share s_0 of the workers are "low-skill" and always choose $i^*(\theta) = 0$, because $\psi_i(\theta) = -\infty$ for $i > 0$.

The remaining $1 - s_0$ of the workers are "skilled" and choose only among the skilled professions because $\psi_0(\theta) = -\infty$. This low-skilled/skilled dichotomy represents the view we used to motivate our analysis that professions creating externalities are primarily confined to talented workers. These workers are more mobile across professions given their greater general education (Murphy, 1986), and most professional externalities identified in the literature are confined to professions dominated by such workers.

Each worker's productivity a_i in i is drawn from a profession-specific distribution F_i^a. We specify the correlation structure of productivity draws for skilled workers by a *Gaussian copula*:

$$f(a_1, ..., a_n) = f_{0,\Sigma}^{\mathcal{N}} \left(\Phi^{-1}(F_1^a(a_1)), ..., \Phi^{-1}(F_n^a(a_n)) \right),$$

where Φ is the CDF of a unidimensional standard normal and $f_{\mu,\Sigma}^{\mathcal{N}}$ is the PDF of a multivariate normal with mean μ and covariance matrix Σ. This specification preserves the marginal productivity distributions F_i^a (i.e., $f \mid_{a_i} = f_i^a$ for all i), but allows correlation specified by Σ. The richest specification we can estimate has a single parameter $\rho \leq 1$ for all off-diagonal elements of Σ governing the correlation of productivity between every distinct pair of professions; all on-diagonal elements are 1. We denote this covariance matrix as Σ_n. When $\rho = 1$, productivity across professions is perfectly correlated so that workers are characterized by a single "talent" parameter that determines their percentile in each profession's productivity distribution. Smaller values of ρ allow sorting on comparative advantage, in which the workers who choose i are those who are most productive in i relative to the other professions, as suggested by the empirical work of Reyes et al. (2013) and Kirkebøen et al. (Forthcoming).

Conditional on the productivity vector $a = (a_1, ..., a_n)$, each preference ψ_i is drawn independently from the distribution

$$\psi_i \sim \beta^{-1} \left(\frac{1}{n} \sum_{j=1}^{n} a_j^{1+\sigma} \right) \left(\overline{\psi}_i + F_\psi \right),$$

where $\overline{\psi}_i$ is a constant and F_ψ is a standard Gumbel distribution given by $F_\psi = e^{-e^{-\psi}}$, generating a standard logit discrete-choice model among individuals with a given ability. The normalization by productivity keeps professional choice scale-invariant with respect to income.[13] Thus, we can interpret β as a uniform-across-ability-levels *allocative sensitivity*, with higher β indicating greater elasticity of profession choice to changes in relative incomes across professions and thus to taxation.

[13]This property holds exactly when taxes are linear and approximately otherwise. Only relative income $y_i / \sum_j y_j$ matters in (13) when $T'(\cdot)$ is constant.

The constants $\overline{\psi}_i$ determine the average relative attractiveness of each profession i; more workers enter i when $\overline{\psi}_i$ is higher.

Our specification of each externality function E_i has the form

$$E_i(Y_0, ..., Y_n) = \prod_{j=0}^{n}(1 + \epsilon_{i,j} Y_j^{\gamma}).$$

γ captures the returns to scale of the externalities; $\gamma = 1$ implies externalities are linear in output; lower values of γ lead to diminishing marginal returns. $\epsilon_{i,j}$ captures the targeting of externalities across professions emphasized by RS. For our estimation, we reduce the dimensionality of these coefficients according to the specification

$$\epsilon_{i,j} = \delta_{i,j}\epsilon_j.$$

For $i \neq j$, $\delta_{i,j}$ equals 1 if j affects output in i and 0 otherwise. The $\epsilon_{i,i}$ remain unrestricted, allowing independence of the own externalities from those on other professions. This independence allows external economies and diseconomies of scale as in Marshall (1890) and Chipman (1970). We restrict externalities coming from profession j to be uniform in magnitude across all professions i on which it has any impact.

The sources for all these inputs are discussed in Section 4.3 below.

4.2 Identification

This section discusses the identification of f and E. The empirical inputs into our estimation are the existing tax schedule T_{2005}, the distributions of income $f_0^y, ..., f_n^y$, the population shares in each profession $s_0, ..., s_n$, and the marginal social products $\partial Y/\partial Y_0, ..., \partial Y/\partial Y_n$ of output in each profession.[14] These inputs come from data we describe in Section 4.3. For the moment, we take the parameters σ, β, ρ, γ, and the matrix $\{\delta_{i,j}\}$ as given, postponing discussion of their selection until Sections 4.3.2 and 4.3.3. The outputs of the present estimation are $f_0^a, ..., f_n^a$, $\overline{\psi}_1, ...\overline{\psi}_n$, and $\epsilon_0, ..., \epsilon_n$.

First, we calculate the aggregate income in each profession and in the economy. For each i, $Y_i = s_i \int_0^{\infty} y f_i^y(y)dy$, and $Y = \sum_{i=0}^{n} Y_i$.

Next, we calculate the externality coefficients $\epsilon_0, ..., \epsilon_n$ using the aggregate income data and the marginal social product measures $\partial Y/\partial Y_j$. As we define it, this derivative gives the cumulative

[14]Rather than use the true non-linear value of T_{2005}, we use a linear approximation in which the marginal tax rate is constant ($T'_{2005} = 0.3$). The true tax schedule T_{2005} features discontinuous marginal rates. Therefore, in a model such as ours in which primitives are smooth and workers are fully optimizing, bunching would result in the income distributions. Because empirical income distributions are smooth, we cannot fit underlying skill distributions to the empirical income distributions using the true T_{2005}. Using the linear version allows us to fit the skill distributions. A a number of optimal tax papers take a similar approach, including Saez (2001, 2002).

increase in the economy's output from a unit increase in output in j, holding labor supply constant in the entire economy. The change to Y_j can be thought of as coming from a small shock to productivity in that profession. As with the externality ratios, the marginal social product includes feedback effects: an increase in Y_j alters output of all professions, inducing further changes to output in the economy and so on. As we show in Appendix A.3, the marginal social product equals

$$\frac{\partial Y}{\partial Y_j} = 1'(I - J)^{-1} 1_j, \tag{10}$$

where $1 = (1, ..., 1)'$, $1_j = (0, ..., 1, ..., 0)'$ with 1 in just the j^{th} spot, I is the identity matrix, and J is the quasi-Jacobian matrix

$$J = \left\{ \frac{Y_i}{Y_k} \frac{\gamma \delta_{i,k} \epsilon_k Y_k^{\gamma}}{1 + \delta_{i,k} \epsilon_k Y_k^{\gamma}} \right\}_{i,k}.$$

Note that when externalities are absent from the economy, $J = 0$ so $\partial Y / \partial Y_j = 1$ for each j: marginal social product coincides with marginal private product. Equation (10) delivers $n + 1$ equations in the $n + 1$ unknowns $\epsilon_0, ..., \epsilon_n$, allowing us to solve for these parameters numerically.

The subsequent step is to infer the empirical productivity distributions $\widetilde{f}_i^a$ that appear in the data. Selection of workers across professions determines these distributions, and hence the $\widetilde{f}_i^a$ differ from the underlying productivity distributions f_i^a we eventually estimate. The following equation delivers a one-to-one mapping between the productivity a_i of a worker in i and her income y_i[15]:

$$a_i = y_i^{\frac{1}{1+\sigma}} \left(1 - T'_{2005}(y_i)\right)^{-\frac{\sigma}{1+\sigma}} E_i(Y_0, ..., Y_n)^{-1}. \tag{11}$$

We define $y_i(a_i)$ to be the unique value of y_i that solves this equation given a_i. Then

$$\widetilde{f}_i^a(a_i) = y_i'(a_i) f_i^y(y_i(a_i)). \tag{12}$$

No selection occurs into or out of the low-skilled profession $i = 0$, so $f_0^a = \widetilde{f}_i^a$.

The penultimate step is to calculate the relative utility $\widetilde{u}_i(a)$ of working in i for a skilled worker with productivity vector a, ignoring profession-preference utility ψ. The relative utility $\widetilde{u}_i(a)$ determines the share of workers with productivity a who choose to work in profession i. It is defined as $\widetilde{u}_i(a) = \left(U_i^*(\theta) - \psi_i(\theta)\right) / \left(n^{-1} \sum_j a_j^{1+\sigma}\right)$, where the productivity component of θ equals a. Appendix A.3 derives the following closed-form expression for relative utility:

$$\widetilde{u}_i(a) = \frac{y_i(a_i) - T_{2005}(y_i(a_i)) + \sigma\left(y_i(a_i) T'_{2005}(y_i(a_i)) - T_{2005}(y_i(a_i))\right)}{(1+\sigma) n^{-1} \sum_j y_j(a_j)(1 - T'_{2005}(y_j(a_j)))^{-\sigma} E_j(Y_0, ..., Y_n)^{-(1+\sigma)}}. \tag{13}$$

Finally, we derive the conditional distribution of a_{-i} given a_i. We use this conditional distribution to back out the underlying productivity distributions f_i^a from the empirical distributions $\widetilde{f}_i^a$,

[15] We can define $y_i(a_i)$ because the right side of (11) strictly increases in y due to Assumption 1.

which are affected by selection. Given $a_i = a$, the conditional distribution of a_{-i} follows a Gaussian copula. The $\Phi^{-1}(F_j^a(a_j))$ for $j \neq i$ are distributed as a multivariate normal with mean $\Phi^{-1}(F_i^a(a_i))\varrho$ and covariance $\Sigma_{n-1} - \varrho'\varrho$, where $\varrho = (\rho, ..., \rho)$ is a $1 \times (n-1)$ vector. We now state the equations that allow us to identify underlying productivity f_i^a and profession preferences $\overline{\psi}_i$ from the data.

Lemma 2. *Given empirical population shares $s_0, s_1, ..., s_n$ and income distributions $f_1^y, ..., f_n^y$, the underlying productivity distributions $f_1^a, ..., f_n^a$ and profession-preference parameters $\overline{\psi}_1, ..., \overline{\psi}_n$ solve the n functional equations*

$$\frac{s_i \widetilde{f}_i^a(a_i)}{1 - s_0} = f_i^a(a_i) \int_{\mathbb{R}_+^{n-1}} \frac{e^{\beta \widetilde{u}_i(a) + \overline{\psi}_i}}{\sum_j e^{\beta \widetilde{u}_j(a) + \overline{\psi}_j}} f^{\mathcal{N}}_{\Phi^{-1}(F_i^a(a_i))\varrho, \Sigma_{n-1} - \varrho'\varrho} \left(\Phi^{-1}(F_1^a(a_1)), ..., \Phi^{-1}(F_n^a(a_n))\right) da_{-i} \tag{14}$$

for $1 \leq i \leq n$ and all $a_i > 0$. Here, $\widetilde{f}_i^a$ is the empirical productivity distribution in i calculated from (12) *and $\widetilde{u}_i(a)$ is the relative utility of working in i for a worker with productivity vector a calculated from* (13)*. These solutions uniquely determine the f_i^a and are unique up to constant for the $\overline{\psi}_i$.*

We solve equation (14) using a numerical solver.

4.3 Data

4.3.1 Income distributions

We follow the classifications of Bakija et al. (2012), whose data we use, in partitioning all US workers into one low-skill profession, which we deem *Other*, and 11 high-skill professions: *Art* (artists, entertainers, writers, and athletes), *Engineering* (computer programmers and engineers), *Finance* (financial managers, financial analysts, financial advisers, and securities traders), *Law* (lawyers and judges), *Management* (executives and managers), *Medicine* (doctors and dentists), *Operations* (consultants and IT professionals), *Real Estate* (brokers, property managers, and appraisers), *Research* (professors and scientists), *Sales* (sales representatives and advertising and insurance agents), and *Teaching* (primary and secondary school teachers). For each profession i, we calculate the share s_i of workers in that profession as well as the empirical distribution of pre-tax income f_i^y in 2005 using two sources of data and several parametric assumptions.

Data on the top of each income distribution come from income-tax filings reported to the IRS. The IRS uses the self-reported profession on personal tax returns (1040s) to assign each filer a Standard Occupation Code (SOC). Bakija et al. (2012) aggregate these codes into the 11 professions we use; we report this classification in Appendix C.1.[16]

[16] Bakija et al. (2012) classify executives in the finance profession (NAICS = 52) as *finance* rather than *management*. Their data on NAICS codes come from examining the profession of the employer on the W-2 of each filer. The BLS provides income and employment data for SOC-NAICS pairs, thus enabling us to match the IRS profession classification. We use the BLS NAICS data only to extract executives in finance. See Appendix C.1.

Their unit of observation is a tax return, of which there are 145,881,000 in 2005.[17] They define the profession of a tax return as that of the primary filer, which is the filer whose social security number is listed first in the case of couples. Bakija et al. (2012) report the number of workers in each profession earning more than \$280,000 and \$1,200,000, as well as the average income of each group of workers above these thresholds.[18]

For each SOC, the BLS reports in the annual Occupational Employment Statistics (OES) database the number of workers as well as the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} income percentiles. The BLS produces the OES using surveys of non-farm establishments. Using these data, we calculate the number of workers in each profession by summing the number in each constituent SOC, and then calculate s_i as the share of all workers in each profession.

To calculate the profession-specific income distribution f_i^y we first assume that for $y \geq$ \$1,200,000, the income distribution is Pareto: $f_i^y(y) = \alpha_i m_i^{\alpha_i}/y^{\alpha_i+1}$. We can uniquely compute the parameters of the Pareto distribution using the mean income of workers earning above this threshold and the number of such workers, both of which are reported by Bakija et al. (2012). Next, we linearly interpolate f_i^y between \$280,000 and \$1,200,000, adding a break point at \$580,000, the geometric average of these income cutoffs.[19] Finally, we solve for the income distribution below \$280,000 under the parametric assumption that over this range, incomes within each profession follow a *Pareto-lognormal distribution* (Colombi, 1990). We denote the standard pdf of a Pareto-lognormal by $f_{\alpha,\mu,\nu}$. This smooth distribution approximates a lognormal with parameters μ and ν at low incomes and a Pareto with parameter α at high values and therefore does a good job of matching both the central tendency and upper tail of the income distribution.

Our precise parametric assumption is that $f_i^y(y) = A_i f_{\alpha_i,\mu_i,\nu_i}(y)$ for $y \leq$ \$280,000. We choose A_i, α_i, μ_i, and ν_i to maximize the likelihood of observing the BLS data, conditional on f_i^y taking the form already estimated for $y \geq$ \$280,000 and conditional on continuity at $y =$ \$280,000. Specifically, for each profession i, the BLS partitions all workers in i into income bins. These bins can be written as $\{s_{i,k}, y_{i,k}^-, y_{i,k}^+\}$, where k indexes the constituent SOCs in i, and $s_{i,k}$ workers in i have incomes in

[17] Bakija et al. (2012) obtain this count from Piketty and Saez (2003), who report this number in an updated table at `http://elsa.berkeley.edu/~saez/TabFig2010.xls`.

[18] To be more precise, Bakija et al. (2012) report data that allow direct computation of these statistics. They report the share of tax returns in the top 1% and top 0.1% in each profession, and write that these income cutoffs are \$280,000 and \$1,200,000, respectively, in 2005 dollars. Because we know the number of tax returns, we can directly compute the number of workers in each profession earning more than each cutoff. Similarly, they report the share of aggregate reported income in the United States earned by workers in each profession in the top 1% and top 0.1% of the income distribution. The aggregate income number comes from Piketty and Saez (2003), who report it as \$6,830,211,000,000 in the spreadsheet referenced in the previous footnote. Using this figure, we directly compute the total income of workers in each profession earning more than \$280,000 and \$1,200,000, and then divide by the counts to arrive at the average.

