

Taxation and the Allocation of Talent

Benjamin B. Lockwood

University of Pennsylvania

Charles G. Nathanson

Northwestern University

E. Glen Weyl

Microsoft Research and Yale University

Taxation affects the allocation of talented individuals across professions by blunting material incentives and thus magnifying nonpecuniary incentives of pursuing a “calling.” Estimates from the literature suggest that high-paying professions have negative externalities, whereas low-paying professions have positive externalities. A calibrated model therefore prescribes negative marginal tax rates on middle-class incomes and positive rates on the rich. 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 of which are very uncertain given existing empirical evidence.

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 National Bureau of Economic Research 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.

Electronically published August 24, 2017

[*Journal of Political Economy*, 2017, vol. 125, no. 5]

© 2017 by The University of Chicago. All rights reserved. 0022-3808/2017/12505-0007\$10.00

I. Introduction

The allocation of talented individuals across professions varies widely over time and space.¹ 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 that such externalities not only exist but are large (Murphy and Topel 2006; French 2008). In this paper, we quantitatively evaluate the impact of nonlinear 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, Saez, and Stantcheva 2014; Rothschild and Scheuer 2014, 2016) 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 nonpecuniary benefits. Higher marginal tax rates incent workers to “follow their passion” by reducing the relative after-tax pecuniary compensation of the more lucrative professions. To the extent that better-paying professions generate negative (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 II and with a simple example in Section III, because they involve discrete jumps in income, these substitution patterns make the first-order approach of Mirrlees (1971) invalid.

¹ According to data from the Harvard and Beyond Project, more than twice as many male Harvard alumni from the 1969–72 cohorts pursued careers in academia and in non-financial management as pursued careers in finance. Twenty years later, careers in finance were 50 percent more common than in academia and were comparable to those in non-financial management. Murphy, Shleifer, and Vishny (1991) document that in 1970 among the 91 countries studied by Barro (1991), the 25th percentile of the share of college students studying engineering equals 3.8 percent and the 75th percentile equals 14.31 percent; the 25th percentile of the share of college students studying law equals 2.7 percent and the 75th percentile equals 11.2 percent.

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 that these externalities, although highly uncertain, are likely to be huge and quite heterogeneous.² Our main findings for optimal policy are the following:

1. The optimal income tax features top rates of about 36 percent, 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.
2. Although the optimal nonlinear, profession-general tax rates differ significantly from zero, they achieve only small welfare gains (1.3 percent) relative to *laissez-faire*. By contrast, profession-targeted policies can achieve much more. We show that an optimal linear subsidy to research professions achieves more than 40 times the welfare gains of our baseline optimal tax.
3. 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 that these understudied patterns are crucial to determining optimal tax policy.

Throughout this paper, we analyze only efficiency, rather than redistributive, gains from taxation. We focus on pure efficiency maximization because it highlights as sharply as possible the role of substitution patterns. In particular, efficiency maximization implies (see Sec. II.B) that 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 affects optimal tax rates.

² Murphy and Topel (2006) estimate that medical research generates a positive externality of more than 20 percent of GDP, whereas French (2008) calculates that the financial profession's income includes 1.5 percent of GDP in rent seeking.

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 that individuals receive their social contribution. Perhaps surprisingly, we show that such a theory is able to account for the broad outlines of existing US income taxation. We also believe that 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 2016).

In addition to our focus on allocation and efficiency, our analysis departs from the literature in several other ways. First, the allocative margin we introduce allows workers to choose between earning distinct income levels in different professions. In contrast, Rothschild and Scheuer (2014, 2016) assume a concave utility function of continuous effort in different activities, which rules out changes in income for marginal workers who switch activities. Second, 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. Third, our analysis is primarily quantitative. The present model incorporates many professions and is estimated using various data sources; previous literature on how income taxation should respond to externalities has involved primarily qualitative, illustrative models. Finally, in contrast to Piketty et al. (2014) and Rothschild and Scheuer (2014, 2016), 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.

II. A Model of Optimal Income Taxation with Externalities

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

A. 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 $a = (a_1, \dots, a_n)$ is a vector of profession-specific productivities and $\psi = (\psi_1, \dots, \psi_n)$ is a vector of the nonpecuniary utility the worker receives in each profession. The distribution of types θ among workers is given by a nonatomic 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 $i(\theta)$.³ 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 $i \neq i(\theta)$.

Following Rothschild and Scheuer (2014, 2016), we assume that externalities in this economy operate through production. 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 $E_i(Y_1, \dots, Y_n)$ as in Rothschild and Scheuer’s studies.⁴

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

$$y_i(\theta) = a_i(\theta)h_i(\theta)E_i(Y_1, \dots, Y_n),$$

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, nonpecuniary utility ψ , and an hours cost function $\phi(\cdot)$ for which $\phi'(\cdot), \phi''(\cdot) > 0$:

$$U(\theta) = y_{i(\theta)}(\theta) - T(y_{i(\theta)}(\theta)) - \phi(h_{i(\theta)}) + \psi_{i(\theta)}(\theta), \tag{1}$$

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 setup 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. In our specification, the cost of effort and the nonpecuniary benefit or cost of a profession are additively separable, thereby ruling out richer interactions between intensive and allocative labor supply decisions.

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

³ As shown below, the choice of i for each θ is unique in the equilibrium we consider.

⁴ A slight difference exists between our setup and that considered by Rothschild and Scheuer. 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 Rothschild and Scheuer’s studies, the i externalities do so only indirectly by drawing labor into j . Online app. D shows that the baseline quantitative results of our paper do not differ greatly when this alternative specification is used.

different 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, with all type θ workers making the same choice $i^*(\theta)$.⁵ We denote the total utility, income, and nonpecuniary 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:

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

In our estimation of the optimal income tax in Section IV, 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.

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 Section A, any solution to the worker’s first-order condition for hours is a strict local maximum (because of a negative second-order condition) when this inequality holds. Thus, the hours choice admits a unique maximum. Intuitively, uniqueness can fail when T' decreases too quickly because a worker can have an interior solution for hours at a low income and high tax rate as well as an interior solution for hours at a high income and low tax rate.

Given the quantitative focus of this paper, we follow Rothschild and Scheuer (2014, 2016) 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.

⁵ The specific choice of $i^*(\theta)$ for such workers is irrelevant for aggregates such as total utility and total output because the measure of indifferent workers equals zero: the dimension of the set of indifferent workers is smaller than the dimension of the set of all workers.

B. *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

$$T(y) = T_0 + \int_0^y T'(\bar{y}) d\bar{y}. \tag{2}$$

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

The equilibrium allocation of output Y_1^*, \dots, 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—because of quasi-linear utility—as the common lump-sum grant T_0 changes. This invariance condition means that 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 .

Owing 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 that 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 zero.

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 $\Theta(y) = \{\theta | y^*(\theta) = y\}$ and the set of such workers in profession i by $\Theta_i(y) = \{\theta | y^*(\theta) = y \text{ and } i^*(\theta) = i\}$. 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

$$\Theta_S(y) \equiv \{\theta | \text{there exist } i, i_h \in I^*(\theta) \text{ such that } y_{i_h}^*(\theta) < y < y_i^*(\theta)\},$$

and we denote their measure by $f_S(y)$.

The perturbation to $T'(\cdot)$ causes additional, secondary labor supply changes. Owing to externalities operating through the E_{is} , the intensive

and allocative margin adjustments just described 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 zero.

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 Section A of the Appendix, 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^*, \dots, 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 (1) and the revenue requirement $\int_{\Theta} T(y^*(\theta)) f(\theta) d\theta = R$, we write the government's objective function as

$$\int_{\Theta} U^*(\theta) f(\theta) d\theta = -R + \int_{\Theta} [y^*(\theta) - \phi(h^*(\theta)) + \psi^*(\theta)] f(\theta) d\theta. \quad (3)$$

The government maximizes the integral on the right: total income less disutility from labor plus nonpecuniary 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.