[19] The break point adds a second degree of freedom in extending the pdf from \$1,200,000 to \$280,000. Using two degrees of freedom, we perfectly match the number of workers in this interval as well as their average income. Matching both statistics is critical for our analysis. The average income in this interval determines much of the aggregate spillover of each profession. The number of workers in the interval determines the average externality of workers earning these incomes, which matters for the optimal income tax at these incomes.

$[y_{i,k}^-, y_{i,k}^+)$; $\sum_k s_{i,k} = s_i$, the total number of workers in i. Let $\widehat{f_i^y}$ denote the income distribution heretofore estimated for $y \geq \$280{,}000$. We use the following likelihood estimator to obtain the Pareto-lognormal parameters:

$$\widehat{A}_i, \widehat{\alpha}_i, \widehat{\mu}_i, \widehat{\nu}_i, = \arg\max_{A,\alpha,\mu,\nu} \sum_k s_{i,k} \log\left(F_i^y(y_{i,k}^+) - F_i^y(y_{i,k}^-)\right), \tag{15}$$

where F_i^y is the cdf corresponding to the pdf f_i^y, and the following constraints bind: $f_i(y) = A_i f_{\alpha_i,\mu_i,\nu_i}(y)$ for $y <$ \$280,000, $f_i^y(y) = \widehat{f_i^y}(y)$ for $y \geq$\$280,000, and $A_i f_{\alpha_i,\mu_i,\nu_i}$(\$280,000) $= \widehat{f_i^y}$(\$280,000).[20]

Table 2 reports summary statistics on the resulting distributions of income for each profession. Skilled professions comprise 18% of all workers, and skilled workers earn 40% of all income. The most populated skilled professions are management and teaching, and the least are real estate, law, and medicine. Substantial heterogeneity in income exists among the skilled professions. Median income ranges from \$40,000 in art to \$201,000 in medicine. Incomes vary even more at the 99th percentile. For instance, engineering and finance have similar median incomes, but the 99th percentile income in finance (\$2,075,000) is more than four times greater than that in engineering (\$452,000).

Figure 1 shows the allocation of workers across professions at each income. Although skilled workers account for only 18% of the total population, they comprise the majority of high earners, as documented in Panel (a). Panel (b) details the composition of skilled workers at each income. At low incomes, the most common profession for skilled workers is art, a result resonant with the image of the "starving artist." Teaching, sales, and operations comprise most of the skilled lower middle class, whereas engineering and management are the largest group in the upper middle class. Nearly all wealthy skilled workers are in finance, law, management, and medicine, and the very wealthy work primarily in management and finance, with some also in law and real estate.

These income distributions by profession are determined in equilibrium by sorting as well as underlying income possibilities. In Appendix C.2, we graph, under our baseline assumption of no comparative advantage, the estimated underlying distributions of income at each skill level, from which individuals choose professions.

4.3.2 Preference and skill parameters

In our baseline analysis, we use a value of $\rho = 1$, which imposes a unidimensional skill distribution on the skilled workers and rules out sorting on comparative advantage. In the broad population and in the short term, this assumption is clearly problematic given the strong evidence of sorting into educational tracks based on comparative advantage shown empirically by Kirkebøen et al. (Forthcoming). However, reconciling a significant, long-term comparative advantage at the top end

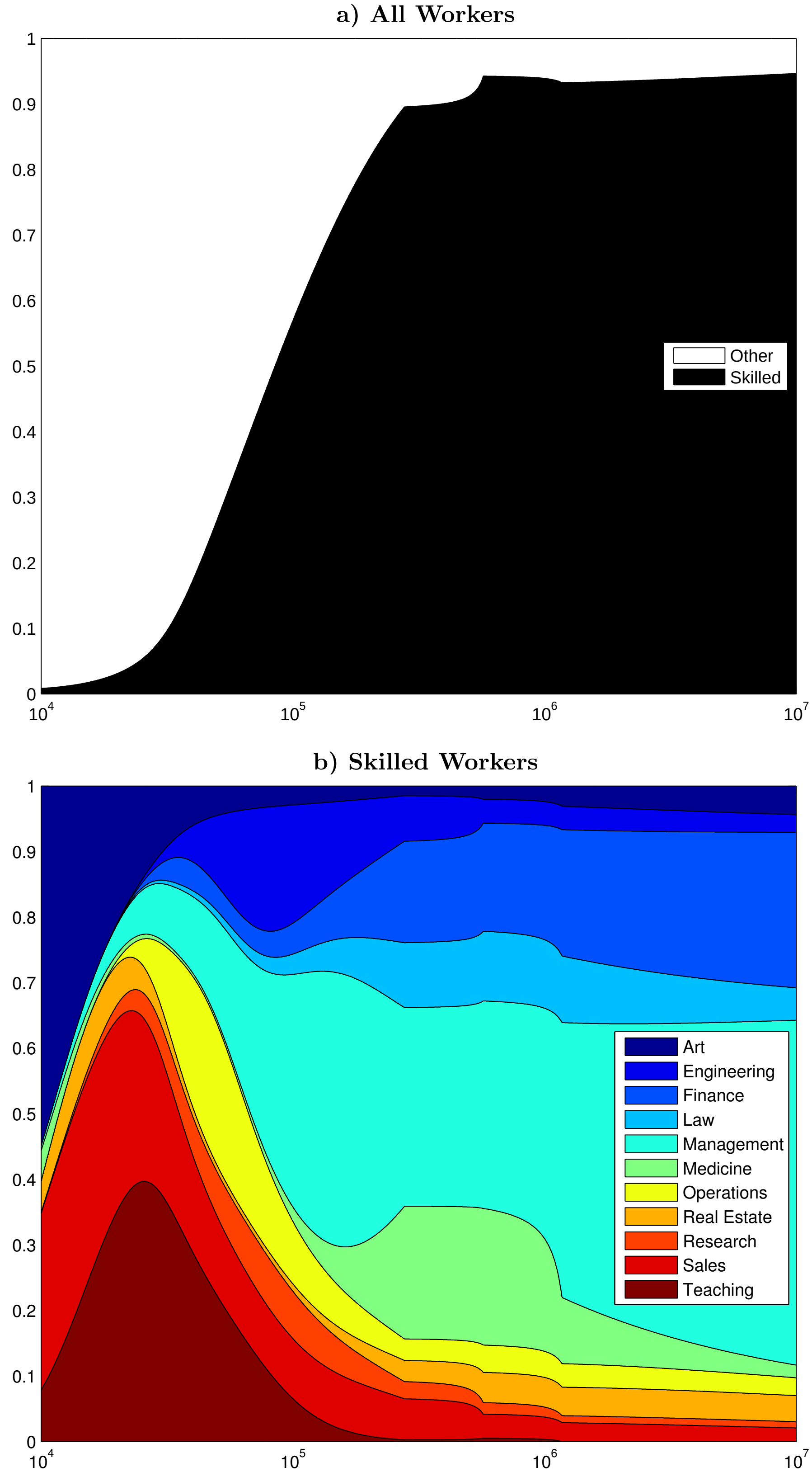
[20] The α_i estimated at this step need not equal the α_i estimated to fit the income distribution over \$1,200,000.

TABLE 2
Summary Statistics of Estimated Professional Income Distributions

	Pop. Share	Inc. Share	Median Inc.	99^{th} Percentile Inc.
Art	1.0%	1.4%	$40,000	$497,000
Engineering	2.0%	3.9%	$73,000	$452,000
Finance	0.9%	4.3%	$85,000	$2,075,000
Law	0.4%	2.1%	$113,000	$1,627,000
Management	3.9%	12.9%	$83,000	$1,273,000
Medicine	0.5%	2.9%	$201,000	$1,346,000
Operations	2.4%	3.6%	$54,000	$368,000
Real Estate	0.3%	1.0%	$50,000	$1,393,000
Research	1.1%	1.8%	$62,000	$399,000
Sales	2.3%	3.3%	$48,000	$414,000
Teaching	3.2%	3.2%	$43,000	$126,000
Other	82.0%	59.7%	$29,000	$118,000

Notes: "Population share" is the fraction of the total workers in each profession, and "Income share" is the fraction of aggregate income earned by workers in each profession. "Median income" and "99^{th} percentile income" are the 50^{th} and 99^{th} percentile incomes within each profession. The results describe the United States in 2005.

FIGURE 1
Distribution of Workers at Each Income Level



Notes: At each income y, the share of workers in profession i is $s_i f_i^y(y) / \sum_j s_j f_j^y(y)$, where s_i is the share of all workers in i. The results describe the United States in 2005.

of the income distribution with the massive reallocations of talent (from the middle-class professions to law and finance) over time observed by Goldin et al. (2013) and Philippon and Reshef (2012) is difficult. The sorting patterns such a comparative advantage would create are counterintuitive. For example, they imply an upward productivity shock in finance will cause mean wages to fall in finance because those who switch in will primarily be workers without large profession-idiosyncratic ability draws. This pattern seems inconsistent with the influx of extremely high-skilled workers that accompanied the growth of the financial profession as documented quantitatively by Philippon and Reshef and discussed ethnographically by Patterson (2010).

We therefore focus on the admittedly very special case of general ability, because of the more plausible sorting patterns it induces and because comparative advantage may be less extreme in the long term when educational curricula and long-term life goals of students may be adjusted. In the sensitivity analysis, we use a smaller value of $\rho = 0.75$ to explore the effects of comparative advantage on optimal tax rates.[21]

We then calibrate σ and β to match two moments of the distribution of income given the parameters and distributions estimated by Lemma 2, which in turn use σ and β. We iterate this step until we reach convergence on a fixed point. The first moment is the elasticity of total economy income with respect to 1 minus the tax rate. A vast literature (Saez et al., 2012) estimates this moment using tax reforms. Chetty (2012) reviews this literature and favors a long-run value for this elasticity of 0.33. We adopt this value as our baseline, and experiment with 0.1 and 0.5 in the sensitivity analysis. To match the moment, we consider the response of aggregate income to a change in taxes, holding profession externalities constant but allowing workers' hours and professional choices to vary. Precisely, we compute $\partial \log Y / \partial \log(1 - T')$, where Y is total income and T' is a constant marginal tax rate. We numerically compute this derivative around the average empirical marginal tax rate T'_{2005}, holding each $E_i(Y_1, ..., Y_n)$ constant.

The second moment is the sensitivity of profession choice with respect to relative income, which helps tie down β, but has not been previously estimated in the literature to our knowledge. To calibrate this sensitivity, we exploit the secular growth in finance wages and employment between 1980 and 2005. As estimated by Philippon and Reshef (2012), the share of all workers in finance grew from 0.35% to 0.87% over this time, while the wages in finance relative to the rest of the (non-farm) economy grew from 1.09 to 3.62.[22] To match these trends, we study

[21] When we vary ρ to 0.75, we continue to use the values of σ and β estimated with $\rho = 1$, because for lower values of ρ, we cannot find β to match the moment below. When comparative advantage is high, a productivity shock in finance actually *lowers* the relative wage in finance, because the shock attracts workers with low productivity to switch into finance. Thus, low ρ rules out a secular increase in finance employment and relative wages as a response to a productivity shock. Rather than try to model these increases differently, we simply hold σ and β constant as we vary ρ. To be clear, we still re-estimate the underlying productivity distributions for the new value of ρ.

[22] These figures use the "other finance" subprofession defined by Philippon and Reshef (2012), because it is constructed similarly to our "finance" profession. The number of workers estimated by Philippon and Reshef (2012) in "other finance" in 2005 equals the number of workers we estimate in "finance" in 2005.

the marginal effect of a productivity shock to finance, which we model as a shock that multiplies each worker's productivity in finance by some constant $\overline{a}$.[23] The relative wage of finance equals $\widetilde{w}_i = \int_{\Theta_i} w_i(\theta)/s_i f(\theta)d\theta / \sum_{j\neq i} \int_{\Theta_j} w_j(\theta)/(1-s_i)f(\theta)d\theta$, where i denotes the index of finance. The moment we match is the fraction

$$\frac{\partial s_i/\partial \overline{a}}{\partial \log \widetilde{w}_i/\partial \overline{a}},$$

where each partial derivative is evaluated at $\overline{a} = 1$. To obtain an empirical value for this moment, we must make an assumption about how frequently new workers replace incumbent ones. Our model is one of long-term professional choice, so s_i is best interpreted as the *flow* of workers into finance; the Philippon and Reshef (2012) data concern the stock. In our baseline analysis, we assume 5% of the worker stock is replaced each period. Appendix C.3 shows this assumption leads to a value of the above derivative of 0.01. In the sensitivity analysis, we use replacement rates of 3% and 10%, which lead to respectively higher and lower values of β.

4.3.3 Externalities

The identification of the externality parameters ϵ_i relies on three inputs: the returns to scale γ of each externality, the marginal social product $\partial Y/\partial Y_i$ of output in each profession, and the $\delta_{i,j}$ linkages.

The literatures we draw on provide no clear guidance on the returns to scale from the various externalities we consider. In our baseline analysis, we therefore choose $\gamma = 1$. The alternate values we use for sensitivity analysis are 0.5, 0.9, and 1.1, which allow us to explore the effects of diminishing and increasing returns to scale of the externalities. Similarly, we set $\delta_{i,j} = 1$ (uniform externalities) for all i and j as a baseline and consider alternate specifications in the sensitivity analysis.

To calculate the marginal social output from each profession, we draw on the literatures that estimate economy-wide externalities from various professions. Although we have done our best to faithfully represent the current literature, we emphasize, and return in our conclusion to, the fact that these estimates are *highly* uncertain extrapolations from heterogeneous and not easily comparable studies primarily aimed at different estimands than those we draw from them. The resulting estimates are listed in Table 3.

To arrive at the marginal social product $\partial Y/\partial Y_i$, we divide each profession's total social product by its total private product.[24] The private product is given by the "income share" column of Table

[23] Specifically, for i corresponding to finance, $a_i(\theta)$ is replaced by $\overline{a}a_i(\theta)$ for all θ.

[24] This empirical ratio gives the average externality rather than the marginal one. However, some of the aggregate spillovers we take from the literature seem better interpreted as marginal effects (Murphy et al., 1991; Chetty et al., 2014). We believe simply dividing the social product by the private product to estimate the marginal externality is most transparent, rather than making further adjustments with the estimates from the literature. We target the marginal externality in the model because the feedback effects among the outputs of the professions make how to calculate a "total" spillover unclear. The marginal externality is local to the equilibrium we are estimating in the

2, and the social product is the sum of this private product and the externality given by Table 3. For example, the marginal social product of teaching equals $(3.2\% + 7.3\%)/3.2\% = 3.28$.

Given the high degree of uncertainty and inevitable subjectivity in these estimates, we devote the remainder of this section to briefly highlighting how we calculate the aggregate externalities in Table 3, with required calculations left to Appendix C.4. Our prior is that Coasian bargaining should eliminate externalities, so when these literatures do not offer a clear finding, we set the aggregate externality to 0. In the cases in which these literatures offer conflicting results, we adopt one value as a baseline and use an alternate value for sensitivity analysis.

Arts Although some evidence, and a number of good theoretical arguments, suggest the arts generate some positive externalities, we are unable to find a plausible basis for estimating the magnitude of these externalities, and consequently assume 0 to be conservative.

Engineering The only study we found of externalities from engineering is a cross-country ordinary least-squares regression by Murphy et al. (1991). They investigate the impact of the allocation of talent on GDP growth rates rather than on GDP levels. To be conservative and fit within our static framework, we interpret these impacts as one-time effects on the level of output rather than impacts on growth rates. We multiply their estimate of the GDP impact of an increase in the fraction of students studying engineering by the number of students studying engineering according to the OECD to obtain an externality of 0.6% of total income.

Finance French (2008) estimates the cost of resources expended to "beat the market" by subtracting passive management fees from active management fees. Bai et al. (Forthcoming) show the informativeness of stock and bond prices (measured in their ability to predict earnings) has stayed constant since 1960, despite a vast growth of the finance profession documented by Philippon (2010). We therefore interpret the entirety of French (2008)'s estimates, which amount to 1.4% of total income in 2005, as negative externalities from finance.

Law Murphy et al. (1991) estimate externalities from law in the same manner they calculate externalities from engineering, and we apply the same methodology to yield a −0.2% externality as a percent of total income. Kaplow and Shavell (1992) present several models of why the provision of legal advice may exceed the social optimum.

Management Two strands in the literature offer competing views on the externalities of management. According to the first strand (Bertrand and Mullainathan, 2001; Malmendier and Tate, 2009), chief executive officer (CEO) compensation shifts resources from shareholders to managers in

data, whereas a "total" externality would depend on the model very far from the point at which it was estimated.

TABLE 3
Aggregate Externalities by Profession: Baseline Estimates

	Externality as share of economy income	Source	Method
Arts	0	—	—
Engineering	0.6%	Murphy et al. (1991)	Cross-country regression of GDP on engineers per capita
Finance	−1.4%	French (2008)	Aggregate fees for active vs. passive investing
Law	−0.2%	Murphy et al. (1991)	Cross-country regression of GDP on lawyers per capita
Management	0	Gabaix and Landier (2008)	Calibrated model indicating CEO pay captures managerial skill and firm characteristics
Medicine	0	—	—
Operations	0	Bloom et al. (2013)	Randomized experiment measuring effect of consultants on plant productivity
Real Estate	0	—	—
Research	16.7%	Murphy and Topel (2006)	Willingness-to-pay for longevity gains from medical research
Sales	0	—	—
Teaching	7.3%	Card (1999)	Returns to education in excess of teacher salaries
Other	0	—	—

Notes: This table reports the total externalities for each profession as a share of the total income in the economy, which is \$6.3 trillion according to our estimates from Table 2. We calculate each externality using results from the listed papers; see the text for a description of how we map each paper's results to an aggregate externality figure. We use these externalities throughout the paper, except in the case of management and research. In sensitivity analysis, we explore the implications of a negative externality for management (taken from Piketty et al., 2014) and a smaller positive externality for research (taken from Jaffe, 1989).

ways that do not actually reflect the CEO's marginal product. Piketty et al. (2014) argue that 60% of the CEO earnings elasticity with respect to taxes represents this rent-seeking behavior, implying the negative externalities from management are 7.7% of total income. The other half of the literature argues market forces can explain CEO compensation (Gabaix and Landier, 2008) and suggests that therefore externalities are 0. Most managers in our sample work at lower levels of firms where the problems of measuring marginal product highlighted by the critics of CEO compensation are less likely to apply, so we take the figure of 0 as our baseline and use the -7.7% figure in sensitivity analysis.

Medicine We could find no literature estimating the externalities of (non-research) medicine and so set the externality to 0 to be conservative.

Operations This profession is comprised of consultants and IT professionals. Bloom et al. (2013) conducted a field experiment to determine the causal impact of management consulting on profits. They interpreted their results as consistent with the view that consultants earn approximately their marginal product, and thus we assume no externality for consulting.

Real Estate We could find no literature estimating the externalities of brokers, property managers, and appraisers and so set the externality to 0 to be conservative.

Research Our baseline estimate for the externalities from research comes from the value of medical research, measured in terms of people's willingness to pay for the additional longevity this research makes possible. Murphy and Topel (2006) estimate the annual gains of medical research equaled 20% of GDP from 1980-2000. Traditional GDP accounting does not capture this externality, in contrast to our model, so we divide it by GDP augmented with this externality to obtain $\frac{.2}{1+.2} = 16.7\%$. Although this externality may be the largest externality from academia and science, this estimate is still conservative in assuming no gains accrue from other research fields.

An alternative measure of research externalities comes from the literature that calculates the social returns to R&D. Jones and Williams (1998) suggest the socially optimal amount of R&D activity is four times the observed amount, which we loosely translate into a three-times externality or 5.4% of GDP. A narrower benchmark for this externality focuses only on the externalities of universities to profits made by geographically proximate firms as studied in Jaffe (1989). We use his estimates to calculate a much smaller 2.7% externality, which we use as a lower-bound estimate in our sensitivity analysis.

Sales Although an extensive theoretical literature argues the welfare effects of advertising can be positive or negative (Bagwell, 2007), we are not aware of any work attempting a comprehensive estimate of externalities, and therefore, as with medicine, we use an externality of 0.

Teaching We calculate the social product of teaching as the impact of an additional year of schooling on aggregate earnings of all workers in the economy. The spillover from teaching is then this social product less the annual earnings of all teachers. As our estimate of the effect of a year of schooling on earnings, we use a 10.5% gain, which equals the midpoint of the numbers collected in Card (1999)'s review and also the estimate from Angrist and Krueger (1991). Because teachers earn 3.2% of economy income, we use a spillover from teaching of 7.3% of economy income.

We also compute the aggregate effect of teaching on earnings using Chetty et al. (2014)'s measure of teacher quality and its long-run impact on eventual student earnings. We use the ratio of total teacher pay to its standard deviation in our data multiplied by the social product Chetty et al. (2014) estimate for a standard deviation in teacher quality to obtain an aggregate effect equal to 9.6% of economy income. This figure leads to a spillover of 6.4% of economy income. Given the similarity between the two spillover estimates and the fact that the estimate based on returns to schooling is more easily interpretable in the aggregate, we use the Card (1999) number as our estimate.

5 Results

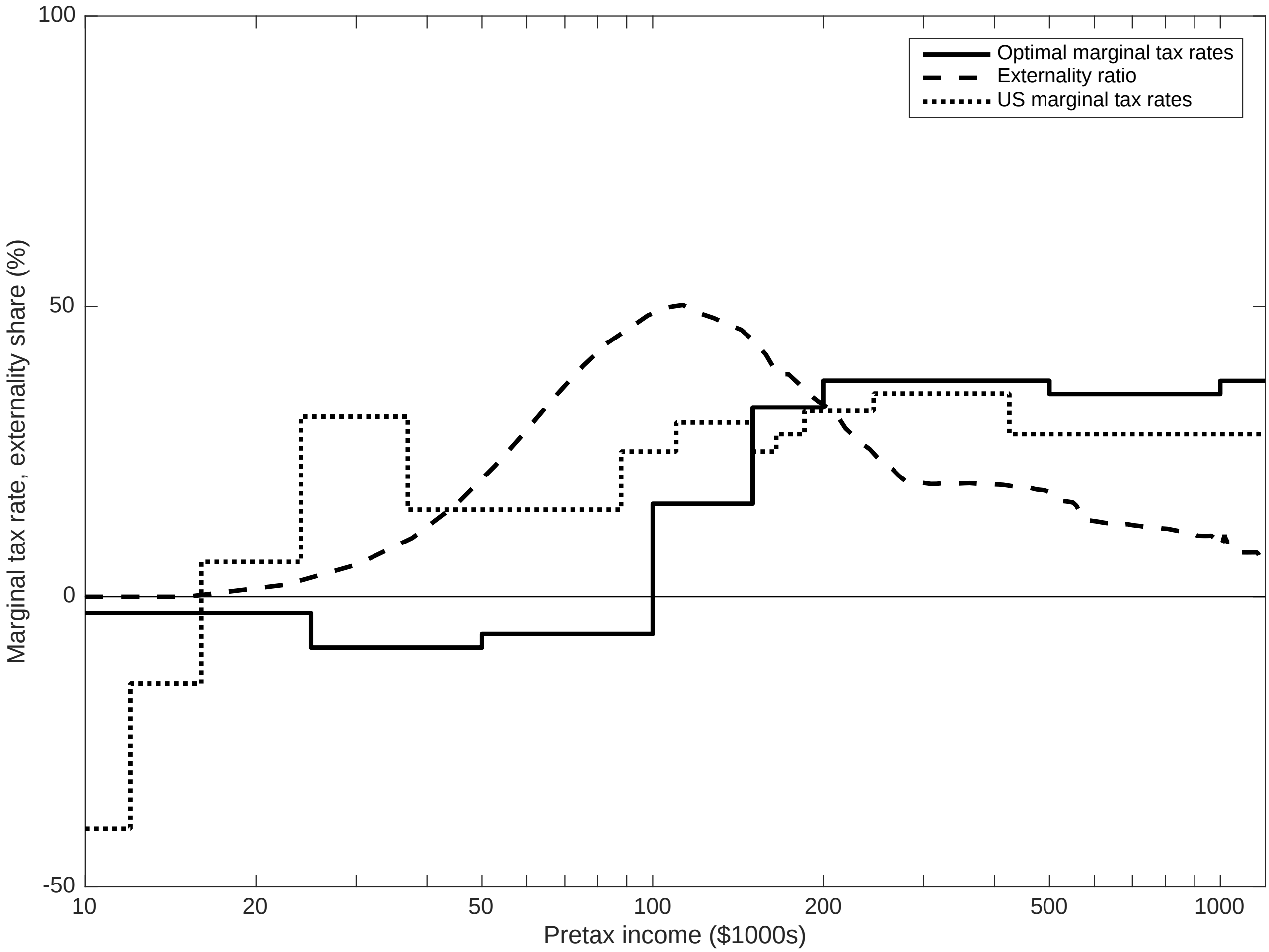
Before investigating optimal taxes, considering the quantitative value of a leading force determining them is instructive: the externality ratio $e(y)$ in the equilibrium at the optimal tax schedule. We defined this externality ratio in Section 2 as the average marginal externality of income earned by those with income equal to y. Proposition 1 showed that in the special case when workers cannot switch professions, the optimal tax schedule satisfies $T'(y) = -e(y)$, thus setting marginal tax rates equal to the average negative externality ratio at each income level. We plot $-e(y)$ as the hashed line along with the optimal tax rates we discuss below in Figure 2. Absent the allocative labor supply margin, these two items in Figure 2 would be mirror images of each other. Interestingly, the results differ markedly from this benchmark.

5.1 Optimal taxes

Given the underlying skill distributions, preference parameters, and externalities we estimate, we numerically calculate the marginal tax schedule that maximizes social welfare. This procedure uses significant computational resources, so we restrict attention to schedules with eight brackets, with cutoffs at $25k, $50k, $100k, $150k, $200k, $500k, and $1m. This restriction clearly violates Assumption 1 but allows for direct optimization at reasonable cost.

Figure 2 presents the results. Optimal taxes (the solid line) begin with negative rates of about 6% on income up to $100,000 and then feature progressively increasing marginal rates after that. The top rate on income above $1m is 37%, and similar marginal rates hold for income above $150k

FIGURE 2
Optimal Marginal Taxes for the United States in 2005



Notes: The externality ratio is the quantity $e(y)$, defined in Section 2 to be the average marginal externality of income earned by workers with income equal to y. The marginal tax rate displays the optimal tax rates over the eight brackets specified, given the data and baseline assumptions explained in Sections 4.1 and 4.3. US Marginal Tax Rates are taken from Figure 4 of CBO (2005) and denote the effective marginal tax rate for a married couple with two children in 2005, accounting for the Earned Income Tax Credit, the Alternative Minimum Tax, the phaseout of itemized deductions, the child tax credit, and personal exemptions.

in other brackets.[25]

To understand this tax schedule, consider the net tax liabilities of workers at different income levels relative to that of a worker with zero income. These *net* tax liabilities are all our optimal schedule identifies; the revenue requirement as explained by Lemma 1 solely determines the overall

[25] Figure 2 presents a local maximum for marginal tax rates. We did find a second local maximum in which welfare was slightly higher ($22 per person). This alternative schedule is nearly identical to the one in Figure 2 except the marginal rates in the $150k-$200k bracket are much higher, over 95%. This optimum is likely an artifact of the way the brackets are constructed. It is present on only the smallest bracket (in log terms), and it disappears when we change the $150k-$200k bracket to $150k-$250k. For these reasons, we do not focus on this optimum. We report it in Appendix C.5.

level of the tax schedule. Due to the negative rates that last until \$100k, net tax liabilities are negative up to \$138k, so that a worker earning \$138k pays the same tax as a worker earning no income. Beyond this point, the marginal rate varies, but on average is about 35%. The smallest tax liability is for a worker earning \$100k, who receives a net income subsidy of \$6,100.

The top tax rates are close to the marginal tax rates the federal government in the United States has applied to top incomes since 1986; the 2005 US federal schedule of marginal rates is pictured in the small dashed lines. Thus, regarding tax rates on the rich, the model's recommendation matches the positive reality. Our model generates these optimal rates without any redistribution motive. The tax rates serve only to increase positive externalities and decrease negative ones.

The model's recommendations differ from policy at lower incomes. Empirically, rates below \$100,000 are much higher than the model's negative optimal rates, both because statutory rates are higher (as depicted in Figure 2) and because benefits to the poor phase out as income increases over this range (CBO, 2005). The model prescribes negative rates on income all the way up to \$100k, which is a much higher threshold than those used by income subsidies in practice, such as the Earned Income Tax Credit. A possible explanation for this failure of the model, but consistent with its spirit, is the negative externalities created by poverty through crime and empathy emphasized by Harberger (1978) and Frankfurt (1987).[26]

To see most sharply the impact of the allocative margin, note that at high incomes, the externality ratio is positive but so are marginal tax rates. Researchers produce the positive externalities at these high incomes—although they constitute a small number of top earners, their externalities are extremely large relative to the negative externalities of law and finance. Yet despite the net positive externalities at high incomes, tax rates are still positive and large there because externalities are even higher at lower incomes. The top tax rates are positive to induce higher earners to switch to lower-paying professions that produce greater externalities.

5.2 Welfare gains and the allocation of talent

We now calculate the gains associated with taxation in our model with respect to two reference points: the empirical US economy in 2005, and a *laissez-faire* economy without any income tax. The latter comparison measures the general efficacy of the income tax for improving welfare, whereas the former provides the marginal improvement that could be obtained from changing the tax already in place. *Laissez-faire* serves as an informative benchmark for the additional reason that it is the optimal marginal tax schedule in our model when externalities are absent.

Table 4 reports the results. Relative to *laissez-faire,* the optimal tax raises average utility by \$815, or 1.2%. The tax achieves a smaller gain of 0.8% relative to the empirical economy, which is not surprising given the tax used to model the empirical economy (a flat 30% tax) is close to the optimal

[26]Incorporating a criminal profession choice available to unskilled individuals but requiring sacrifice of their state subsidies, and thus creating a motive for the state to subsidize the poor, is an interesting direction for future research.

TABLE 4
Per Capita Welfare Gains Relative to *Laissez-Faire*

	2005 US Data	Optimal Nonlinear Income Tax
Levels	\$312	\$815
Percent	0.4%	1.2%

Notes: Gains are computed relative to a *laissez-faire* simulation that has no income tax. We evaluate the social-welfare function in the observed 2005 US economy as well as in the economy at the optimal tax schedule shown in Figure 2. The first row is the difference in per-capita utility relative to *laissez-faire*, and the second row is the percent improvement relative to this benchmark.

tax we calculate. These gains are significant, but still small relative to the externalities calculated in Table 3. These large externalities—for instance, research at 16.7% of the economy—suggest a reallocation of talent to more productive professions could increase welfare by much more than the 1.2% achieved by the optimal income tax. Our findings that welfare gains are quite small are robust to all scenarios we consider, except those with targeted subsidies to research; they never exceed 2%, and in some scenarios are as small as 0.4%. The largest welfare gains come when the allocative margin is strongest (when individuals switch elastically across professions or the intensive-margin elasticity is small) and the smallest come when we assume a smaller externality of research.

A possible reason for the inefficacy of the income tax is that it possesses a limited ability to induce switching between professions, given that workers' tax liabilities are independent of their professions. To investigate this idea, we calculate the allocation of talent under *laissez-faire* and under the optimal tax. Table 5 reports the share of skilled workers in each profession in the data and in each of these two simulations. Relative to *laissez-faire*, the optimal tax decreases the share of workers in negative externality professions (finance and law) and increases the share in positive externality professions (engineering, research, and teaching). However, none of these changes are very large, and the broad allocation of talent stays the same. Relative to the status quo, the optimal tax primarily shifts individuals out of low-earning professions (e.g., art, sales and teaching) and into middle income professions (e.g., engineering and management). These changes result from the marginal rates in the status quo being much higher on the working and middle class than in the optimum. This reallocation does some good, mostly by raising incomes rather than externalities per unit income, but allocates workers out of teaching. Another reason taxes are ineffectual is that the intensive margin pushes them in the opposite direction, implying that high taxes that have allocative benefits also come at intensive margin costs.