1. Intensive Margin

Consider a worker in profession i for whom $y^*(\theta) = y$. Denote the wage of this worker by $w_i(\theta) = a_i(\theta)E_i(Y_1^*, \dots, Y_n^*)$. 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 (3) 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 that $dY_i^* = w_i(\theta)dh^*(\theta)$, 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

$$\partial^{\text{int}} \int_{\Theta} U^*(\theta) f(\theta) d\theta = - \frac{\sigma y f(y)}{1 - T'(y) + \sigma y T''(y)} [T'(y) + e(y)] dT' dy. \quad (4)$$

2. Allocative Margin

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_i in which she earns $y_{i_i}^*(\theta)$, with $y_{i_i}^*(\theta) < y < y_{i_h}^*(\theta)$. The tax perturbation decreases the after-tax income, and hence utility, in i_h by $dT' dy$ while leaving utility in i_i unchanged. As a result, the worker switches from i_h to i_i .

This switch directly changes the value of the government's objective function (3) by

$$y_{i_i}^*(\theta) - \phi(h_{i_i}^*(\theta)) + \psi_{i_i}^*(\theta) - [y_{i_h}^*(\theta) - \phi(h_{i_h}^*(\theta)) + \psi_{i_h}^*(\theta)].$$

By the envelope theorem (because the worker receives the same utility in i_i and i_h), this difference equals the fiscal externality $T'(y_{i_i}^*(\theta)) - T'(y_{i_h}^*(\theta))$.

We define the average proportional tax change from switching workers by

$$\Delta_T(y) \equiv \int_{\Theta_s(y)} \frac{T(y_h^*(\theta)) - T(y_i^*(\theta))}{y} \frac{f(\theta)}{f_s(y)} d\theta.$$

The worker’s switch from i_h to i_i 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_i is therefore $e_i y_i^*(\theta) - e_h y_h^*(\theta)$. We define the average proportional externality change from switching workers by

$$\Delta_e(y) \equiv \int_{\Theta_s(y)} \frac{e_i y_i^*(\theta) - e_h y_h^*(\theta)}{y} \frac{f(\theta)}{f_s(y)} d\theta.$$

Recall that 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_i is $dT' dy$, the complete allocative margin change in the government objective from the perturbation to the tax schedule is

$$\partial^{\text{all}} \int_{\Theta} U^*(\theta) f(\theta) d\theta = -y f_s(y) [\Delta_T(y) + \Delta_e(y)] dT' dy. \tag{5}$$

In Saez’s (2002) analysis of the effect of taxes on the extensive margin of labor supply, the analogous expression depends only on the aggregate taxes paid by each individual and the density of workers indifferent to exiting the labor force.

3. Total First-Order Condition

The government’s first-order condition holds when the intensive margin change (4) and allocative margin change (5) to the government’s objective resulting from the tax perturbation sum to zero. Because we arbitrarily chose the income y at which $T'(\cdot)$ was perturbed, the first-order condition must hold for all y . Proposition 1 produces the first-order condition by adding (4) and (5) and then dividing by $-y dT' dy$.

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

$$0 = \underbrace{\frac{\sigma f(y)}{1 - T'(y) + \sigma y T''(y)}}_{\text{intensive}} [T'(y) + e(y)] + \underbrace{f_s(y) [\Delta_T(y) + \Delta_e(y)]}_{\text{allocative}}$$

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 zero, 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. Rothschild and Scheuer (2014, 2016) 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 that the optimal tax is related only to the relative size of the allocative and intensive responses,

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

and not to the level of these responses. For example (assuming a linear tax for the moment), suppose that σ and f_s doubled so that the sizes 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.

III. 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, $\lim_{y \rightarrow \infty} T'(y)$, for the optimal T . This rate measures the marginal tax rate the top earners face (although possibly only at extremely high incomes) and is similar to the object explored in Saez (2001) and Saez, Slemrod, and Giertz (2012). 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.

A. 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 .⁶ 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 . Above some level \bar{a}_i , the distribution of a_i in each profession is Pareto, with conditional probability distribution $\Pr(a_i(\theta) \geq a | a_i(\theta) \geq \bar{a}_i) = (\bar{a}_i/a)^{\alpha(1+\sigma)}$ for some $\alpha > 0$; in equilibrium, the Pareto exponent for the income distribution will equal α . For skilled workers, nonpecuniary utility ψ_i of working in $i = H$ or L is distributed as

$$\psi_i | a \sim \beta^{-1} [(a_L^{1+\sigma} + a_H^{1+\sigma})/2] (\bar{\psi}_i + F_\psi),$$

where the $\bar{\psi}_i$ are constants and F_ψ is a standard Gumbel distribution given by $F_\psi = e^{-e^{-\psi}}$. The term $(a_L^{1+\sigma} + a_H^{1+\sigma})/2$ is a normalization to ensure that nonpecuniary 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 one, whereas $E_U(Y_L, Y_H)$ 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 IV. In our baseline analysis, engineering, teaching, and research professions cause positive externalities, whereas law and finance lead to negative externalities. We find that, in the upper tail of the income distribution, the incomes in the first set of professions are lower than those in the second.

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 $\tau_{\text{int}} = -\lim_{y \rightarrow \infty} \epsilon(y)$. Hence,

$$\tau_{\text{int}} = -(s_H e_H + s_L e_L),$$

⁶ Formally, $\psi_H(\theta) = \psi_L(\theta) = -\infty$ for the unskilled workers and $\psi_U(\theta) = -\infty$ for the skilled workers.

where e_H and e_L are the externality ratios and s_i is the share of workers at top incomes in profession i .⁷ 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 Rothschild and Scheuer (2014, 2016), is optimal when profession choices are fixed.

The allocative optimal top rate looks quite different from τ_{int} . From proposition 1, $\Delta_T(y) + \Delta_\epsilon(y) = 0$ 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 $y_i^*(\theta) = a_i^{1+\sigma}(\theta)[1 - T'(y_i^*(\theta))]^\sigma$, $y_H^*(\theta) = r y_L^*(\theta)$ at high incomes.⁸ 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 $\Delta_\epsilon(y)$ is $(re_H - e_L)y_L^*(\theta)$, where τ is the top tax rate. The optimum sums these to zero and is

$$\tau_{\text{all}} = - \frac{re_H - e_L}{r - 1}.$$

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 Rothschild and Scheuer’s (2014, 2016) 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 as long as $e_H < e_L$. To see this relationship between τ_{all} and r dramatically, note that as $r \rightarrow 1$, 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 Section B of the Appendix.

⁷ Formally, $s_i = \lim_{y \rightarrow \infty} \int_{\theta(y)} f(\theta) d\theta / \int_{\theta(y)} f(\theta) d\theta$.

⁸ This statement requires $\lim_{y \rightarrow \infty} T'(y) < 1$ or $\sigma = 0$.

B. Calibration

To calculate τ^* , we need values of α , σ , β , r , e_H , e_L , $\bar{\psi}_H - \bar{\psi}_L$, 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 online appendix E. We also explore the different values of τ^* generated by a reasonable range of the parameters.

We first set α , the Pareto parameter for the tail of the US income distribution, to 1.5 on the basis of the ratio of the total income earned by the top 1 percent of the US income distribution to the 99th percentile of the income distribution. We set $\sigma = 0.23$ and $\beta = 1.3$ on the basis of 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 productivities in H and L are perfectly correlated, a worker in the 99th percentile of H incomes will also be in the 99th percentile of L incomes. We use a 99th 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 .

To calculate the externality ratios e_H and e_L , we take a weighted average of approximate externality ratios of the professions constituting each of H and L . As we discuss in online appendix E, dividing a profession's aggregate spillover by its aggregate income provides an accurate approximation of its externality ratio. In Section IV, 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, -0.08 for law, 0.14 for engineering, 11.06 for research, and 2.01 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 $e_L = 3.08$. Although these ratios are endogenous to the tax structure, we take them as fixed for the purposes of this calibration exercise. In the exercise in Section IV we allow the externality ratios to depend on the tax structure.

Finally, we set $\bar{\psi}_H - \bar{\psi}_L$ to match the share of workers in H and L , given the data and the tax rate in 2005. According to Bakija, Cole, and Heim (2012), 7.2 percent of the top 1 percent of earners in 2005 were in L and 22.3 percent were in H .

TABLE 1
OPTIMAL TOP TAX RATE FOR DIFFERENT PARAMETER VALUES (%)

	r	σ	β	e_H	e_L
Half of baseline	26	35	14	20	20
Baseline	24	24	24	24	24
Double baseline	18	14	30	30	28

NOTE.—This table reports the optimal top tax rate τ^* in the three-profession example defined in Sec. III. In each column, we hold all but the header parameter constant, varying that parameter to 50 percent, 100 percent, and 200 percent of its baseline value and reporting the optimal top tax rates. The baseline parameters are $r = 4.7$, $\sigma = 0.23$, $\beta = 1.3$, $e_H = -0.24$, and $e_L = 3.08$.

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 16 percent 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 $\tau = \tau^*$, $s_H = 0.19$ and $s_L = 0.08$, leading to an intensive optimal tax rate of $\tau_{\text{int}} = -0.20$. 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 , as H has only three times greater representation at high incomes. By contrast, $\tau_{\text{all}} = 1.12$, confiscating more than all of the marginal income of top earners. The total optimum τ^* balances the intensive and allocative optima at a rate of 0.24. This rate is reasonably close to the top tax rate of 0.39 that we calculate in Section IV.

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.⁹ 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 conflicting effects on τ^* .

⁹ 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.

IV. Empirical Strategy

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

A. Specification

Our specification of f is as follows. We separate the professions into n skilled professions $i = 1, \dots, n$ and one 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 $\psi_0(\theta) = -\infty$.

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*:

$$f(a_1, \dots, a_n) = \int_{\mu, \Sigma}^{\mathcal{N}} (\Phi^{-1}(F_1^a(a_1)), \dots, \Phi^{-1}(F_n^a(a_n))),$$

where Φ is the cumulative distribution function (CDF) of a unidimensional standard normal and $f_{\mu, \Sigma}^{\mathcal{N}}$ is the probability density function (PDF) of a multivariate normal with mean μ and covariance matrix Σ . This specification preserves the marginal productivity distributions F_i^a (i.e., $f|_{a_i} = f_i^a$ for all i) but allows correlation specified by Σ . We assume that all the diagonal elements of Σ are 1 and all off-diagonal elements are equal to some ρ that governs the correlation of productivity between every distinct pair of professions. 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, Rodríguez, and Urzúa (2013) and Kirkeboen, Leuven, and Mogstad (2016).

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

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

where $\bar{\psi}_i$ 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.¹⁰ Thus, we can interpret β as the *allocative sensitivity*, with higher β indicating greater elasticity of profession choice to changes in relative incomes across professions and thus to taxation. The constants $\bar{\psi}_i$ determine the average relative attractiveness of each profession i ; more workers enter i when $\bar{\psi}_i$ is higher.

Our specification of each externality function E_i has the form

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

where γ captures the returns to scale of the externalities; $\gamma = 1$ implies that externalities are linear in output; lower values of γ lead to diminishing marginal returns; and ϵ_{ij} captures the targeting of externalities across professions emphasized by Rothschild and Scheuer (2014, 2016). For our estimation, we reduce the dimensionality of these coefficients according to the specification

$$\epsilon_{ij} = \delta_{ij} \epsilon_j,$$

where $\delta_{i,j} \in \{0, 1\}$ if $i \neq j$. We thus restrict externalities coming from profession j to be uniform in magnitude across all professions i on which profession j has any impact. The $\epsilon_{i,i}$ remain unrestricted, allowing independence of the own externalities from those on other professions.

B. 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, \dots, f_n^y , the population shares in each profession s_0, \dots, s_n , and the marginal social products $\partial Y / \partial Y_0, \dots, \partial Y / \partial Y_n$ of output in each profession.¹¹ These inputs come from data we describe in Section IV.C. For the moment, we take the parameters $\sigma, \beta, \rho, \gamma$, and the matrix $\{\delta_{i,j}\}$ as given, postponing discussion of their selection until Sections IV.C.2 and IV.C.3. The outputs of the present estimation are $f_0^a, \dots, f_n^a, \bar{\psi}_1, \dots, \bar{\psi}_n$, and $\epsilon_0, \dots, \epsilon_n$.

¹⁰ This property holds exactly when taxes are linear, as only relative income $y_i / \sum_j y_j$ matters in (9) when $T'(\cdot)$ is constant.

¹¹ Rather than use the true nonlinear 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 number of optimal tax papers take a similar approach, including Saez (2001, 2002).

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 $\epsilon_0, \dots, \epsilon_n$ 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 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 Section C of the Appendix, the marginal social product equals

$$\frac{\partial Y}{\partial Y_j} = 1'(I - J)^{-1} 1_j, \tag{6}$$

where $1 = (1, \dots, 1)'$, $1_j = (0, \dots, 1, \dots, 0)'$ 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 (6) delivers $n + 1$ equations in the $n + 1$ unknowns $\epsilon_0, \dots, \epsilon_n$, allowing us to solve for these parameters numerically.

The subsequent step is to infer the empirical productivity distributions \tilde{f}_i^a that appear in the data. Selection of workers across professions determines these distributions, and hence the \tilde{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 :

$$a_i = y_i^{1/(1+\sigma)} [1 - T_{2005}^\sigma(y_i)]^{-\sigma/(1+\sigma)} E_i(Y_0, \dots, Y_n)^{-1}. \tag{7}$$

We define $y_i(a_i)$ to be the unique value of y_i that solves this equation given a_i .¹² Then

$$\tilde{f}_i^a(a_i) = y_i'(a_i) f_i^y(y_i(a_i)). \tag{8}$$

No selection occurs into or out of the low-skilled profession $i = 0$, so $f_0^a = \tilde{f}_0^a$.

The penultimate step is to calculate the relative utility $\tilde{u}_i(a)$ of working in i for a skilled worker with productivity vector a , ignoring profession

¹² That the solution to (7) is unique follows because the right side of (7) strictly increases in y_i because of assumption 1.

preference utility ψ . The relative utility $\tilde{u}_i(a)$ determines the share of workers with productivity a who choose to work in profession i . It is defined as

$$\tilde{u}_i(a) = [U_i^*(\theta) - \psi_i(\theta)] / \left(n^{-1} \sum_j a_j^{1+\sigma} \right),$$

where the productivity component of θ equals a . Section C of the Appendix derives the following closed-form expression for relative utility:

$$\tilde{u}_i(a) = \frac{y_i(a_i) - T_{2005}(y_i(a_i)) + \sigma(y_i(a_i) T'_{2005}(y_i(a_i)) - T_{2005}(y_i(a_i)))}{(1 + \sigma) n^{-1} \sum_j y_j(a_j) [1 - T'_{2005}(y_j(a_j))]^{-\sigma} E_j(Y_0, \dots, Y_n)^{-(1+\sigma)}}. \tag{9}$$

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 \tilde{f}_i^a , which are affected by selection. Given a_i , 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 $\Sigma_{-1} - \varrho'\varrho$, where Σ_{-1} is the top $(n - 1) \times (n - 1)$ block of Σ and $\varrho = (\rho, \dots, \rho)$ 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 $\bar{\psi}_i$ from the data.