These results suggest that historical tax reductions are unlikely to have played a large role in

TABLE 5
Share of Skilled Workers in Each Profession

	2005 US Data	Optimal Nonlinear Income Tax	*Laissez-Faire*	Pre-Reagan Income Tax
Art	5.1%	4.3%	4.3%	5.4%
Engineering	11.8%	13.3%	12.4%	11.6%
Finance	5.1%	5.3%	5.7%	4.7%
Law	2.7%	3.1%	3.5%	2.3%
Management	22.5%	24.4%	24.0%	21.4%
Medicine	2.9%	3.2%	4.7%	2.1%
Operations	13.3%	13.2%	12.7%	13.9%
Real Estate	1.7%	1.6%	1.6%	1.8%
Research	6.2%	6.2%	5.9%	6.3%
Sales	11.9%	10.8%	10.6%	12.6%
Teaching	16.7%	14.7%	14.6%	17.8%

Notes: Columns after the first report results from simulations under the optimal tax in Figure 2, under the *laissez-faire* economy with no income tax and under 1980 US tax policy.

the shifts in talent allocation. To confirm this hypothesis, we use the Tax Foundation's US Federal Individual Income Tax Rates history to simulate talent allocation and welfare under the 1980 ("Pre-Reagan") income-tax schedule. This schedule involves much higher rates and a more progressive structure; it provides a more extreme departure from *laissez-faire* than the 2005 schedule. Welfare is substantially lower under the pre-Reagan rates relative to the status quo and by 0.8% relative to *laissez-faire*. The allocation of talent under this schedule is shown in the final column of Table 5. As expected, the allocation of talent is only slightly different than *laissez-faire* under the pre-Reagan schedule.

TABLE 6
Optimal Tax Rates for Different Assumptions

	\$0-\$25k	\$25k-\$50k	\$50k-\$100k	\$100k-\$150k	\$150k-\$200k	\$200k-\$500k	\$500k-\$1m	\$1m+
Baseline	**−2.8%**	**−8.8%**	**−6.4%**	**16.0%**	**32.6%**	**37.2%**	**34.9%**	**37.2%**
Elasticities								
Low σ	−19.5%	11.5%	46.6%	71.7%	74.8%	61.7%	59.8%	65.0%
High σ	−2.3%	−9.4%	−16.5%	0.4%	23.3%	32.3%	29.8%	32.4%
Low β	−2.6%	−11.2%	−17.2%	2.3%	22.6%	30.8%	28.7%	31.6%
High β	−2.6%	−6.2%	−0.1%	24.8%	37.5%	41.6%	39.8%	41.8%
Externalities								
Smaller rsch. ext.	−1.4%	−0.7%	7.1%	20.6%	26.4%	24.8%	20.8%	24.4%
Neg. mgmt. ext.	−2.3%	−5.7%	4.5%	31.2%	43.9%	46.2%	50.0%	60.3%
Fin. on fin.	−2.8%	−10.4%	−8.7%	13.3%	28.9%	32.2%	28.0%	27.5%
Eng. on eng.	−2.2%	−6.3%	−0.3%	18.4%	31.9%	36.3%	34.1%	36.3%
Rsch. on eng.	−1.2%	−1.0%	6.4%	19.5%	26.0%	24.2%	20.7%	23.3%
$\gamma = 0.5$	−2.7%	−8.3%	−5.7%	15.8%	31.8%	36.1%	33.4%	35.7%
$\gamma = 0.9$	−2.8%	−8.4%	−5.6%	16.9%	33.1%	37.4%	35.0%	37.3%
$\gamma = 1.1$	−2.8%	−9.1%	−7.1%	15.4%	32.2%	37.1%	34.8%	37.1%
Neg. own ext.	−1.6%	−3.7%	1.6%	19.9%	30.0%	29.8%	27.7%	30.8%
Comparative Advantage								
$\rho = 0.75$	−3.2%	−12.5%	−19.5%	9.9%	32.6%	29.5%	16.0%	11.7%
Tax Instruments								
Rsch. subsidy (neg. own ext.)	−3.1%	−3.2%	4.3%	18.1%	30.9%	43.4%	28.8%	42.0%
Rsch. subsidy ($\gamma = 0.5$)	0.0%	0.0%	−0.1%	−1.1%	−3.1%	−0.1%	80.0%	80.0%

5.3 Sensitivity to alternate assumptions

Table 6 reports the optimal tax rates under various alternate assumptions, which we now discuss.

5.3.1 Elasticities

We begin by varying our input for the elasticity of taxable income with respect to 1 minus the tax rate. We experiment with values of 0.1 and 0.5 (our baseline was 0.33). These inputs lead to estimated σ values of 0.01 and 0.4; our baseline estimate was 0.24. The estimated β changes only slightly to 1.47 and 1.42 relative to the baseline of 1.5. These changes are small because we are holding constant the separate moment that mostly determines β. Thus, this experiment alters the *relative* importance of the intensive-margin elasticity rather than just the overall elasticity that, as we noted in Section 2.2, plays no role in determining optimal tax rates under our theory.

Optimal tax rates are much higher with a lower σ, as shown by the rates above 70%, especially on the upper-middle-class range that makes the largest difference between being in the wealthy and middle-class professions. These high rates are consistent with the logic of our calibrated example; profession switching becomes more important relative to hours in determining optimal taxes when σ is small. The effect is more dramatic here, however, because the mixed sorting created by the richer substitution patterns in our analysis here means that the value of the allocative margin is smaller. Unless the intensive margin is very weak, it has a strong influence on optimal taxes, implying that weakening it significantly raises optimal taxes by leaving the weak allocative margin to determine tax rates uncontested. At the higher value of $\sigma = 0.4$, which is at the high end of estimates obtained from microeconomic studies (Chetty, 2012), optimal rates are still progressive. The top rates are slightly smaller, and the negative rates on low earners are more extreme.

We next vary the profession-switching sensitivity β. As discussed earlier, we vary the assumed replacement rates of workers into finance in our calibration to 10% and 3% from the baseline of 5%. These alternate assumptions lead to values of β of 1.0 and 2.0 versus the baseline value of 1.5. Consistent with our argument that only the relative size of the intensive and allocative margin elasticities matters, the lower value of β gives results similar to the higher value of σ. The higher value of β moves toward results for the low value of σ, though not as dramatically, because it involves a much smaller change in the ratio of the two forces (β increases by one third while σ fell by an order of magnitude). These results provide another quantitative confirmation that discrete profession switches are central to the progressive structure of taxes we find.

5.3.2 Externalities

We vary the externalities in numerous ways, given our substantial uncertainty over both their magnitude and functional form.

We begin with two specifications that change the magnitude of the externalities. The research externality is the largest externality. To investigate the degree to which this externality drives the results, we use a much smaller aggregate externality of 2.7% instead of 16.7% of economy income. As discussed earlier, this smaller number is calibrated from the literature on R&D externalities. This smaller research externality does indeed produce smaller top tax rates and higher rates for low-income workers, as can be seen in Table 6. The negative rates end earlier (at $50k), and rates on higher earners are lower (between 20% and 30%). But the basic structure of the tax system stays intact. Changing the management externality from 0 to an aggregate negative externality of 7.7% of economy income makes a much larger difference by making nearly all high-earning professions have negative externalities. Tax rates on income between $150k and $1m jump from about 35% to about 45%, and the tax rate on income over $1m rises to 60%. This result is analogous to our finding in Section 3.2 that raising the negative externality of the high-earning profession is more important than raising the positive externality of the low-earning profession.

The range of externality magnitudes explored here—which reflects the opinions of various economists—generates as much variation in optimal tax rates as differences in the elasticity of taxable income. As Table 6 shows, varying this elasticity generates top rates between 32% and 65%, whereas varying the externalities from research and management generate top rates between 24% and 60%. Furthermore, the aggregate elasticity of taxable income (ETI), which is more relevant than the individual components of labor supply to the standard theory (Feldstein, 1999), is completely irrelevant in our theory, as shown in Section 2.2. Estimating externalities and the decomposition of ETI into its components have attracted far less attention than estimating ETI (Saez et al., 2012), yet the first two determine tax rates in our framework rather than the third.

We now alter the functional form of the externalities. First we consider what happens when the externality from finance falls entirely on itself by setting $\delta_{i,j} = 0$ for $i \neq j$ when j indexes finance. The resulting top tax rates, especially at the very top, are smaller than the baseline optimal rates. For instance, the rate above $1m falls from 37% to 27.5%. This decline in rates is consistent with the theoretical results of RS, who show the social planner has little incentive to tax rent-seeking when the rent-seekers compete against each other, which is the case when finance externalities fall entirely on finance. The decline here is much less dramatic than in their quantitative calibration where finance is the only profession with externalities and only two professions exist. In our estimation, the largest externalities are positive and from the middle-class professions.

Next we alter the functional form to allow all engineering externalities to fall on engineering. This specification is motivated by industrial research clusters like Silicon Valley in which engineering firms create new ideas that enhance the productivity of other engineering firms (Saxenian, 2006). This specification leaves optimal tax rates essentially unchanged. We also consider a specification in which research externalities fall entirely on engineering. To be consistent with how we calibrate research externalities, we use the smaller externalities calibrated from the R&D literature for this exercise. Relative to the optimal rates under that calibration, the rates when research externalities fall entirely on engineering are largely unchanged. In principle, these different linkages could lead to larger rates by causing feedback effects that increase the net benefit of profession switching. This effect appears to be balanced by the lower aggregate externalities implied by Jaffe (1989)'s much lower externality estimates, suggesting that even his estimates, correctly interpreted, would lead to quite similar results.

Our baseline analysis assumed externalities were linear in output by setting the returns to scale parameter γ to 1. We explore the possibility of economies or diseconomies of scale in externalities by setting $\gamma = 0.5$ and 0.9. We also set $\gamma = 1.1$ to investigate the possibility of slightly increasing returns. None of these values materially change the optimal rates, although the low value of $\gamma = 0.5$ does slightly reduce top tax rates. Our tax schedule is sufficiently similar to the status quo that a linear approximation to externalities makes little difference to the results.

Finally, we consider congestion effects wherein the arrival of new workers lowers the productivity of existing workers in a given profession. We implement these congestion effects by assuming each dollar of private product in teaching or research raises the aggregate output of the profession by only 50 cents.[27] Optimal rates do diminish, but the effect is slight, with top rates falling from 37% to 31%. In contrast to the work of RS, we find the multi-profession nature of our economy likely significantly mitigates congestion effects. Negative externalities within a profession are only a small part of the overall impact of individuals migrating into a profession, compared to the impact of that profession on the broader economy.

5.3.3 Comparative advantage

We next consider the impact of allowing comparative advantage, which changes the patterns of substitution across professions. Absent comparative advantage, taxes induce shifts of the very skilled across fields. With comparative advantage, most substitution will occur among lower-ability individuals because higher-ability individuals will tend to have much lower ability in another field.

To explore this effect, we change ρ from 1 to 0.75. We draw from Kirkebøen et al. (Forthcoming) a sample statistic, which we call comparative advantage, to give a sense of the sorting caused by this lower value of ρ. For each skilled worker, define $i_1(\theta) = \arg\max_i F_i^a(a_i(\theta))$ to be the profession in which she is (relatively) most productive and $i_2(\theta) = \arg\max_{i \neq i_1(\theta)} F_i^a(a_i(\theta))$ to be the profession in which her (relative) productivity is second highest. The formula for comparative advantage is

[27] We choose $\delta_{j,j}$ for j corresponding to research and teaching so that the relevant diagonal entries in the quasi-Jacobian matrix J defined in Section 4.1 equal 0.5.

given by

$$\sum_{\substack{1\leq i,j\leq n\\ i\neq j}} Pr\left[i_1(\theta)=i, i_2(\theta)=j\right]\times$$

$$\left(\mathbb{E}\left[\log\left(\frac{y_i^*(\theta)}{y_j^*(\theta)}\right)\middle| i_1(\theta)=i, i_2(\theta)=j\right]-\mathbb{E}\left[\log\left(\frac{y_i^*(\theta)}{y_j^*(\theta)}\right)\middle| i_1(\theta)=j, i_2(\theta)=i\right]\right).$$

This formula gives the average relative income premium of skilled workers in their most skilled profession. At $\rho = 0.75$, comparative advantage is equal to 0.4, representing an average premium of 40 log points of income, close to the figures observed empirically, though in a very different setting, by Kirkebøen et al..[28] When $\rho = 1$, comparative advantage equals 0.

The tax schedule with comparative advantage features declining marginal rates for top incomes, with the rates at \$150k-\$200k similar to before but top tax rates much lower. The new rate on income above \$1m is 11%, and the rate between \$500k and \$1m is 14%. The negative rates for low earners actually increase to a maximum of 21.4%. Comparative advantage makes profession switching unattractive to those earning very high incomes because they are likely to have high idiosyncratic incomes in their present profession. Thus, comparative advantage brings optimal rates for the wealthy closer to the (negative) intensive-margin optimum. Rates remain largely unchanged at middle incomes because individuals with low idiosyncratic ability may still substitute across professions.

The fact that comparative advantage changes the structure of taxes more than any other feature we analyze demonstrates the importance of profession-substitution patterns for optimal taxes and should provoke more research on this topic, which is almost completely unstudied in previous literature. What likely matters is the correlation between productivity in high-earning professions and that in high-externality professions. For instance, productivity may in general be weakly correlated, but if top earners in finance and law would be highly productive researchers, as we anecdotally believe is the case based on accounts as in Patterson (2010), the optimal tax would likely still be progressive.

5.3.4 Tax instruments

We argued the small gains from taxation result from an untargeted income tax struggling to precisely reallocate individuals. To explore targeted policies, we introduce a linear income tax (or subsidy) to supplement the non-linear income tax that the government can levy directly on re-

[28] They estimate comparative advantage using the field-of-study choice of students, as opposed to the ability levels, which they cannot observe. They also find an average value of 0.4, but their number is not closely analogous to ours because of the different definition, because their sample is limited to students on the margin between professions, and because they focus on fields of study rather than professions. We experimented with defining comparative advantage using profession choices but were not able to match their estimate even for $\rho = 0$. Sorting on non-pecuniary utility ψ significantly blunts comparative advantage given the β we have estimated.

search (perhaps through the National Science Foundation), which is the profession we estimate produces the strongest externalities.

Optimizing these instruments in our baseline model does not yield an optimum: it becomes optimal for the planner to provide a very large subsidy to research, which pays for itself through positive externalities. Eventually even equilibrium existence fails as a large research profession puts the economy on an explosive path.[29] Thus, although our linear functional-form assumptions make little difference for policies such as our optimal non-discriminatory tax that are close to the status quo, they make a very large difference when we consider significant departures from the status quo.

The behavior of the economy away from the status quo can be understood by comparing two methods of avoiding this explosive result. The first and perhaps most plausible is to choose each $\delta_{i,i}$ so that the externality of teaching and research on themselves equals -0.1.[30]

In this case, we find an optimal research tax of -392.1%, which would multiply salaries by four times even beyond their subsidized 2005 levels. An important part of this subsidy is to offset the negative effect on salaries of the crowding induced by the negative value of $\delta_{i,i}$. Table 6 reports the optimal income-tax rates accompanying the optimal subsidy, which hardly change from our baseline, even getting a bit higher at the top. Other professions still produce enough externalities that targeting research does not significantly change the picture. Furthermore, because all the negative effects of research fall onto research, the targeted subsidy can offset these burdens.