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

$$\begin{aligned} \frac{s_i \tilde{f}_i^a(a_i)}{1 - s_0} &= f_i^a(a_i) \int_{\mathbb{R}^{n-1}} \frac{e^{\beta \tilde{u}_i(a) + \bar{\psi}_i}}{\sum_j e^{\beta \tilde{u}_j(a) + \bar{\psi}_j}} \\ &\times \int_{\Phi^{-1}(F_i^a(a_i))\varrho, \Sigma_{-1} - \varrho'\varrho} (\Phi^{-1}(F_1^a(a_1)), \dots, \Phi^{-1}(F_n^a(a_n))) da_{-i} \end{aligned} \tag{10}$$

for $1 \leq i \leq n$ and all $a_i > 0$. Here, \tilde{f}_i^a is the empirical productivity distribution in i calculated from (8) and $\tilde{u}_i(a)$ is the relative utility of working in i for a worker with productivity vector a calculated from (9). These solutions uniquely determine the f_i^a and are unique up to constant for the $\bar{\psi}_i$.

We solve equation (10) using a numerical solver.

C. Data

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 information technology 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 pretax 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 online appendix F.1.¹³

Their unit of observation is a tax return, of which there are 145,881,000 in 2005.¹⁴ 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.¹⁵

For each SOC, the BLS reports in the annual Occupational Employment Statistics (OES) database the number of workers as well as the 10th, 25th, 50th, 75th, and 90th income percentiles. The BLS produces the OES using surveys of nonfarm establishments. Using these data, we calculate the number of workers in each profession by summing the num-

¹³ One exception to Bakija et al.'s (2012) unique assignment of SOCs to professions concerns the SOCs for chief executives (11-1011) and general and operations managers (11-1021). Bakija et al. assign such workers to finance if the industry of the employer listed on the W-2 is finance and insurance (North American Industry Classification System [NAICS] code 52); they assign such workers to management otherwise.

¹⁴ 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>.

¹⁵ 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 percent and top 0.1 percent 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 percent and top 0.1 percent 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.

ber in each constituent SOC and then calculate s_i as the share of all workers in each profession.¹⁶

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.¹⁷ 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 $[y_{i,k}^-, y_{i,k}^+]$; $\sum_k s_{i,k} = s_i$, the total number of workers in i . Let \hat{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:

$$\hat{A}_i, \hat{\alpha}_i, \hat{\mu}_i, \hat{\nu}_i = \arg \max_{A,\alpha,\mu,\nu} \sum_k s_{i,k} \log [F_i^y(y_{i,k}^+) - F_i^y(y_{i,k}^-)],$$

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 $y < \$280,000$, $f_i^y(y) = \hat{f}_i^y(y)$ for $y \geq \$280,000$, and $A_i f_{\alpha_i,\mu_i,\nu_i}(\$280,000) = \hat{f}_i^y(\$280,000)$.¹⁸

¹⁶ To match Bakija et al.'s (2012) splitting of SOCs 11-1011 and 11-1021 into finance and management (see n. 13), we use BLS data for SOC-NAICS pairs to split these SOCs into a category in which the NAICS = 52 and one in which the NAICS \neq 52.

¹⁷ 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.

¹⁸ 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 reports summary statistics on the resulting distributions of income for each profession. Skilled professions constitute 18 percent of all workers, and skilled workers earn 42 percent 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 \$203,000 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 percent of the total population, they constitute 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 constitute 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 online appendix F.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.

TABLE 2
SUMMARY STATISTICS OF ESTIMATED PROFESSIONAL INCOME DISTRIBUTIONS

	Population Share (%)	Income Share (%)	Median Income	99th Percentile Income
Art	1.0	1.5	\$40,000	\$497,000
Engineering	2.0	4.1	\$73,000	\$452,000
Finance	.9	4.6	\$92,000	\$2,075,000
Law	.4	2.2	\$113,000	\$1,627,000
Management	3.9	13.4	\$78,000	\$1,273,000
Medicine	.5	3.0	\$203,000	\$1,348,000
Operations	2.4	3.7	\$51,000	\$368,000
Real estate	.3	1.1	\$50,000	\$1,393,000
Research	1.1	1.9	\$59,000	\$399,000
Sales	2.3	3.5	\$48,000	\$414,000
Teaching	3.2	3.4	\$43,000	\$126,000
Other	82.0	57.6	\$26,000	\$111,000

NOTE.—“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 “99th percentile income” are the 50th and 99th percentile incomes within each profession. The results describe the United States in 2005.

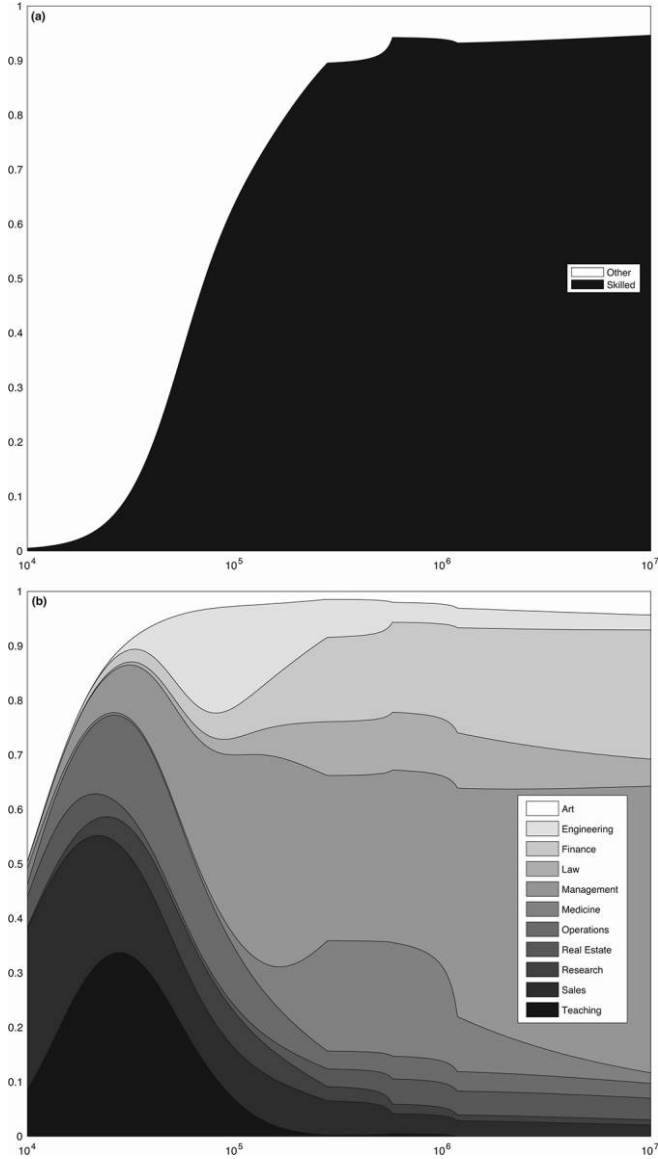


FIG. 1.—Distribution of workers at each income level: (a) all workers; (b) skilled workers. At each income y , the share of workers in profession i is $s_i f_i^s(y) / \sum_j s_j f_j^s(y)$, where s_i is the share of all workers in i and f_i^s is the PDF for income in i . The results describe the United States in 2005.

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. (2016). However, reconciling a significant, long-term comparative advantage at the top end of the income distribution with the massive reallocations of talent over time observed by Goldin and Katz (2008) and Philippon and Reshef (2012) is difficult. The sorting patterns such a comparative advantage would create are counterintuitive. For example, they imply that 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.¹⁹

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

¹⁹ When we vary ρ to 0.75, we reestimate the productivity distributions but we continue to use the values of σ and β estimated with $\rho = 1$. We do this because for $\rho = 0.75$ (and other similar values) we cannot find β to match the moment in (11). When comparative advantage is high, a positive productivity shock to 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 .

this derivative around the average empirical marginal tax rate T'_{2005} , holding each $E_i(Y_1, \dots, Y_n)$ 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 percent to 0.87 percent over this time, while the wages in finance relative to the rest of the (nonfarm) economy grew from 1.09 to 3.62.²⁰ To match these trends, we study 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 \bar{a} .²¹ The relative wage of finance equals

$$\tilde{w}_i = \int_{\theta_i} w_i(\theta)/s_i f(\theta) d\theta / \sum_{j \neq i} \int_{\theta_j} w_j(\theta)/(1 - s_j) f(\theta) d\theta,$$

where i denotes the index of finance. The moment we match is the fraction

$$\frac{\partial s_i / \partial \bar{a}}{\partial \log \tilde{w}_i / \partial \bar{a}}, \tag{11}$$

where each partial derivative is evaluated at $\bar{a} = 1$.²²

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 alternative 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. Simi-

²⁰ 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 in “other finance” in 2005 equals the number of workers we estimate in “finance” in 2005.

²¹ Specifically, for i corresponding to finance, $a_i(\theta)$ is replaced by $\bar{a}a_i(\theta)$ for all θ .

²² 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 that 5 percent of the worker stock is replaced each period. Online app. F.3 shows that this assumption leads to a value of the above derivative of 0.01. In the sensitivity analysis, we use replacement rates of 3 percent and 10 percent, which lead to respectively higher and lower values of β .

larly, we set $\delta_{i,j} = 1$ (uniform externalities) for all i and j as a baseline and consider alternative specifications in the sensitivity analysis.

To calculate the marginal social output from each profession, we draw on the literatures that estimate economywide externalities from various professions. Although we have done our best to faithfully represent the current literature, we emphasize that these estimates are highly uncertain extrapolations from heterogeneous and not easily comparable studies primarily aimed at estimands different from 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.²³ The private product is given by the "income share" column of table 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 $(3.4\% + 6.9\%)/3.4\% = 3.03$.

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 online appendix F.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 zero. In the cases in which these literatures offer conflicting results, we adopt one value as a baseline and use an alternative value for sensitivity analysis.

Arts.—Although some evidence, and a number of good theoretical arguments, suggest that the arts generate some positive externalities, we are unable to find a plausible basis for estimating the magnitude of these externalities and consequently assume zero 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 percent 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, Philippon, and Savov (2012) show that the infor-

²³ 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 that 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.

TABLE 3
AGGREGATE EXTERNALITIES BY PROFESSION: BASELINE ESTIMATES

	Externality as Share of Economy Income	Source	Method
Art	0
Engineering	.6%	Murphy et al. (1991)	Cross-country regression of GDP on engineers per capita
Finance	-1.5%	French (2008)	Aggregate fees for active vs. passive investing
Law	-.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	20.5%	Murphy and Topel (2006)	Willingness to pay for longevity gains from medical research
Sales	0
Teaching	6.9%	Card (1999)	Returns to education in excess of teacher salaries
Other	0

NOTE.—This table reports the total externalities for each profession as a share of the total income in the economy, which is \$5.9 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]).

mativeness 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).²⁴ We therefore interpret the entirety of French's estimates, which amount to 1.5 percent of total income in 2005, as negative externalities from finance.

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

²⁴ After this paper's acceptance, Bai, Philippon, and Savov (2016) revised the finding from Bai et al. (2012) to a statistically significant increase in the informational content of stock prices, while noting that it is difficult to determine how much of this increase reflects welfare-relevant revelatory price efficiency.

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 ways that do not actually reflect the CEO's marginal product. Piketty et al. (2014) argue that 60 percent of the CEO earnings elasticity with respect to taxes represents this rent-seeking behavior, implying that the negative externalities from management are 8.1 percent of total income. The other half of the literature argues that market forces can explain CEO compensation (Gabaix and Landier 2008) and suggests that therefore externalities are zero. 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 zero as our baseline and use the -8.1 percent figure in sensitivity analysis.

Medicine.—We could find no literature estimating the externalities of (nonresearch) medicine and so set the externality to zero to be conservative.

Operations.—This profession comprises 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 zero 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 that the annual gains of medical research equaled 25.72 percent of GDP from 1980 to 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 $.2572/(1 + .2572) = 20.5$ percent. Although this externality may be the largest externality from academia and science, this estimate is still conservative in assuming that 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 that the socially optimal amount of R&D activity is four times the observed amount, which we loosely translate into a three-times externality or 5.6 percent 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 3.0 percent externality, which we use as a lower-bound estimate in our sensitivity analysis.

Sales.—Although an extensive theoretical literature argues that 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 zero.

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 10.3 percent gain, which equals the midpoint of the numbers collected in Card's (1999) review. Because teachers earn 3.4 percent of economy income, we use a spillover from teaching of 6.9 percent of economy income.

We also compute the aggregate effect of teaching on earnings using Chetty, Friedman, and Rockoff's (2014) 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. estimate for a standard deviation in teacher quality to obtain an aggregate effect equal to 10.2 percent of economy income. This figure leads to a spillover of 6.8 percent 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.

V. 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 II as the average marginal externality of income earned by those with income equal to y . Proposition 1 showed that in the special case in which 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. Without 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.

A. Optimal Taxes

Given the underlying skill distributions, preference parameters, and externalities we estimate, we numerically calculate the marginal tax schedule

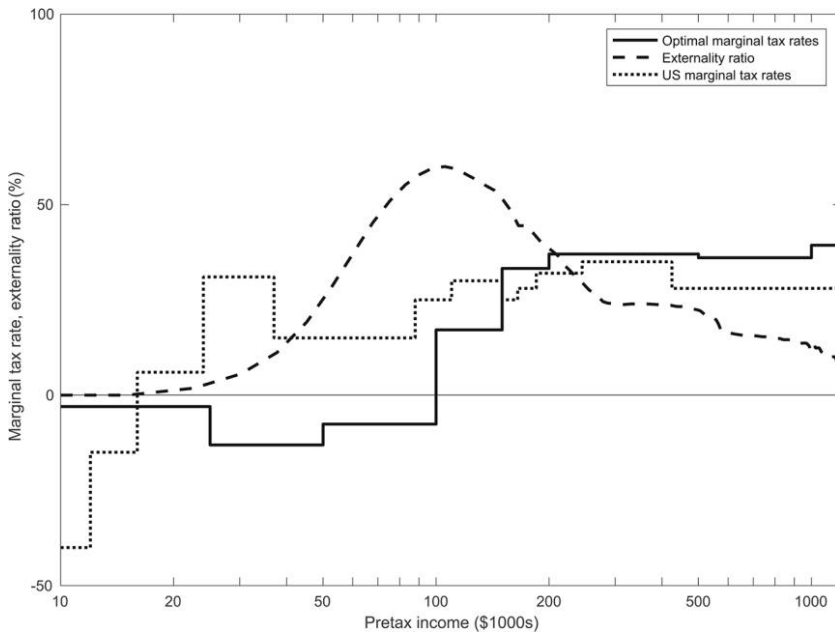


FIG. 2.—Optimal marginal taxes for the United States in 2005. The externality ratio is the quantity $e(y)$, defined in Section II to be the average marginal externality of income earned by workers with income equal to y . The marginal tax rate displays the optimal tax rates over the eight brackets specified, given the data and baseline assumptions explained in Sections IV.A and IV.C. US marginal tax rates are taken from figure 4 of Congressional Budget Office (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.

that maximizes social welfare. This procedure uses significant computational resources, so we restrict attention to schedules with eight brackets, with cutoffs at \$25,000, \$50,000, \$100,000, \$150,000, \$200,000, \$500,000, and \$1 million. 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 8 percent on income up to \$100,000 and then feature progressively increasing marginal rates after that. The top rate on income above \$1 million is 39.3 percent, and similar marginal rates hold for income above \$150,000 in other brackets.²⁵

²⁵ Figure 2 presents a local maximum for marginal tax rates. We did find a second local maximum in which welfare was slightly higher (\$24 per person). This alternative schedule is nearly identical to the one in fig. 2 except that the marginal rates in the \$150,000–\$200,000 bracket are much higher, over 95 percent. This optimum is driven by the piecewise linear structure of the tax code that induces sharp bunching at \$150,000 and a gap in incomes around \$250,000. The superiority of this local optimum is very fragile—if the

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 level of the tax schedule. Owing to the negative rates that last until \$100,000, net tax liabilities are negative up to \$146,000, so that a worker earning \$146,000 pays the same tax as a worker earning no income. Beyond this point, the marginal rate varies but on average is about 36 percent. The smallest tax liability is for a worker earning \$100,000, who receives a net income subsidy of \$7,800.