Welfare is much higher under the research subsidy. Relative to *laissez-faire*, welfare is 37.3% higher. The subsidy allocates 44.7% of skilled workers to research in the equilibrium, almost 10 times the baseline amount. Targeted support for certain key professions can thus greatly raise welfare, and a progressive income tax can still be optimal even in the presence of such targeting.

Another way of avoiding an explosive result is to impose diminishing returns in the production of externalities ($\gamma = 0.5$). Unlike within-research crowding, such diminishing returns do not diminish the private returns to research. They also have an equal effect on externalities in *all* professions, rather than specifically affecting research and teaching. This linearity of private returns makes much larger welfare gains possible, even with a less extreme subsidy. In particular, the optimal research subsidy is now 120%, a large number but much smaller than the previous case, and this subsidy achieves a *much* larger welfare gain of 99.5%. However, this smaller (if still very large) targeted research subsidy is supplemented radical change in the optimal income tax. As reported in Table 6, the optimal top tax rates are a nearly confiscatory 80%, the figure at which we capped rates for convergence; optimal unconstrained rates are likely higher. Rates below the top two are essentially 0.

Intuitively, a large targeted subsidy and confiscatory rates are two methods of inducing greater

[29]This failure provides a nice example of the problem of equilibrium existence, uniqueness, and stability we discussed in Section 2.1.

[30]Relative to the baseline specification, these changes to the form of externalities only slightly alters the optimal tax rates, changing them to -2.5%, -5.1%, -0.3%, 19.8%, 32.0%, 32.8%, 30.4%, and 33.0%.

movement into professions, especially research and teaching, with large positive externalities. When teaching and research have negative externalities on themselves, targeted subsidies are a more effective tool because they offset the reduction in private returns from negative self-externalities as untargeted taxes cannot. However, when the production of externalities merely produces decreasing returns, teaching and research remain competitive professions that near-confiscatory taxes can be effective in inducing individuals to enter. In particular, once a moderate subsidy has been applied to research, progressive untargeted taxes become a much more desirable tool because the subsidy raises the attractiveness of research, ensuring that most substitution out of high-earning professions occurs into research rather than into a field generating fewer positive externalities. Employing a smaller targeted subsidy and larger untargeted taxes is thus optimal because they also induce migration into teaching, which cannot be targeted.

The general message appears quite robust: targeted support for certain key professions can greatly raise welfare, and a progressive income tax can still be optimal even in the presence of such targeting. However, the exact form of optimal policy in the presence of targeting is highly sensitive to very uncertain features of how externalities accrue and to the returns to scale of professions—details that were not crucial with non-discriminatory taxes. Our quantitative prescriptions of optimal taxes with targeting should be interpreted much more cautiously than our baseline results, which prescribe policy close to the status quo and thus are much less sensitive to structural assumptions used for extrapolation.

6 Conclusion

This paper proposes an alternative framework for the optimal taxation of income relative to the standard redistributive theory of Vickrey (1945) and Mirrlees (1971). Income taxation acts as an implicit Pigouvian tax that is used to reallocate talented individuals from professions that cause negative externalities to those that cause positive externalities. Optimal tax rates are highly sensitive to which professions generate what externalities and to the labor-substitution patterns across professions. They do not depend on the overall elasticity of taxable income. In our baseline calibration, our theory does a reasonable job of accounting for the general pattern of income taxation in the United States.

We interpret our results as a starting point for three important lines of inquiry to determine certain parameters more precisely.

The first and most important line of inquiry is work on the externalities created by different professions. Such research is likely to be highly profession-specific, as is the degree of potential for improvement over existing literature. For example, Card (1999) and Chetty et al. (2014) yield consistent and persuasive numbers on the spillovers from teaching, whereas our extrapolations from the cross-country regressions of Murphy et al. (1991) to determine the externalities of engineering and law are speculative at best. Careful empirical analysis could dramatically improve the quality of externality estimates in the latter fields.

For example, simple decomposition of legal activities between adversarial and compliance expenditures could already be useful. Kaplow and Shavell (1992) argue that an important component of an arms race exists in adversarial expenditures, whereas spending on compliance may be helpful in ensuring rules are correctly implemented to avoid harmful externalities. Combining such an analysis with estimates of the impact of litigation on improving economic incentives could generate an account nearly as persuasive as that on education mentioned above. Similarly, output in engineering could be disaggregated into three components: new product development, where theory suggests imperfect appropriability creates positive externalities (Spence, 1976); operations, where externalities should be limited; and reverse engineering, where negative business-stealing externalities predominate (Hirshleifer, 1971).

Second, to our knowledge, very little is known about labor-substitution patterns across professions. The closest evidence known to us comes from the causal impact on earnings of quasi-random assignment across fields of study at universities (Hastings et al., 2013; Kirkebøen et al., Forthcoming). But many of the professional choices studied in our paper are made conditional on a given undergraduate degree. Neither Hastings et al. (2013) nor Kirkebøen et al. (Forthcoming) identify substitution patterns in response to changes in material rewards. Studies of such substitution patterns are critical to determining optimal tax policy, but progress will likely require difficult-to-obtain long-term exogeneous variation in professional wages.

Finally, more work is needed on potential problems in implementing the large profession-specific interventions suggested by our analysis. For instance, misrepresenting income earned in one profession as coming from another might be feasible, limiting the ability of the government to enforce such targeting. Opening the door to profession-specific taxes and subsidies might offer the state flexibility in industrial policy that could be abused. Investigating the history of such challenges, such as targeted scholarships at universities for pursuing socially valuable professions, could shed light on the feasibility of larger-scale corrective taxes and subsidies.

Beyond these new lines of inquiry, future research could relax the assumptions of our analysis in two ways.

First, Piketty et al. (2014) and RS consider models in which individuals simultaneously engage in both rent-seeking and productive activities. By contrast, in our model, each unit of output from a profession causes the same externality. Yet the greatest benefit from reallocation might arise within professions. Take finance, for example. Hirshleifer (1971) argues that high-speed trading is oversupplied, whereas Posner and Weyl (2013) show long-term price discovery of large bubbles is just as likely to be undersupplied as innovative breakthroughs. Uniform income taxation, even by profession, is unlikely to be a sufficient tool to achieve such reallocation. Mechanisms that do are an exciting direction for future research.

Second, this paper assumes all statutory taxes are paid. Tax avoidance would substantially change the analysis, especially if avoidance is profession-specific. For example, if financiers can avoid labor-income taxation by representing their income as capital income against which a lower rate is charged, income taxation might make finance more attractive rather than less. Incorporating calibrated avoidance considerations into our model is an interesting direction for future research.

References

Angrist, Joshua D. and Krueger, Alan B. (1991). 'Does Compulsory School Attendance Affect Schooling and Earnings?', *Quarterly Journal of Economics* 106(4), 979–1014.

Arrow, K. J. and Hahn, F. H. (1971), *General Competitive Analysis*, Amsterdam: North-Holland.

Atkinson, Anthony B., Piketty, Thomas and Saez, Emmanuel. (2011). 'Top Incomes in the Long Run of History', *Journal of Economic Literature* 49(1), 3–71.

Bagwell, Kyle. (2007), The Economic Analysis of Advertising, *in* **Mark Armstrong and Robert H. Porter**., eds, 'Handbook of Industrial Organization', Vol. 3, Amsterdam: North-Holland.

Bai, Jennie, Philippon, Thomas and Savov, Alexi. (Forthcoming). 'Have Financial Markets Become More Informative?', *Journal of Financial Economics* .

Bakija, Jon, Cole, Adam and Heim, Bradley T. (2012), Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U. S. Tax Return Data. http://web.williams.edu/Economics/wp/BakijaColeHeimJobsIncomeGrowthTopEarners.pdf.

Baumol, William J. (1990). 'Entrepreneurship: Productive, Unproductive and Destructive', *Journal of Political Economy* 98(5), 893–921.

Becker, Gary S. and Murphy, Kevin M. (1993). 'A Simple Theory of Advertising as a Good or a Bad', *Quarterly Journal of Economics* 108(4), 941–964.

Bertrand, Marianne and Mullainathan, Sendhil. (2001). 'Are CEOs Rewarded for Luck? The Ones Without Principals Are', *Quarterly Journal of Economics* 116(3), 901–932.

Bloom, Nicholas, Eifert, Benn, Mahajan, Aprajit, McKenzie, David and Roberts, John. (2013). 'Does Management Matter? Evidence from India', *Quarterly Journal of Economics* 128(1), 1–51.

Brown, Peter. (2011). 'American Voters Split on Government Shutdown', *Quinnipiac University Poll* .

Card, David. (1999), The Causal Effect of Education on Earnings, *in* **Orley Ashenfelter and David Card**., eds, 'Handbook of Labor Economics', Vol. 3, Amsterdam: Elsevier, chapter 30, pp. 1801–1863.

CBO. (2005), Effective Marginal Tax Rates on Labor Income, Technical report, Congressional Budget Office.

Chetty, Raj. (2012). 'Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply', *Econometrica* 80(3), 969–1018.

Chetty, Raj, Friedman, John and Rockoff, Jonah E. (2014). 'Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes', *American Economic Review* 104(9), 2633–2679.

Chipman, John S. (1970). 'External Economies of Scale and Competitive Equilibrium', *Quarterly Journal of Economics* 84(3), 347–385.

CNN/ORC. (2011), Wall Street Attitudes Poll. http://i2.cdn.turner.com/cnn/2011/images/10/24/rel17e.pdf.

Colombi, Roberto. (1990), A New Model of Income Distributions: The Pareto Lognormal Differentstribution, *in* **Camilo Dagum and Michele Zenga**., eds, 'Income and Wealth Distribution, Inequality and Poverty', Berlin: Springer, pp. 18–32.

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

Dixit, Avinash and Norman, Victor. (1978). 'Advertising and Welfare', *Bell Journal of Economics* 9(1), 1–17.

Feldstein, Martin. (1999). 'Tax avoidance and the deadweight loss of the income tax', *Review of Economics and Statistics* 81(4), 674–680.

Frankfurt, Harry. (1987). 'Equality as a Moral Ideal', *Ethics* 98(1), 21–43.

French, Kenneth R. (2008). 'The Cost of Active Investing', *Journal of Finance* 63(4), 1537–1573.

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

Goldin, Claudia, Katz, Lawrence F., Hausman, Naomi and Ward, Bryce. (2013), Harvard and Beyond Project, Survey data project, Harvard University.

Gul, Faruk and Pesendorfer, Wolfgang. (2008), The Case for Mindless Economics, *in* **Andrew Caplin and Andrew Schotter**., eds, 'The Foundations of Positive and Normative Economics: A Handbook', New York: Oxford University Press, pp. 3–42.

Gurun, Umit G. and kirand Amit Seru, Gregor Matvos. (Forthcoming). 'Advertising Expensive Mortgages', *Journal of Finance* .

Harberger, Arnold C. (1978). 'On the Use of Distributional Weights in Social Cost-Benefit Analysis', *Journal of Political Economy* 86(2), S87–S120.

Hastings, Justine S, Neilson, Christopher A and Zimmerman, Seth D. (2013), Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile. http://goo.gl/rlgAcc.

Hicks, J. R. (1939), *Value and Capital*, Oxford: Clarendon.

Hirshleifer, Jack. (1971). 'The Private and Social Value of Information and the Reward to Inventive Activity', *American Economic Review* 61(4), 561–574.

Jaffe, Adam B. (1989). 'Real Effects of Academic Research', *American Economic Review* 79(5), 957–970.

Jones, Charles I and Williams, John C. (1998). 'Measuring the Social Return to R & D', *Quarterly Journal of Economics* 113(4), 1119–1135.

Kaplow, Louis and Shavell, Steven. (1992). 'Private versus Socially Optimal Provision of Ex Ante Legal Advice', *Journal of Law, Economics, and Organization* 8(2), 306–320.

Kirkebøen, Lars, Leuven, Edwin and Mogstad, Magne. (Forthcoming). 'Field of Study, Earnings, and Self-Selection', *Quarterly Journal of Economics* .

Malmendier, Ulrike and Tate, Geoffrey. (2009). 'Superstar CEOs', *Quarterly Journal of Economics* 124(4), 1593–1638.

Mankiw, N. Gregory. (2010). 'Spreading the Wealth Around: Reflections Inspired by Joe the Plumber', *Eastern Economic Journal* 36, 285–298.

Marshall, Alfred. (1890), *Principles of Economics*, New York: Macmillan.

Marx, Karl. (1867), *Das Kapital, Kritik der Politischen Ökonomie*, Hamburg, Germany: Verglag von Otto Meissner.

Mirrlees, J. A. (1971). 'An Exploration in the Theory of Optimal Taxation', *Review of Economic Studies* 38(2), 175–208.

Murphy, Kevin M. (1986), Specialization and Human Capital. unpublished dissertation, University of Chicago.

Murphy, Kevin M., Shleifer, Andrei and Vishny, Robert W. (1991). 'The Allocation of Talent: Implications for Growth', *Quarterly Journal of Economics* 106(2), 503–530.

Murphy, Kevin M. and Topel, Robert H. (2006). 'The Value of Health and Longevity', *Journal of Political Economy* 114(5), 871–904.

Parker, Kim. (2012). 'Yes, the Rich Are Different', *Pew Social and Demographic Trends* August 27.

Patterson, Scott. (2010), *The Quants*, New York: Random House.

Philippon, Thomas. (2010). 'Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?', *American Economic Journal: Microeconomics* 2(3), 158–182.

Philippon, Thomas and Reshef, Ariell. (2012). 'Wages and Human Capital in the U.S. Finance Industry: 1909–2006', *Quarterly Journal of Economics* 127(4), 1551–1609.

Piketty, Thomas and Saez, Emmanuel. (2003). 'Income Inequality in the United States, 1913–1998', *Quarterly Journal of Economics* 118(1), 1–41.

Piketty, Thomas, Saez, Emmanuel and Stantcheva, Stefanie. (2014). 'Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities', *American Economic Journal: Economic Policy* 6(1), 230–271.

Posner, Eric and Weyl, E. Glen. (2013). 'Benefit-Cost Analysis for Financial Regulation', *American Economic Review Papers and Proceedings* 103(3), 393–397.

Rand, Ayn. (1957), *Atlas Shrugged*, New York: Random House.

Reyes, Loreto, Rodríguez, Jorge and Urzua, Sergio S. (2013), Heterogeneous Economic Returns to Postsecondary Degrees: Evidence from Chile. https://www.econ.umd.edu/research/papers/621.

Rizzo, John A. (1999). 'Advertising and Competition in the Ethical Pharmaceutical Industry: The Case of Antihypertensive Drug', *Journal of Law and Economics* 42(1), 89–116.

Rothschild, Casey and Scheuer, Florian. (2014*a*), Optimal Taxation with Rent-Seeking. http://www.stanford.edu/∼scheuer/rent_seeking.pdf.

Rothschild, Casey and Scheuer, Florian. (2014*b*), A Theory of Income Taxation under Multidimensional Skill Heterogeneity. http://www.stanford.edu/∼scheuer/multidim.pdf.

Saez, Emmanuel. (2001). 'Using Elasticities to Derive Optimal Income Tax Rates', *The Review of Economic Studies* 68(1), 205–229.

Saez, Emmanuel. (2002). 'Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses', *Quarterly Journal of Economics* 117(3), 1039–1073.

Saez, Emmanuel, Slemrod, Joel and Giertz, Seth H. (2012). 'The Elasticity of Taxable Income With Respect to Marginal Tax Rates: A Critical Review', *Journal of Economic Literature* 50(1), 3–50.

Saez, Emmanuel and Stantcheva, Stefanie. (Forthcoming). 'Generalized Social Welfare Weights for Optimal Tax Theory', *American Economic Review* .

Saxenian, AnnaLee. (2006), *The New Argonauts*, Cambridge, MA: Harvard University Press.

Slemrod, Joel, Yitzhaki, Shlomo, Mayshar, Joram and Lundholm, Michael. (1994). 'The Optimal Two-Bracket Linear Income Tax', *Journal of Public Economics* 53(2), 269–290.