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 fig. 2) and because benefits to the poor phase out as income increases over this range (Congressional Budget Office 2005). The model prescribes negative rates on income all the way up to \$100,000, which is a much higher threshold than those used by income subsidies, such as the Earned Income Tax Credit.

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.

B. 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

\$200,000 bracket threshold is raised to \$250,000, then this local optimum generates welfare substantially lower than our baseline specification—and thus we do not focus on it for our main results, although we report it in online app. F.5.

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.

Panel A of table 4 reports the results. Relative to laissez-faire, the optimal tax raises average utility by \$898, or 1.3 percent. The tax achieves a smaller gain of 0.5 percent relative to the empirical economy, which is not surprising given that the tax used to model the empirical economy (a flat 30 percent tax) is close to the optimal 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 20.5 percent of the economy—suggest that a reallocation of talent to more productive professions could increase welfare by much more than the 1.3 percent achieved by the optimal income tax. Our findings that welfare gains are quite small are robust to all scenarios we consider in table 5 below except those with targeted subsidies to research; they never exceed 2.5 percent and in some scenarios are smaller than 0.5 percent. The largest welfare gains come when the allocative margin is strongest (when individuals switch elastically across professions or the intensive-margin elas-

TABLE 4
WELFARE AND THE ALLOCATION OF TALENT UNDER DIFFERENT TAX REGIMES

	Laissez-Faire	2005 US Data	Optimal Nonlinear Income Tax	Pre-Reagan Income Tax
A. Per Capita Welfare Gains Relative to Laissez-Faire				
Levels	...	\$343	\$898	-\$460
Percent5	1.3	-.6
B. Share of Skilled Workers in Each Profession (%)				
Art	4.5	5.2	4.5	5.5
Engineering	12.6	12.0	13.7	11.8
Finance	6.1	5.4	5.7	4.9
Law	3.6	2.9	3.3	2.5
Management	23.0	22.0	23.6	21.2
Medicine	5.1	3.1	3.5	2.3
Operations	12.0	12.9	12.2	13.5
Real estate	1.7	1.9	1.7	1.9
Research	6.0	6.3	6.2	6.5
Sales	10.7	11.9	10.9	12.5
Teaching	14.7	16.5	14.7	17.5

NOTE.—Each column reports results from simulating the estimated economy under different tax regimes. “Laissez-faire” denotes no marginal income tax, and “2005 US data” denotes the approximate 2005 marginal tax schedule used to estimate the model. “Optimal nonlinear income tax” refers to the optimal tax schedule shown in fig. 2, and “Pre-Reagan income tax” equals the 1980 US tax schedule.

ticity 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 induces little switching between professions, as 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. Panel B of table 4 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.

These results suggest that historical tax reductions are unlikely to have played a large role in 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 lower under the pre-Reagan rates relative to the status quo and is lower by 0.6 percent relative to *laissez-faire*. The allocation of talent under this schedule is shown in the final column of table 4. As expected, the allocation of talent is only slightly different from *laissez-faire* under the pre-Reagan schedule.

C. *Sensitivity to Alternate Assumptions*

Table 5 reports the optimal tax rates under various alternate assumptions, which we now discuss.

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 0.004 and 0.4; our baseline estimate was 0.23. The estimated β changes only slightly to 1.42 and 1.35 relative to the baseline of 1.33. These

TABLE 5
OPTIMAL TAX RATES AND WELFARE GAINS RELATIVE TO LAISSEZ-FAIRE FOR DIFFERENT ASSUMPTIONS (%)

	TAX RATE BRACKET								WELFARE GAIN	
	\$0-\$25k	\$25k-\$50k	\$50k-\$100k	\$100k-\$150k	\$150k-\$200k	\$200k-\$500k	\$500k-\$1m	\$1m+	\$1m+	
Baseline	-3.0	-13.1	-7.6	17.1	33.2	37.0	36.1	39.3		1.3
	A. Elasticities									
Low σ	-51.7	-4.9	64.2	83.6	83.9	63.0	67.2	72.7		2.4
High σ	-2.1	-12.1	-18.6	2.5	23.5	32.6	30.9	34.3		1.6
Low β	-2.5	-14.4	-20.0	2.9	22.8	30.4	29.1	32.8		.7
High β	-3.0	-10.1	3.2	28.3	39.6	42.8	42.7	45.2		2.2
	B. Externalities									
Smaller research externality	-1.4	-2.3	6.5	20.1	25.2	22.8	20.1	24.7		.4
Negative management externality	-2.4	-9.1	3.7	31.2	41.9	46.4	51.0	61.7		2.1
Finance on finance	-2.9	-13.9	-7.5	16.8	30.7	33.4	30.5	30.5		1.0
Engineering on engineering	-2.9	-9.3	-1.3	19.3	33.1	36.3	35.3	38.5		1.2
Research on engineering	-.9	-2.4	5.4	18.2	25.4	22.2	20.4	23.0		.4
$\gamma = .5$	-2.9	-12.1	-5.9	17.8	32.9	36.0	34.8	37.8		1.2
$\gamma = .9$	-3.0	-12.8	-6.9	17.7	33.5	37.0	36.1	39.2		1.3
$\gamma = 1.1$	-3.0	-13.2	-7.4	17.6	33.7	37.5	36.7	39.9		1.3
Negative own externality	-1.8	-6.8	.1	18.7	28.4	28.9	27.9	31.4		.8
	C. Comparative Advantage									
$\rho = .75$	-4.9	-18.3	-15.5	17.4	32.7	25.8	15.5	13.0		.7
	D. Tax Instruments									
Research subsidy (negative own externality)	-12.0	-5.8	.2	15.6	30.0	42.6	24.6	44.5		46.7
Research subsidy ($\gamma = .5$)	22.4	.2	.0	.1	.3	4.3	20.8	71.0		152

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 II.B, 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 80 percent, 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 percent and 3 percent from the baseline of 5 percent. These alternative assumptions lead to values of β of 0.9 and 1.9 versus the baseline value of 1.3. 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.

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 3.0 percent instead of 20.5 percent 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 5. The negative rates end earlier (at \$50,000), and rates on higher earners are lower (between 20 percent and 30 percent). But the basic structure of the tax system stays intact. Changing the management externality from zero to an aggregate negative externality of 8.1 percent of aggregate income makes a much larger difference by making nearly all high-earning professions have negative externalities. Tax rates on income between \$150,000 and \$1 million jump from about 35 percent to about 46 percent, and the tax rate on income over \$1 million rises to 62 percent. This result is analogous to our finding in Section III.B 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.

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 \$1 million falls from 39 percent to 30.5 percent. This decline in rates is consistent with the theoretical results of Rothschild and Scheuer (2014, 2016), who show that 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.

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's (1989) much lower externality estimates, suggesting that even his estimates, correctly interpreted, would lead to quite similar results.

Our baseline analysis assumed that externalities were linear in output by setting the returns to scale parameter γ to one. 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 that each dollar of private product in teaching or research raises the aggregate output of the profession by only 50 cents.²⁶ Optimal rates do diminish, but the effect is slight, with top rates falling from 39 percent to 31 percent. In contrast to the work of Rothschild and Scheuer (2014, 2016), we find that the multiprofession 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.

3. Comparative Advantage

We next consider the impact of allowing comparative advantage, which changes the patterns of substitution across professions. Without 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 Kirkeboen et al. (2016) 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 given by

$$\sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \Pr[i_1(\theta) = i, i_2(\theta) = j] \\ \times \left\{ \mathbb{E} \left[\log \left[\frac{y_i^*(\theta)}{y_j^*(\theta)} \right] \middle| i_1(\theta) = i, i_2(\theta) = j \right] \right. \\ \left. - \mathbb{E} \left[\log \left[\frac{y_i^*(\theta)}{y_j^*(\theta)} \right] \middle| i_1(\theta) = j, i_2(\theta) = i \right] \right\}.$$

²⁶ We choose δ_{ij} for j corresponding to research and teaching so that the relevant diagonal entries in the quasi-Jacobian matrix J defined in Sec. IV.A equal 0.5.

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. (2016).²⁷ When $\rho = 1$, comparative advantage equals zero.

The tax schedule with comparative advantage features declining marginal rates for top incomes, with the rates at \$150,000–\$200,000 similar to before but top tax rates much lower. The new rate on income above \$1 million is 13 percent, and the rate between \$500,000 and \$1 million is 15.5 percent. The negative rates for low earners actually increase to a maximum of 18.3 percent. 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.

4. Tax Instruments

We argued that 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 nonlinear income tax that the government can levy directly on research, which is the profession we estimate produces the strongest externalities.

Under our baseline assumptions these instruments can fail to have an optimum, so we modify the baseline parameters in two ways. First, we choose each $\delta_{i,i}$ so that the externality of teaching and research on themselves equals -0.1 .²⁸

²⁷ 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 nonpecuniary utility ψ significantly blunts comparative advantage given the β we have estimated.

²⁸ When these parameters change but the available tax instrument remains only a nonlinear income tax, the optimal rates stay close to the optimum in the baseline specification, changing to -2.3 percent, -8.0 percent, -1.3 percent, 18.8 percent, 30.9 percent, 31.7 percent, 30.7 percent, and 33.8 percent.

In this case, we find an optimal research tax of -477.4 percent, which would multiply salaries by five 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 $\delta_{i,r}$. Table 5 reports the optimal income tax rates accompanying the optimal subsidy, which change only subtly 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 46.7 percent higher. The subsidy allocates 46.2 percent 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 208 percent, a large number but much smaller than in the previous case, and this subsidy achieves a much larger welfare gain of 152 percent. However, this smaller (if still very large) targeted research subsidy is accompanied by a radical change in the optimal income tax. As reported in table 5, the optimal top tax rate is very high at 71 percent. Rates below the top two are essentially zero, and the rate for the second-highest bracket is lower than in our baseline.

Intuitively, a large targeted subsidy and high marginal tax rates are two methods of inducing greater 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 high marginal tax rates 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.

VI. 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. Optimal taxes in our baseline calibration are not radically different from the US federal income tax schedule.

Our estimates of optimal tax rates depend crucially on several empirical objects whose value is highly uncertain. The first and most important of these are the externalities created by different professions. Our extrapolations from the cross-country regressions of Murphy et al. (1991) to determine the externalities of engineering and law are speculative at best and could be greatly improved by further empirical analysis. 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 that 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 in Card (1999) and Chetty et al.'s (2014) estimates of the external effects of schooling. Similarly, output in engineering could be disaggregated into three components: new product development, where theory suggests that imperfect appropriability creates positive externalities (Spence 1976); operations, where externalities should be limited; and reverse engineering, where negative business-stealing externalities predominate (Hirshleifer 1971).

The second uncertain empirical object is the pattern of labor substitution 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, Neilson, and Zimmerman 2013; Kirkebøen et al. 2016). But many of the professional choices studied in our paper are made conditional on a given undergraduate degree. Neither Hastings et al. nor Kirkebøen et al. 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 exogenous variation in professional wages.

Future research could relax the assumptions of our analysis in two ways. First, Piketty et al. (2014) and Rothschild and Scheuer (2014, 2016) 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 that 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 that 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 avoidance considerations into our model is an interesting direction for future research.

Appendix

Proofs and Derivations

A. Section II

1. 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 $\phi(h) = h^{1/\sigma+1}/(1/\sigma + 1)$, $\phi''(h) = \phi'(h)/(\sigma h)$. 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.

2. Proof of Lemma 1

Note that $h^*(\theta)$ 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 (2), 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 $(T^a)' = (T^b)'$.

3. Calculation of Externality Ratios

We show that the externality ratios solve the system of equations

$$e_j = \sum_{i=1}^n \frac{\partial \log E_i(Y_1^*, \dots, Y_n^*)}{\partial Y_j} \left(a_i + \sum_{k=1}^n b_{i,k} e_k \right), \tag{A1}$$

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, and $b_{i,k}$ 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. Define $\Theta_i = \{\theta | \{i\} \subset I^*(\theta)\}$ to be the set of workers in i (or indifferent) and de-

note $\partial\Theta_i = \{\theta | \{i\} \subsetneq I^*(\theta)\}$ 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 posttax utility of these professions, the change in the pretax 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

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

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

Returning to (A1), note that it takes the form $e = a + Be$, where lowercase letters are n -dimensional column vectors and the uppercase 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.

B. Section III: 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

$$0 = \frac{\sigma}{1 - \tau^*} (\tau^* + s_H e_H + s_L e_L) + \frac{2\beta \tilde{s}_H s_L (r^\alpha - 1)(1 - \tau^*)^\sigma}{\alpha(r + 1)} [\tau^*(r - 1) + r e_H - e_L], \tag{A2}$$

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 that 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 $f_s(y)/f(y)\Delta_r(y)$ 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 $\max_{y, \theta} T(y) - \phi(y/a_i(\theta))$ 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

$$\tilde{s}_L = F^L(-2\beta(1 - \tau)^{1+\sigma}(r - 1)(1 + \sigma)^{-1}(r + 1)^{-1} - \Delta\bar{\psi}),$$

where $\Delta\bar{\psi} = \bar{\psi}_H - \bar{\psi}_L$ and F^L is the CDF of the standard logistic distribution. Note that this share is independent of income.

Conditional on being skilled with productivity a_L , the CDF of ψ_i is $F_\psi(2\beta(1 + r)^{-1} a_L^{-(1+\sigma)} \psi_i - \bar{\psi}_i)$. Therefore, the conditional CDF of $\Delta\psi = \psi_H - \psi_L$ equals $F^L(2\beta(1 + r)^{-1} a_L^{-(1+\sigma)} \Delta\psi - \Delta\bar{\psi})$. The PDF of $\Delta\psi$ equals $2\beta(1 + r)^{-1} a_L^{-(1+\sigma)} f^L$. A standard fact about the logistic distribution is that $f^L = F^L(1 - F^L)$. 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

$$\begin{aligned} \Delta_T(y) &= \int_{y/r}^y \frac{\tau(r - 1)y'}{y} \frac{2\beta(1 - \tau)^\sigma \tilde{s}_H s_L f(y')}{(1 + r)y'f(y)} dy' \\ &= \frac{\tau(r - 1)2\beta(1 - \tau)^\sigma \tilde{s}_H s_L (r^\alpha - 1)}{\alpha(1 + r)}, \end{aligned}$$

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,

$$\begin{aligned} \Delta_e(y) &= \int_{y/r}^y \frac{(re_H - e_L)y'}{y} \frac{2\beta(1 - \tau)^{\sigma} \tilde{s}_H s_L f(y')}{(1 + r)y'f(y)} dy' \\ &= \frac{(re_H - e_L)2\beta(1 - \tau)^{\sigma} \tilde{s}_H s_L (\tau^{\alpha} - 1)}{\alpha(1 + r)}. \end{aligned}$$

Putting these equations together and factoring yields the desired result. QED

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 that τ^* is always in the interval between τ_{int} and τ_{all} because only in this interval will the two terms in (A2) have opposite signs and thus only there can the equation be satisfied. Thus, τ^* must be a convex combination of τ_{int} and τ_{all} , though no simple closed-form solution exists for the relevant weights.

C. Section IV

1. Identification of Externality Coefficients

First, we derive (6). Note $Y_i = H_i E_i(Y_1, \dots, Y_n)$, 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:

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

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 i th entry equals $\partial Y_i / \partial H_j$. Then the above equation can be written in matrix form as

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

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

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

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.