Spence, A. Michael. (1976). 'Product Selection, Fixed Costs and Monopolistic Competition', *Review of Economic Studies* 43(2), 217–235.

Vickrey, William. (1945). 'Measuring Marginal Utility by Reactions to Risk', *Econometrica* 13(4), 319–333.

Weinzierl, Matthew. (2014). 'The Promise of Poistive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice', *Journal of Public Economics* 118, 128–142.

Wilson, Robert B. (1993), *Nonlinear Pricing*, Oxford: Oxford University Press.

Appendix

A Proofs and Derivations

A.1 Section 2

Formalization of no-bunching condition Let w denote the total productivity of a worker. His optimal income choice is $y^* = \arg\max_y y - T(y) - \phi(y/w)$. This optimal y^* moves smoothly in response to perturbations in T as long as it strictly maximizes utility for all w. This property holds when the second-order condition is strictly satisfied when the first-order condition holds.

The first-order condition is $1 - T'(y) - \phi'(y/w)/w = 0$, and the second-order condition is $-T''(y) - \phi''(y/w)/w^2 < 0$. Because $\phi(h) = h^{1/1+\sigma}/(1/1+\sigma)$, $\phi''(h) = \phi'(h)/(\sigma h)$. Applying this equality, we find $\phi''(y/w)/w^2 = \phi'(y/w)/(\sigma y w) = (1 - T'(y))/(\sigma y)$, where the last equality used the first-order condition. The second-order condition thus simplifies to $-T''(y) - (1 - T'(y))/(\sigma y) < 0$, which reduces to the inequality in Assumption 1.

Proof of Lemma 1. Note $h^*(\theta)$ depends on $T(\cdot)$ only through $T'(\cdot)$. A worker prefers profession i over j if and only if

$$y_i^*(\theta) - T(y_i^*(\theta)) - \phi(h_i^*(\theta)) + \psi_i(\theta) > y_j^*(\theta) - T(y_j^*(\theta)) - \phi(h_j^*(\theta)) + \psi_j(\theta).$$

This equation depends on T through the intensive margin and through the difference $T(y_i^*(\theta)) - T(y_j^*(\theta))$, but from (3), this difference depends only on $T'(\cdot)$ and not on T_0. Therefore, $i^*(\theta)$ depends on $T'(\cdot)$ and not on T_0, so the equilibrium depends only on $T'(\cdot)$.

Let R^a and R^b be two revenue requirements, and let T^a and T^b be the respective optimal tax rates. Let $\mathcal{U}^a$ and $\mathcal{U}^b$ be the respective values of the government's objective function under T^a and T^b. Consider the tax schedule $T^a + R^b - R^a$ formed by adding $R^b - R^a$ to T_0^a but leaving $(T^a)'$ unchanged. This tax schedule raises R^b in revenue, and the value of the objective function under it is $\mathcal{U}^a + R^b - R^a$ because the equilibrium is the same as under T^a. By the optimality of T^b, $\mathcal{U}^a + R^b - R^a \leq \mathcal{U}^b$. We can make the same argument with a and b reversed to obtain $\mathcal{U}^b + R^a - R^b \leq \mathcal{U}^a$. It follows that $\mathcal{U}^a + R^b - R^a = \mathcal{U}^b$, so $T^a + R^b - R^a = T^b$ and $(T^a)' = (T^b)'$.

□

Calculation of externality ratios We show the externality ratios solve the system of equations

$$e_j = \sum_{i=1}^{n} \frac{\partial \log E_i(Y_1^*, ..., Y_n^*)}{\partial Y_j} \left(a_i + \sum_{k=1}^{n} b_{i,k} e_k \right), \tag{16}$$

where a_i and $b_{i,k}$ are constants that depend on the equilibrium under consideration. Each a_i represents the direct effect of an increase in productivity in i on welfare. $b_{i,k}$ measure the changes to output in each

k, which themselves cause externalities. All of these coefficients depend on both intensive and allocative margin labor-supply adjustments.

To derive these constants, consider the effect of increasing log productivity in i on hours, income, and utility. The first-order condition for each worker is $h_i^* = w_i^\sigma(1 - T'(y_i^*))^\sigma$, so $y_i^* = w_i^{1+\sigma}(1 - T'(y_i^*))^\sigma$. Differentiating this equation yields $dy_i^*/d\log w_i = (1+\sigma)(1 - T'(y_i^*))y_i^*/(1 - T'(y_i^*) + \sigma y_i^* T''(y_i^*))$. The change in the cost of effort is $\phi'(h_i^*)dh_i^*/d\log w_i = y_i^*(1-T'(y_i^*))d\log h_i^*/d\log w_i$. Solving for this derivative and substituting yields a total change in the effort cost of $\sigma y_i^*(1 - T'(y_i^*))(1 - T'(y_i^*) - y_i^* T''(y_i^*))/(1 - T'(y_i^*) + \sigma y_i^* T''(y_i^*))$. Finally, the change in utility is simply $y_i^*(1 - T'(y_i^*))$ from the envelope theorem.

The change in productivity induces switching on the allocative margin. Denote $\Theta_i = \{\theta \mid \{i\} \subset I^*(\theta)\}$ to be the set of workers in i (or indifferent) and denote $\partial\Theta_i = \{\theta \mid \{i\} \subsetneq I^*(\theta)\}$ to be the set of workers indifferent between i and another profession. For these latter type of workers, define $i'(\theta)$ to be a uniquely chosen element of $I^*(\theta)$ not equal to i. The productivity change induces a switch between i and $i'(\theta)$. Because the worker is indifferent to the post-tax utility of these professions, the change in the pre-tax utility is $T(y_i^*(\theta)) - T(y_{i'(\theta)}^*(\theta))$. The switch also causes an externality. Output rises in i by $y_i^*(\theta)$ and falls in $i'(\theta)$ by $y_{i'(\theta)}^*(\theta)$, leading to a change in social welfare of $e_i y_i^*(\theta) - e_{i'} y_{i'(\theta)}^*(\theta)$.

We can now calculate the constants. For ease of notation, we define f_i on Θ_i by $f_i(\theta) = y_i^*(\theta)(1 - T'(y_i^*(\theta)))f(\theta)$. Then

$$
\begin{aligned}
a_i &= \int_{\Theta_i} \frac{1 + \sigma T'(y_i^*(\theta)) + \sigma y_i^*(\theta) T''(y_i^*(\theta))}{1 - T'(y_i^*(\theta)) + \sigma y_i^*(\theta) T''(y_i^*(\theta))} f_i(\theta) d\theta + \int_{\partial\Theta_i} \left(T(y_i^*(\theta)) - T\left(y_{i'(\theta)}^*\right)\right) f_i(\theta) d\theta \\
b_{i,i} &= \int_{\Theta_i} \frac{1+\sigma}{1 - T'(y_i^*(\theta)) + \sigma y_i^*(\theta) T''(y_i^*(\theta))} f_i(\theta) d\theta + \int_{\partial\Theta_i} y_i^*(\theta) f_i(\theta) d\theta \\
b_{i,k} &= -\int_{\Theta_i \cap \Theta_k} y_{i'(\theta)}^* f_i(\theta) d\theta,
\end{aligned}
$$

where the last equation is defined for $k \neq i$.

Returning to Equation 16, note it takes the form $e = a + Be$, where lowercase letters are n dimensional column vectors and the upper case B is an $n \times n$, not necessarily symmetric, matrix. This has solution $e = [I - B]^{-1} a$. Because B need not be symmetric, neither does $I - B$ need to be. Properties of $I - B$ likely play an important role in the existence, uniqueness, and stability in this model. A natural conjecture by analogy to classical general equilibrium theorem (Arrow and Hahn, 1971) is that a sufficient condition for at most a single equilibrium to exist, which is stable, is that $-[I - B]$ is globally stable (stable for every value of the vector Y) in the sense of Hicks (1939) that all the principal minors of $I - B$ are positive. This condition, combined with some boundary conditions, likely ensures existence of such an equilibrium. This conjecture is consistent with our empirical findings that when externalities (and thus B) become too large, we cannot find an equilibrium, or multiple local steady states exist. Investigating these issues in general equilibrium theory at a general level with greater depth is beyond the scope of this paper, however.

A.2 Section 3

Optimal top tax in general three-profession model The following lemma gives the first-order condition τ^* must satisfy, which is the equation in Proposition 1 computed for the current example.

Lemma 3. *In this example, the optimal top tax rate τ^* solves the equation*

$$0 = \frac{\sigma}{1-\tau^*}(\tau^* + s_h e_h + s_l e_l) + \frac{2\beta\tilde{s}_h s_l (r^\alpha - 1)(1-\tau^*)^\sigma}{\alpha(r+1)}(\tau^*(r-1) + r e_h - e_l), \tag{17}$$

where $\tilde{s}_h$ is the share of skilled top earners that choose H, conditional on ability.

Proof. The equation follows from Proposition 1 in the limit of large y. The intensive part follows immediately. We show a constant limiting tax rate is optimal, which shows $T'' = 0$ at high income levels. For the allocative part, we must calculate $f_S(y)/f(y)\Delta_T(y)$ and $f_S(y)/f(y)\Delta_e(y)$ for large y.

We first solve for the profession shares for skilled workers. No externalities affect L or H, so the total productivity of a worker in either profession i is private productivity a_i. The solution to the optimization $\max_y y - T(y) - \phi(y/a_i(\theta))$ is $a_i(\theta)^{1+\sigma}(1-\tau)^{1+\sigma}/(1+\sigma)$. A skilled worker chooses H if and only if $\psi_h(\theta) - \psi_l(\theta) > -(1-\tau)^{1+\sigma} a_l(\theta)^{1+\sigma}(r-1)/(1+\sigma)$. The difference between two variables following Gumbel distributions with the same scale parameter is logistically distributed, so $\tilde{s}_l = F^{\mathcal{L}}(-2\beta(1-\tau)^{1+\sigma}(r-1)(1+\sigma)^{-1}(r+1)^{-1} - \Delta\overline{\psi})$, where $\Delta\overline{\psi} = \overline{\psi}_h - \overline{\psi}_l$ and $F^{\mathcal{L}}$ is the CDF of the standard logistic distribution. Note this share is independent of income.

Conditional on being skilled with productivity a_l, the CDF of ψ_i is $F_\psi(2\beta(1+r)^{-1}a_l^{-(1+\sigma)}\psi_i - \overline{\psi}_i)$. Therefore, the conditional CDF of $\Delta\psi = \psi_h - \psi_l$ equals $F^{\mathcal{L}}(2\beta(1+r)^{-1}a_l^{-(1+\sigma)}\Delta\psi - \Delta\overline{\psi})$. The PDF of $\Delta\psi$ equals $2\beta(1+r)^{-1}a_l^{-(1+\sigma)}f^{\mathcal{L}}$. A standard fact about the logistic distribution is that $f^{\mathcal{L}} = F^{\mathcal{L}}(1-F^{\mathcal{L}})$. Therefore, the conditional measure of indifferent workers equals $2\beta(1+r)^{-1}a_l^{1+\sigma}\tilde{s}_h\tilde{s}_l$. Using the formula for income in the text, we simplify this expression to $2\beta(1+r)^{-1}(1-\tau)^\sigma y_l^{-1}\tilde{s}_h\tilde{s}_l$.

At income y, the share of workers in L is s_l. The measure of workers who are skilled and for whom $y_l^*(\theta) = y$ equals $s_l f(y)/\tilde{s}_l$. Therefore, the measure of such workers who are indifferent between H and L is $2\beta(1+r)^{-1}(1-\tau)^\sigma y^{-1}\tilde{s}_h s_l f(y)$. It follows that

$$\Delta_T(y) = \int_{y/r}^{y} \frac{\tau(r-1)y'}{y}\frac{2\beta(1-\tau)^\sigma\tilde{s}_h s_l f(y')}{(1+r)y' f(y)}dy' = \frac{\tau(r-1)2\beta(1-\tau)^\sigma\tilde{s}_h s_l (r^\alpha - 1)}{\alpha(1+r)},$$

where we have used the fact that the distribution of income is Pareto with parameter α. This fact follows because the ability distributions are all Pareto with parameter $\alpha(1+\sigma)$ and log income equals $1+\sigma$ times log ability. Similarly,

$$\Delta_e(y) = \int_{y/r}^{y} \frac{(re_h - e_l)y'}{y}\frac{2\beta(1-\tau)^\sigma\tilde{s}_h s_l f(y')}{(1+r)y' f(y)}dy' = \frac{(re_h - e_l)2\beta(1-\tau)^\sigma\tilde{s}_h s_l (r^\alpha - 1)}{\alpha(1+r)}.$$

Putting these equations together and factoring yields the desired result.

□

As can be seen from Lemma 3, the relative weight on τ_{int} scales with the labor-supply elasticity σ, whereas the relative weight on τ_{all} scales with the profession-switching sensitivity β. The larger β is, the more sensitive profession choices are to relative income and the greater the importance of the allocative margin in the optimal tax.

Note $\tau^\star$ is always in the interval between τ_{int} and τ_{all} because only in this interval will the two terms in equation 3 have opposite signs and thus only there can the equation be satisfied. Thus, $\tau^\star$ must be a convex combination of τ_{int} and τ_{all}, though no simple closed-form solution exists for the relevant weights.

A.3 Section 4

Identification of externality coefficients First, we derive (10). Note $Y_i = H_i E_i(Y_1, ..., Y_n)$, where $H_i = \int_\Theta a_i(\theta) h_i(\theta) d\theta$. Given how it is defined, the partial derivative $\partial Y/\partial Y_j$ equals $(\partial Y/\partial H_j)/(\partial Y_j/\partial H_j)$, where these partial derivatives are calculated holding each H_i constant:

$$\frac{\partial Y_i}{\partial H_j} = 1_{i,j} E_j + H_i \sum_k \frac{\partial E_i}{\partial Y_k} \frac{\partial Y_k}{\partial H_j} = 1_{i,j} E_j + Y_i \sum_k \frac{\partial E_i / E_i}{\partial Y_k} \frac{\partial Y_k}{\partial H_j}.$$

Define the quasi-Jacobian matrix J by $J = \{(Y_i/E_i)\partial E_i/\partial Y_k\}_{i,k}$. Let $\partial \mathcal{Y}/\partial H_j$ be the column matrix whose ith entry equals $\partial Y_i/\partial H_j$. Then the above equation can be written in matrix form as

$$\frac{\partial \mathcal{Y}}{\partial Y_j} = 1_j E_j + J \frac{\partial \mathcal{Y}}{\partial Y_j},$$

where 1_j is the vector with a 1 in the jth spot and 0 otherwise. Therefore,

$$\frac{\partial \mathcal{Y}}{\partial Y_j} = (I - J)^{-1} 1_j E_j \implies \frac{\partial Y}{\partial Y_j} = 1'(I - J)^{-1} 1_j,$$

where we have used the facts that $\partial Y_j/\partial H_j = E_j$ and $\partial Y/\partial Y_j = 1'\partial \mathcal{Y}/\partial Y_j$. Note that when externalities are absent, J is identically 0 so $\partial Y/\partial Y_j = 1$. Finally, directly taking the derivatives of E using our specification gives the equation for J in the text.

Proof of Lemma 2. We begin by proving statements made in the text before the lemma. Consider (11). Conditional on $i^*(\theta) = i$, the worker's maximization is $\max_{y_i} y_i - T(y_i) - \phi\left(y_i a_i^{-1} E_i(Y_0, ..., Y_n)^{-1}\right)$. The solution satisfies $y_i = (1 - T'(y_i))^\sigma (a_i E_i(Y_0, ..., Y_n))^{1+\sigma}$. By using $T = T_{2005}$ and solving for a_i, we immediately obtain (11).

Now we prove (13). The result of the maximization just described is $y_i - T(y_i) - (\sigma/(1+\sigma))(1 - T'(y_i))y_i$. Using this equation and (11), as well as the definition for relative utility in the text, we derive (13).