2. Proof of Lemma 2

We begin by proving statements made in the text before the lemma. Consider (7). Conditional on $i^*(\theta) = i$, the worker's maximization is $\max_y y_i - T(y_i) - \phi(y_i a_i^{-1} E_i(Y_0, \dots, Y_n)^{-1})$. The solution satisfies $y_i = [1 - T'(y_i)]^{\sigma} [a_i E_i(Y_0, \dots, Y_n)]^{1+\sigma}$. By using $T = T_{2005}$ and solving for a_i , we immediately obtain (7).

Now we prove (9). 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 (7), as well as the definition for relative utility in the text, we derive (9).

Next, we prove that 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 zero and covariance Σ . 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 $\Sigma_{-1} - \varrho\varrho'$.

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

$$U_i^*(\theta) - \psi_i(\theta) + \psi_i(\theta) = n^{-1} \left(\sum_j a_j^{1+\sigma} \right) \tilde{u}_i(a) + \psi_i(\theta),$$

which is equivalent to maximizing $\tilde{u}_i(a) + \psi_i(\theta)/(n^{-1}\sum_j a_j^{1+\sigma})$. This latter term is distributed as $\beta^{-1}(\tilde{\psi}_i + F_\psi)$, where F_ψ is a standard Gumbel distribution. If we let this Gumbel draw be $\tilde{\psi}_i(\theta)$, the worker is choosing i to maximize $\beta\tilde{u}_i(a) + \tilde{\psi}_i + \psi_i(\theta)$. 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 that the share of workers with productivity a who choose i is

$$\Pr(i^*(\theta) = i | \theta|_a = a) = \frac{e^{\beta\tilde{u}_i(a) + \tilde{\psi}_i}}{\sum_j e^{\beta\tilde{u}_j(a) + \tilde{\psi}_j}}. \tag{A3}$$

To prove (10), 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 $\tilde{f}_i^a(a_i)$, 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 (10). 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 (A3) as the probability of choosing i for each productivity profile. The result is the right side of (10).

References

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

Bagwell, Kyle. 2007. "The Economic Analysis of Advertising." In *Handbook of Industrial Organization*, vol. 3, edited by Mark Armstrong and Robert H. Porter. Amsterdam: North-Holland.

Bai, Jennie, Thomas Philippon, and Alexi Savov. 2012. "Have Financial Markets Become More Informative?" Staff Report no. 578, Fed. Reserve Bank New York.

- . 2016. "Have Financial Markets Become More Informative?" *J. Financial Econ.* 122 (3): 625–54.
- Bakija, Jon, Adam Cole, and Bradley T. Heim. 2012. "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data." Manuscript, Dept. Econ., Williams Coll.
- Barro, Robert J. 1991. "Economic Growth in a Cross-Section of Countries." *Q.J.E.* 106 (2): 407–43.
- Baumol, William J. 1990. "Entrepreneurship: Productive, Unproductive, and Destructive." *J.P.E.* 98 (5): 893–921.
- Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Are CEOs Rewarded for Luck? The Ones without Principals Are." *Q.J.E.* 116 (3): 901–32.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Q.J.E.* 128 (1): 1–51.
- Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, vol. 3, edited by Orley Ashenfelter and David Card, 1801–63. Amsterdam: Elsevier.
- 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, John Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes." *A.E.R.* 104 (9): 2633–79.
- Colombi, Roberto. 1990. "A New Model of Income Distributions: The Pareto Lognormal Distribution." In *Income and Wealth Distribution, Inequality and Poverty*, edited by Camilo Dagum and Michele Zenga, 18–32. Berlin: Springer.
- Congressional Budget Office. 2005. "Effective Marginal Tax Rates on Labor Income." Technical report, Congressional Budget Office, Washington, DC.
- Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *A.E.R.* 88 (1): 83–95.
- French, Kenneth R. 2008. "The Cost of Active Investing." *J. Finance* 63 (4): 1537–73.
- Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Q.J.E.* 123 (1): 49–100.
- Goldin, Claudia, and Lawrence F. Katz. 2008. "Transitions: Career and Family Life Cycles of the Educational Elite." *A.E.R. Papers and Proc.* 98 (2): 363–69.
- Gul, Faruk, and Wolfgang Pesendorfer. 2008. "The Case for Mindless Economics." In *The Foundations of Positive and Normative Economics: A Handbook*, edited by Andrew Caplin and Andrew Schotter, 3–42. New York: Oxford Univ. Press.
- Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman. 2013. "Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile." Working Paper no. 19241, NBER, Cambridge, MA.
- 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." *A.E.R.* 61 (4): 561–74.
- Jaffe, Adam B. 1989. "Real Effects of Academic Research." *A.E.R.* 79 (5): 957–70.
- Jones, Charles I., and John C. Williams. 1998. "Measuring the Social Return to R&D." *Q.J.E.* 113 (4): 1119–35.
- Kaplow, Louis, and Steven Shavell. 1992. "Private versus Socially Optimal Provision of Ex Ante Legal Advice." *J. Law, Econ., and Org.* 8 (2): 306–20.
- Kirkeboen, Lars, Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *Q.J.E.* 131 (3): 1057–1111.

- Malmendier, Ulrike, and Geoffrey Tate. 2009. "Superstar CEOs." *Q.J.E.* 124 (4): 1593–1638.
- Mankiw, N. Gregory. 2010. "Spreading the Wealth Around: Reflections Inspired by Joe the Plumber." *Eastern Econ. J.* 36:285–98.
- Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimal Taxation." *Rev. Econ. Studies* 38 (2): 175–208.
- Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1991. "The Allocation of Talent: Implications for Growth." *Q.J.E.* 106 (2): 503–30.
- Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *J.P.E.* 114 (5): 871–904.
- Patterson, Scott. 2010. *The Quants*. New York: Random House.
- Philippon, Thomas. 2010. "Financiers versus Engineers: Should the Financial Sector Be Taxed or Subsidized?" *American Econ. J.: Microeconomics* 2 (3): 158–82.
- Philippon, Thomas, and Ariell Reshef. 2012. "Wages and Human Capital in the U.S. Finance Industry: 1909–2006." *Q.J.E.* 127 (4): 1551–1609.
- Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Q.J.E.* 118 (1): 1–41.
- Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva. 2014. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities." *American Econ. J.: Econ. Policy* 6 (1): 230–71.
- Posner, Eric, and E. Glen Weyl. 2013. "Benefit-Cost Analysis for Financial Regulation." *A.E.R. Papers and Proc.* 103 (3): 393–97.
- Reyes, Loreto, Jorge Rodríguez, and Sergio S. Urzúa. 2013. "Heterogeneous Economic Returns to Postsecondary Degrees: Evidence from Chile." Working Paper no. 18817, NBER, Cambridge, MA.
- Rothschild, Casey, and Florian Scheuer. 2014. "A Theory of Income Taxation under Multidimensional Skill Heterogeneity." Manuscript, Dept. Econ., Stanford Univ.
- . 2016. "Optimal Taxation with Rent-Seeking." *Rev. Econ. Studies* 83 (1): 1225–62.
- Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Rev. Econ. Studies* 68 (1): 205–29.
- . 2002. "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses." *Q.J.E.* 117 (3): 1039–73.
- Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *J. Econ. Literature* 50 (1): 3–50.
- Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Welfare Weights for Optimal Tax Theory." *A.E.R.* 106 (1): 24–45.
- Saxenian, AnnaLee. 2006. *The New Argonauts*. Cambridge, MA: Harvard Univ. Press.
- Spence, A. Michael. 1976. "Product Selection, Fixed Costs and Monopolistic Competition." *Rev. Econ. Studies* 43 (2): 217–35.
- Vickrey, William. 1945. "Measuring Marginal Utility by Reactions to Risk." *Econometrica* 13 (4): 319–33.
- Weinzierl, Matthew. 2014. "The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice." *J. Public Econ.* 118:128–42.
- Wilson, Robert B. 1993. *Nonlinear Pricing*. Oxford: Oxford Univ. Press.