Next, we prove the distribution of a_{-i} conditional on a_i follows a Gaussian copula. By definition, the $\Phi^{-1}(F_i^a(a_i))$ are jointly normal with mean 0 and covariance Σ_n. A standard result is that a multivariate normal conditioned on some of the variates is also multivariate normal, with mean and covariance given by formulas. Applying these formulas, we obtain that conditional on $\Phi^{-1}(F_i^a(a_i))$, the remaining $\Phi^{-1}(F_j^a(a_j))$ are multivariate normal with mean $\Phi^{-1}(F_i^a(a))\varrho$ and covariance $\Sigma_{n-1} - \varrho'\varrho$.

We finally move on to the lemma itself. Consider a worker with productivity vector a. She chooses i to maximize $U_i^*(\theta)$, where θ restricted to productivity is a. This optimization is equivalent to maximizing $U_i^*(\theta) - \psi_i(\theta) + \psi_i(\theta) = n^{-1}\left(\sum_j a_j^{1+\sigma}\right)\widetilde{u}_i(a) + \psi_i(\theta)$, which is equivalent to maximizing $\widetilde{u}_i(a) + \psi_i(\theta)/\left(n^{-1}\sum_j a_j^{1+\sigma}\right)$. This latter term is distributed as $\beta^{-1}(\overline{\psi}_i + F_\psi)$, where F_ψ is a standard Gumbel distribution. If we let this Gumbel draw be $\widetilde{\psi}_i(\theta)$, the worker is choosing i to maximize $\beta\widetilde{u}_i(a)+\overline{\psi}_i+\widetilde{\psi}_i(\theta)$. A result from Gumbel distributions is that the probability that $A_i + B_i > A_j + B_j$ for all j when B_j are independent standard Gumbel distributions is $e^{A_i}/\sum_j e^{A_j}$. Applying this result, we conclude the share of workers with productivity a who choose i is

$$\Pr(i^*(\theta) = i \mid \theta|_a = a) = \frac{e^{\beta\widetilde{u}_i(a)+\overline{\psi}_i}}{\sum_j e^{\beta\widetilde{u}_j(a)+\overline{\psi}_j}}. \tag{18}$$

To prove (14), we compute in two different ways the share of all skilled workers such that the worker is in i at productivity a_i. First is the density of the i empirical productivity distribution $\widetilde{f}_i^a(a_i)$, times the share s_i of all workers in i, divided by the measure of skilled workers $1 - s_0$. This product gives the left side of (14). Alternatively, consider the probability that any worker would have productivity in i equal to a_i were she to choose i. This probability is $f_i^a(a_i)$. But only some of such workers choose i. To compute that conditional probability, we integrate over the conditional distribution of a_{-i}, using (18) as the probability of choosing i for each productivity profile. The result is the right side of (14).

$\square$

B Calibration Details

The Pareto parameter α is calibrated as follows. Using data from Bakija et al. (2012) described in Section 4, we use the fact that 16.97% of US income in 2005 went to those earning at least \$280,000, in order to calculate that the average income of such earners equals \$800,000. We calculate this average using the aggregate income and number of earners covered in the Bakija et al. (2012) data. In a Pareto distribution, the average value over a threshold (within the support of the distribution) equals the value of that threshold times $\alpha/(\alpha-1)$. Therefore, this fraction equals about 2.86 (the value we use does not involve intermediate rounding), leading to an α of 1.5.

We use engineering, research, and teaching to represent L, and finance and law to represent H. The 99$^{\text{th}}$ percentile incomes are the average of these statistics across these professions. Specifically, we use the average weighted by the share of each profession among earners in the top 1% of the total income distribution. These shares, given by Bakija et al. (2012), are 4.6% for engineering, 1.8% for research, 0.8% for teaching, 13.9% for finance, and 8.4% for law. The 99$^{\text{th}}$ percentile incomes are reported in Table 2.

Our simplified approximation to externality ratios is just the ratio of a profession's aggregate spillover to that profession's aggregate income. The reason this ratio is an approximation is because such externality estimates consider only income, whereas the true externality ratios should consider utility, which includes the cost of labor. Using the formulas for the externality ratios from the previous appendix, we calculate that in the case in which the tax is linear, using income externalities underestimates the true externality ratios by a factor of $1+\sigma T'$. Because $\sigma = 0.24$ and the average tax rate in the United States is around 30%, we underestimate the externality ratios by only 7%, which is small enough to ignore for this illustrative example. Section 4 uses a more complex method that does not involve an approximation.

According to data we use from the BLS, 6.3% of the labor force in 2005 were in engineering, research, or teaching, and 1.3% were in finance or law. Therefore, we assume 7.6% of the total population is skilled and 17% of skilled workers choose H over L. They make this choice given r and the prevailing taxes in 2005, which we assume for simplicity are constant at a 30% marginal rate. The resulting preference parameters imply the share $\tilde{s}_h$ of highly skilled workers choosing H was 23.3%.[31] From $\tilde{s}_h$, we infer $\overline{\psi}_h - \overline{\psi}_l$, which allows us to recalculate the s_i as tax rates move around.[32]

C Estimation Details

C.1 Professional classifications

We map the IRS profession classifications in Bakija et al. (2012) to ours in the following manner. *Art* is "Arts, media, sports," *Engineering* is "Computer, math, engineering, technical (nonfinance)," *Finance* is "Financial professions, including management," *Law* is "Lawyers," *Management* is "Executive, non-finance, salaried" plus "Executive, non-finance, closely held business" plus "Manager, non-finance, salaried" plus "Manager, non-finance, closely held business," *Medicine* is "Medical," *Operations* is "Business operations (nonfinance)," *Real Estate* is "Real estate," *Research* is "Professors and scientists," and *Sales* is "Skilled sales (except finance or real estate)." Bakija et al. (2012) use a combined category "Government, teachers, social services." We apportion worker counts from this category to *Teaching* and *Other* using the ratio in the BLS data of teachers (SOCs below) to government workers (NAICS = 92). We subtract teachers in government from the count of government workers (this adjustment is de minimis). The remainder of *Other* is "Blue collar or lminimsl service" plus "Unknown" plus "Farmers & ranchers" plus "Pilots" plus "Supervisor, non-finance, salaried" plus "Supervisor, non-finance, closely held business." We use Tables 2 and 3, "Percentage of primary taxpayers in the top one [resp. 0.1] percent of the distribution of income (excluding capital gains) that are in each profession."

In the BLS data, we aggregate SOCs into our professions using a classification similar to that in Bakija et al. (2012). This similarity justifies matching the BLS and IRS data. The exact list of SOCs we use to define each profession in the BLS is below:

Art: Art directors (27-1011), Craft artists (27-1012), Fine artists, including painters, sculptors, and illustrators (27-1013), Multi-media artists and animators (27-1014), Artists and related workers, all other

[31] The measure of workers in H earning y equals $\tilde{s}_h/\tilde{s}_l$ times the measure of workers in L earning y/r. Thus, $s_h/s_l = r^\alpha \tilde{s}_h/\tilde{s}_l$. This equation allows us to obtain $\tilde{s}_h$ and $\tilde{s}_l$ from s_h and s_l, as the former two sum to 1.

[32] As shown in the proof of Lemma 3, $\tilde{s}_l = F^{\mathcal{L}}(-2\beta(1-\tau)^{1+\sigma}(r-1)(1+\sigma)^{-1}(r+1)^{-1} - \overline{\psi}_h + \overline{\psi}_l)$, where $F^{\mathcal{L}}$ is the standard logistic distribution. We solve this equation using $\tilde{s}_l = 76.7\%$ and $\tau = 35\%$. To recalculate the s_i, pick an ability level and let s_u^* be the share of workers at that ability in either U or L who are in U. This share does not depend on ability or the tax rate at high ability levels. Then $s_u = s_u^*/(s_u^* + (1-s_u^*)\tilde{s}_l + (1-s_u^*)r^\alpha \tilde{s}_h)$. Using this equation, we calculate s_u^* at $\tau = 35\%$ and then use the updated $\tilde{s}_h$ and $\tilde{s}_l$ to update s_u at different tax levels, and then use $s_l = s_u(1-s_u^*)\tilde{s}_l/s_u^*$ and $s_h = s_u(1-s_u^*)r^\alpha \tilde{s}_h/s_u^*$.

(27-1019), Commercial and industrial designers (27-1021), Fashion designers (27-1022), Floral designers (27-1023), Graphic designers (27-1024), Interior designers (27-1025), Merchandise displayers and window trimmers (27-1026), Set and exhibit designers (27-1027), Designers, all other (27-1029), Actors (27-2011), Producers and directors (27-2012), Athletes and sports competitors (27-2021), Dancers (27-2031), Choreographers (27-2032), Music directors and composers (27-2041), Musicians and singers (27-2042), Radio and television announcers (27-3011), Public address system and other announcers (27-3012), Broadcast news analysts (27-3021), Reporters and correspondents (27-3022), Public relations specialists (27-3031), Editors (27-3041), Technical writers (27-3042), Writers and authors (27-3043), Interpreters and translators (27-3091), Media and communication workers, all other (27-3099).

Engineering: Computer programmers (15-1021), Computer software engineers, applications (15-1031), Computer software engineers, systems software (15-1032), Aerospace engineers (17-2011), Agricultural engineers (17-2021), Biomedical engineers (17-2031), Chemical engineers (17-2041), Civil engineers (17-2051), Computer hardware engineers (17-2061), Electrical engineers (17-2071), Electronics engineers, except computer (17-2072), Environmental engineers (17-2081), Health and safety engineers, except mining safety engineers and inspectors (17-2111), Industrial engineers (17-2112), Marine engineers and naval architects (17-2121), Materials engineers (17-2131), Mechanical engineers (17-2141), Mining and geological engineers, including mining safety engineers (17-2151), Nuclear engineers (17-2161), Petroleum engineers (17-2171), Engineers, all other (17-2199).

Finance: Chief executives (11-1011) in Finance and Insurance (NAICS = 52), General and operations managers (11-1021) in Finance and Insurance (NAICS = 52), Financial managers (11-3031), Financial analysts (13-2051), Personal financial advisors (13-2052), Securities, commodities, and financial services sales agents (41-3031).

Law: Lawyers (23-1011), Administrative law judges, adjudicators, and hearing officers (23-1021), Arbitrators, mediators, and conciliators (23-1022), Judges, magistrate judges, and magistrates (23-1023).

Management: Chief executives (11-1011) outside Finance and Insurance (NAICS $\neq$ 52), General and operations managers (11-1021) outside Finance and Insurance (NAICS $\neq$ 52), Advertising and promotions managers (11-2011), Marketing managers (11-2021), Sales managers (11-2022), Public relations managers (11-2031), Administrative services managers (11-3011), Computer and information systems managers (11-3021), Compensation and benefits managers (11-3041), Training and development managers (11-3042), Human resources managers, all other (11-3049), Industrial production managers (11-3051), Purchasing managers (11-3061), Transportation, storage, and distribution managers (11-3071), Farm, ranch, and other agricultural managers (11-9011), Farmers and ranchers (11-9012), Construction managers (11-9021), Education administrators, preschool and child care center/program (11-9031), Education administrators, elementary and secondary school (11-9032), Education administrators, postsecondary (11-9033), Education administrators, all other (11-9039), Engineering managers (11-9041), Food service managers (11-9051), Funeral directors (11-9061), Gaming managers (11-9071), Lodging managers (11-9081), Medical and health services managers (11-9111), Natural sciences managers (11-9121), Social and community service managers (11-9151), Managers, all other (11-9199).

Medicine: Chiropractors (29-1011), Dentists, general (29-1021), Oral and maxillofacial surgeons (29-1022), Orthodontists (29-1023), Prosthodontists (29-1024), Dentists, all other specialists (29-1029), Anesthesiologists (29-1061), Family and general practitioners (29-1062), Internists, general (29-1063), Obstetricians and gynecologists (29-1064), Pediatricians, general (29-1065), Psychiatrists (29-1066), Surgeons (29-1067), Physicians and surgeons, all other (29-1069), Podiatrists (29-1081).

Operations: Agents and business managers of artists, performers, and athletes (13-1011), Purchasing agents and buyers, farm products (13-1021), Wholesale and retail buyers, except farm products (13-1022), Purchasing agents, except wholesale, retail, and farm products (13-1023), Claims adjusters, examiners, and investigators (13-1031), Insurance appraisers, auto damage (13-1032), Compliance officers, except agriculture, construction, health and safety, and transportation (13-1041), Cost estimators (13-1051), Emergency management specialists (13-1061), Employment, recruitment, and placement specialists (13-1071), Compensation, benefits, and job analysis specialists (13-1072), Training and development specialists (13-1073), Human resources, training, and labor relations specialists, all other (13-1079), Logisticians (13-1081), Management analysts (13-1111), Meeting and convention planners (13-1121), Business operations specialists, all other (13-1199).

Real Estate: Property, real estate, and community association managers (11-9141), Appraisers and assessors of real estate (13-2021), Real estate brokers (41-9021), Real estate sales agents (41-9022).

Research: Computer and information scientists, research (15-1011), Animal scientists (19-1011), Food scientists and technologists (19-1012), Soil and plant scientists (19-1013), Biochemists and biophysicists (19-1021), Microbiologists (19-1022), Zoologists and wildlife biologists (19-1023), Biological scientists, all other (19-1029), Conservation scientists (19-1031), Epidemiologists (19-1041), Medical scientists, except epidemiologists (19-1042), Life scientists, all other (19-1099), Astronomers (19-2011), Physicists (19-2012), Atmospheric and space scientists (19-2021), Chemists (19-2031), Materials scientists (19-2032), Environmental scientists and specialists, including health (19-2041), Geoscientists, except hydrologists and geographers (19-2042), Hydrologists (19-2043), Physical scientists, all other (19-2099), Economists (19-3011), Sociologists (19-3041), Urban and regional planners (19-3051), Anthropologists and archeologists (19-3091), Geographers (19-3092), Historians (19-3093), Political scientists (19-3094), Social scientists and related workers, all other (19-3099), Business teachers, postsecondary (25-1011), Computer science teachers, postsecondary (25-1021), Mathematical science teachers, postsecondary (25-1022), Architecture teachers, postsecondary (25-1031), Engineering teachers, postsecondary (25-1032), Agricultural sciences teachers, postsecondary (25-1041), Biological science teachers, postsecondary (25-1042), Forestry and conservation science teachers, postsecondary (25-1043), Atmospheric, earth, marine, and space sciences teachers, postsecondary (25-1051), Chemistry teachers, postsecondary (25-1052), Environmental science teachers, postsecondary (25-1053), Physics teachers, postsecondary (25-1054), Anthropology and archeology teachers, postsecondary (25-1061), Area, ethnic, and cultural studies teachers, postsecondary (25-1062), Economics teachers, postsecondary (25-1063), Geography teachers, postsecondary (25-1064), Political science teachers, postsecondary (25-1065), Psychology teachers, postsecondary (25-1066), Sociology teachers, postsecondary (25-1067), Social sciences teachers, postsecondary, all other (25-1069), Health specialties teachers, postsecondary (25-1071), Nursing instructors and teachers, postsecondary (25-1072), Education teachers, postsecondary (25-1081), Library science teachers, postsecondary (25-1082), Criminal justice and law enforcement teachers, postsecondary (25-1111), Law teachers, postsecondary (25-1112), Social work teachers, postsecondary (25-1113), Art, drama, and music teachers, postsecondary (25-1121), Communications teachers, postsecondary (25-1122), English language and literature teachers, postsecondary (25-1123), Foreign language and literature teachers, postsecondary(25-1124), History teachers, postsecondary (25-1125), Philosophy and religion teachers, postsecondary (25-1126).

Sales: Advertising sales agents (41-3011), Insurance sales agents (41-3021), Sales representatives, services, all other (41-3099), Sales representatives, wholesale and manufacturing, technical and scientific products (41-4011), Sales representatives, wholesale and manufacturing, except technical and scientific products (41-4012), Sales engineers (41-9031), Sales and related workers, all other (41-9099).

Teaching: Preschool teachers, except special education (25-2011), Kindergarten teachers, except special education (25-2012), Elementary school teachers, except special education (25-2021), Middle school teachers, except special and vocational education (25-2022), Vocational education teachers, middle school (25-2023), Secondary school teachers, except special and vocational education (25-2031), Vocational education teachers, secondary school (25-2032), Special education teachers, preschool, kindergarten, and elementary school (25-2041), Special education teachers, middle school (25-2042), Special education teachers, secondary school (25-2043).

Other: All SOCs not listed above.

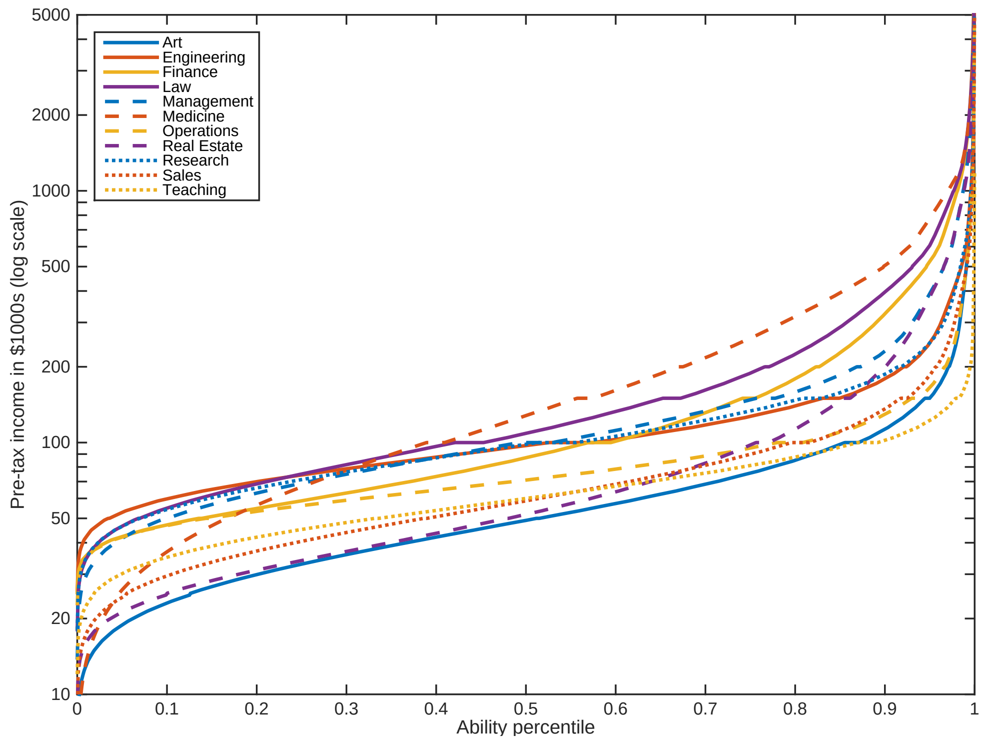
C.2 Income by ability

Figure 3 plots our estimated marginal distributions of pre-tax income for each quantile of the income distribution in our baseline estimation, where $\rho = 1$ so that individuals have a single dimension of ability. The figure therefore represents the pre-tax earnings from which an individual at the given quantile of the ability distribution could choose if she entered each of the professions.

The patterns are quite intuitive. At low levels of ability, stable professions, such as Engineering and Law, have the highest earnings, whereas "starving artists" are at the bottom. Toward the top end of the income distribution, finance, law and medicine are most lucrative, but even art does well given superstar effects. Teaching is at the bottom given the limited upside.

FIGURE 3

Income by Ability Levels



Notes: This figure plots, for each profession, the income earned by a worker in that profession whose ability a_i is at each percentile of the underlying distribution of ability across all workers in the economy. We compute the realized income at equilibrium under the optimal income tax given in Figure 2.

C.3 Calibration of profession-switching sensitivity

Let $s_{i,t}$ denote the share of the population in (other) finance at time t. We denote the share of workers flowing into finance by $s^f_{i,t}$. The stock $s_{i,t}$ and flow $s^f_{i,t}$ are related by the differential equation $\dot{s}_{i,t} = \delta(s^f_{i,t} - s_{i,t})$, where $\delta > 0$ is a replacement rate. Solving this equation from some reference time 0 yields

$$s_{i,t} = e^{-\delta t} s_{i,0} + \int_0^t \delta e^{-\delta(t-\tau)} s^f_{i,\tau} d\tau.$$

The share of the stock that is replaced in a year equals $1 - e^{-\delta}$, and we use this expression to calibrate δ. For instance, if 1/30 of the stock is replaced annually because people work for 30 years, we choose δ such that $1/30 = 1 - e^{-\delta}$.

Some elasticity exists that expresses the flows into finance as a function of relative log wages. The specification we adopt is $s^f_{i,t} = b_0 + b_1 \log \widetilde{w}_{i,t}$, where $\widetilde{w}_{i,t}$ is the relative wage in finance. If relative wages are a random walk, a worker's best predictor of lifetime relative wages is the current value, and hence present relative wages alone guide labor flows. We want to compute b_1, which we will use as our moment to match.

To estimate b_1, we use Philippon and Reshef (2012)'s annual data on $\widetilde{w}_{i,t}$ and $s_{i,t}$. They show that in the period 1950-1980, these series were roughly flat. Both increased sharply after 1980, and the increases are close to linear in time. We therefore assume that in 1980, which we denote $t = 0$, finance employment was in equilibrium, so that $s^f_{i,0} = s_{i,0}$. It follows that $s^f_{i,t} = s_{i,0} + b_1(\log \widetilde{w}_{i,t} - \log \widetilde{w}_{i,0})$. We also assume that log relative wages increased linearly, so that $\log \widetilde{w}_{i,t} = \log \widetilde{w}_{i,0} + b_w t$. The data strongly bear out this linearity. When we regress $\log \widetilde{w}_{i,t}$ on time, the estimated coefficient for $b_w = 0.0478$ and has a t-stat of 22; the $R^2 = 95\%$. It follows that $s_{i,t} = s_{i,0} + b_1 b_w \int_0^t \delta e^{-\delta(t-\tau)} \tau d\tau$, so that

$$b_1 = \frac{\delta}{t\delta + e^{-\delta t} - 1} \frac{s_{i,t} - s_{i,0}}{b_w}.$$

The quantity $s_{i,t} - s_{i,0}$ equals the increase in the share of the population in finance between 1980 and 2005. Philippon and Reshef (2012) directly reports these numbers; the share increased from 0.35% to 0.87%, yielding $s_{i,t} - s_{i,0} = 0.52\%$. In the above formula, $t = 25$. Therefore, b_1 emerges as a function of our input for δ. Using the calibration method described earlier, we arrive at the following table:

Tenure (years)	δ	$\widehat{b_1}$
0	∞	0.0044
1	0.63	0.0045
5	0.18	0.0054
10	0.095	0.0069
20	0.049	0.0101
30	0.033	0.0135
40	0.025	0.0170
50	0.020	0.0204
100	0.010	0.0378

C.4 Externality calculations

Engineering We use Murphy et al. (1991)'s preferred estimates that are restricted to the 55 countries in which more than 10,000 students are in college. Their result is that a one percentage-point increase in the share of students studying engineering raises real per-capita GDP growth by 0.054% points. According to the OECD, 10.7% of college students in the United States in 2005 studied engineering. The total externalities from engineering therefore amount to $(10.7)(0.054) = 0.6\%$ of GDP. In our context, we interpret the externalities from engineering as 0.6% of total income.[33]

Finance We interpret the entirety of French (2008)'s estimates as negative externalities from finance. In 2005, he estimates these externalities at 0.63% of US stock market capitalization, or \$90.7 billion. This externality amounts to -1.4% of the total income we calculate in Table 2, which is \$6.3 trillion.

[33] The total income we measure in Table 2 is less than GDP because it excludes capital gains, transfers, and investment. We assume the engineering externality raises each component of GDP by the same proportion, so it raises the total income measure we focus on by 0.6%.

Law Murphy et al. find a one percentage-point increase in the share of students in a country studying law lowers real per-capita GDP growth by 0.078% points. We again interpret this effect as a one-time change to the level of output. According to the OECD, 2.4% of students in the United States study law. Externalities from law therefore equal $-(2.4)(0.078) = -0.2\%$ of GDP.

Research Jaffe's model allows university research to have a direct effect on commercial patents as well as an indirect effect through influencing industrial R&D. His preferred estimate is that the total elasticity of patents with respect to university research is 0.6. The direct elasticity of industrial R&D on patents is 0.94, and industrial R&D expenditures are six times larger than university research expenditures. Therefore, a dollar in university research is equivalent to $6(0.6)/0.94 = 3.83$ dollars spent on industrial R&D in terms of resulting patents. According to the National Science Foundation, $45 billion was spent on university R&D in 2005. Using the estimate from Jaffe, we conclude the total externality from this activity was $172 billion, which amounts to 2.7% of total income.

Sales Informative and purely rational consumption theories of advertising imply advertising will tend to be undersupplied in most cases (Becker and Murphy, 1993), whereas persuasive theories suggest it will be oversupplied (Dixit and Norman, 1978). But although empirical efforts have sought to quantify the welfare effects of advertising in particular markets, such as pharmaceuticals (Rizzo, 1999) and subprime mortgages (Gurun and kirand Amit Seru, Forthcoming), none attempts a comprehensive, profession-wide study, so we are hesitant to use these estimates.

Teaching Card (1999) reviews the literature on the causal effect of education on earnings and finds results between a 0.05 and 0.15 log increase for each year of schooling. We use the midpoint of this interval, 0.1, which also equals the estimate of Angrist and Krueger (1991). The annual social product of teaching therefore equals $e^{0.1} - 1 = 10.5\%$ of total income. According to our estimates of profession-specific income distributions in Table 2, the private product of teachers equals 3.2% of total income. The total externalities of teachers amount to the difference, which is 7.3%.

Chetty et al. (2014) calculate that a one-standard-deviation increase in teacher quality raises eventual student earnings by 1.34%. They also calculate that the present value of future earnings for a middle-school student is $468,000 in 2005 dollars.[34] In our model, all variations in teacher pay come from differences in quality $a_i(\theta)$. The standard deviation of teacher pay in our data equals $27,000. The total pay equals $203 billion. According to the Digest of Education Statistics published by the National Center for Education Statistics, the number of students in all elementary and secondary schools in the United States in 2005 was 54 million.[35] Our data contain 4.2 million teachers. The total social product of teachers is

$$(\$203 \text{ billion}/\$27,000)(1.34\% * \$468,000)(54 \text{ million}/4.2 \text{ million}) = \$606 \text{ billion}.$$

[34] They use a 5% discount rate and assume earnings grow 2% annually.

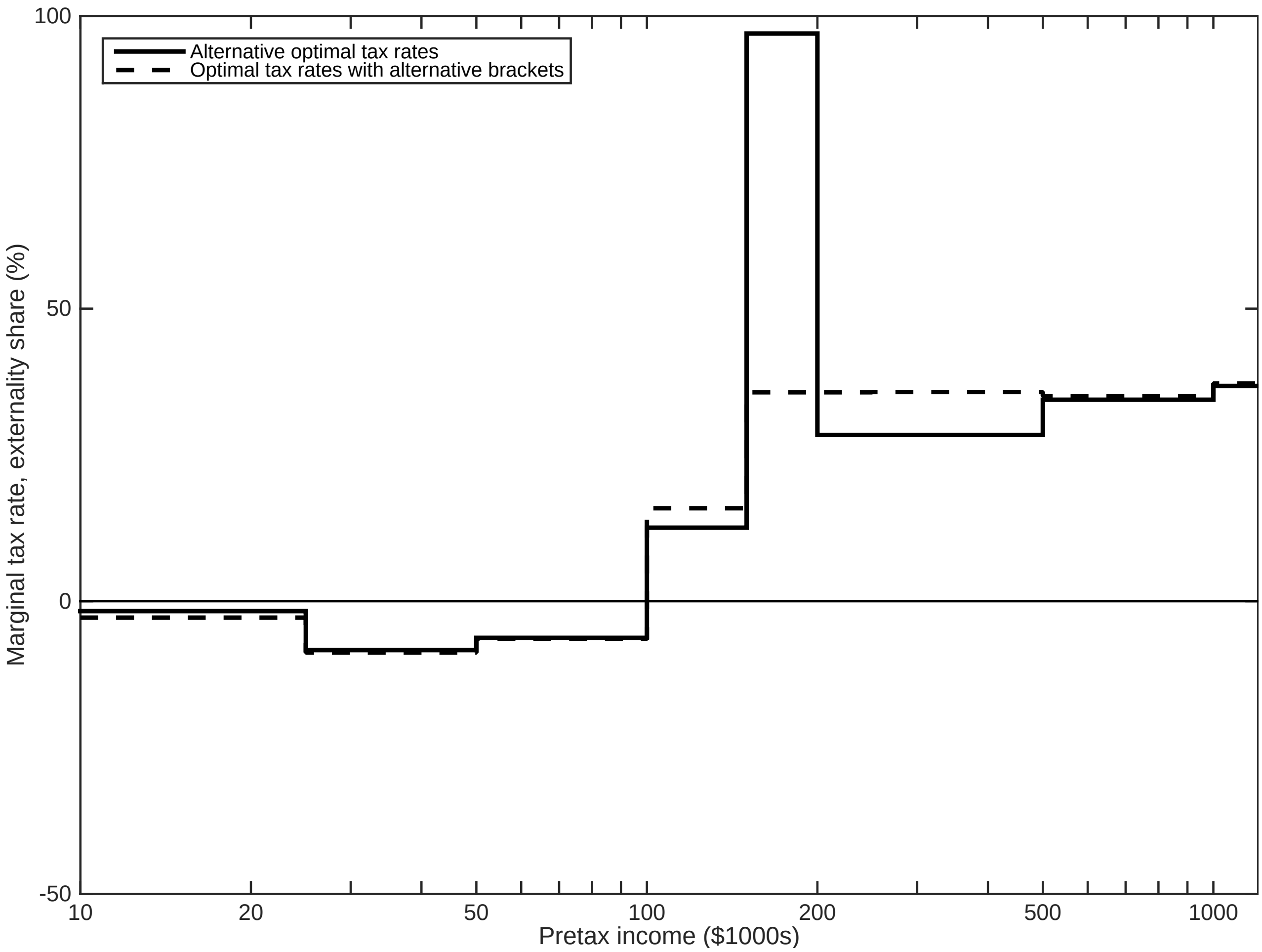
[35] The NCES reports enrollments of 53.4 million in 2000 and 54.9 million in 2010, which we average. Data accessed at `http://nces.ed.gov/programs/digest/d13/tables/dt13_105.20.asp`.

This social product equals 9.6% of the $6.3 trillion of total income in the economy. This number is slightly less than the 10.5% figure we calculated using the social returns to education, but it is very close.

C.5 Alternative local optimum for marginal tax rates

As discussed in the text, a second local optimum exists for marginal tax rates under the baseline parameters in which welfare is slightly higher ($22 per person). We choose not to focus on it because it is likely an artifact of the way the brackets are constructed. It is present on only the smallest bracket (in log terms), and it disappears when we change the $150k-$200k bracket to $150k-$250k. The figures below reproduce Figure 2 for the secondary optimum and also for the case in which the bracket is enlarged.

FIGURE 4
Alternative Optimal Marginal Taxes Schedules



Notes: This figure displays an alternative local maximum for the marginal tax schedule using the same baseline assumptions and parameters behind Figure 2. The dashed line displays the only optimum we find when the $150k-$200k bracket is changed to $150k-$250k. While other local optima cannot be ruled out with certainty, we confirm that using the vector of optimal tax rates from the solid line as the seed value for the optimization when computing the dotted line does alter it.