Optimal Income Taxation with Spillovers from Employer Learning

Ashley C. Craig*

Harvard University

January 2019

JOB MARKET PAPER

Click here for most recent version

Abstract

I study optimal income taxation when human capital investment is imperfectly observable by employers. In my model, Bayesian employer inference about worker productivity drives a wedge between the private and social returns to human capital investment by compressing the wage distribution. The resulting positive externality from worker investment implies lower optimal marginal tax rates, all else being equal. To quantify the significance of this externality for optimal taxation, I calibrate the model to match empirical moments from the United States, including new evidence on how the speed of employer learning about new labor market entrants varies over the worker productivity distribution. Taking into account the spillover from human capital investment introduced by employer inference reduces optimal marginal tax rates by 13 percentage points at around 100,000 dollars of income, with little change in the tails of the income distribution. The welfare gain from this adjustment is equivalent to raising every worker's consumption by one percent.

JEL classification: D62, D82, H21, H23, I21, I24, J24
Keywords: optimal taxation, income taxation, human capital, imperfect information, externalities, statistical discrimination, signaling

*ashleycraig@fas.harvard.edu
I am indebted to my advisors, Roland Fryer, Lawrence Katz, David Laibson and Stefanie Stantcheva. I am also grateful to my informal advisors, mentors and colleagues for valuable comments and suggestions: Lisa Abraham, Ole Agersnap, Alex Bell, Raj Chetty, Oren Danieli, Adriano Fernandes, Edward Glaeser, Benjamin Golub, Nathaniel Hendren, Albert Jan Hummel, Louis Kaplow, Hyunjin Kim, Musab Kurnaz, Jonathan Libgober, Weiling Liu, Benjamin Lockwood, Philip Marx, Gregory Mankiw, David Martin, Namrata Narain, Casey Rothschild, Heather Sarsons and Matthew Weinzierl. Thanks also to participants at several Harvard lunch talks and seminars, the Young Economists Symposium, the NBER Junior Researcher Series and the National Tax Association Annual Conference. Financial support from the Harvard Kennedy School Women and Public Policy Program, and a James M. and Cathleen D. Stone Ph.D Scholarship from the Multidisciplinary Program in Inequality & Social Policy at Harvard University is gratefully acknowledged.

1 Introduction

Employers base hiring and remuneration decisions on imperfect information. When evaluating workers, they rely on noisy correlates of productivity such as references, academic transcripts, and job market papers. Although employers' beliefs about a given worker become more accurate over time, there can be a substantial delay before the worker's wage reflects her marginal product (Farber and Gibbons 1996, Altonji and Pierret 2001, Lange 2007, Kahn and Lange 2014). Until then, employer inference based on imperfect information compresses the wage distribution, which drives a wedge between the present discounted private and social returns to raising one's productivity.

In this way, rational inference by employers introduces a positive externality from human capital investment. Intuitively, a student who studies harder obtains higher future wages by improving her test scores, recommendations, and other indications of her ability. But with imperfect employer information, she also benefits from the hard work of *other* similar students: if her peers were to invest more, employers would tend to look more favorably on her as well.[1] Her peers do not internalize this spillover when choosing how hard to work, and invest less than is socially optimal. This principle applies to learning by any worker while at high school or college, and to investments later in life.

I study the role of income taxation to correct this type of externality. First, I develop a model of optimal taxation with imperfectly observable human capital investment. Next, I show with a simple example how Bayesian inference by employers compresses the wage distribution. This drives a wedge between the private and social returns to investment and lowers the optimal tax rate. Third, I generalize to non-linear taxation, and show that the downward adjustment to marginal tax rates is concentrated at intermediate levels of income. Finally, I calibrate the model to match empirical moments from the United States, introducing new evidence on how employer learning varies over the productivity distribution. Taking into account the spillover introduced by employer inference reduces optimal marginal tax rates by 13 percentage points at around 100,000 dollars of income, with little change in the tails of the income distribution. The welfare gain from this adjustment is equivalent to raising every worker's consumption by one percent.

After observing her investment cost, each worker in my model makes an imperfectly observable investment in human capital, which determines her productivity.[2] Employers

[1] This suggests that an encounter with one worker will affect assessments of other observably similar workers. Sarsons (2018) shows this occurs, although her results are hard to reconcile with full rationality.

[2] Viewed through the lens of the model, obtaining a formal qualification is an imperfect signal of having raised one's productivity. This is a useful approximation of the world to the extent that marginal increases in human capital accumulation require costly effort rather than simply arising from attendance at school.

cannot directly observe the worker's true productivity level. Instead, they infer it based on a noisy but informative signal, combined with a prior belief. As a direct consequence of Bayesian inference by employers, every worker's equilibrium wage is a weighted average of her own productivity and the productivity of other similar workers. An increase in investment by one group of workers therefore has the side effect of altering employers' perceptions of other workers who send similar signals.

Taxation in this model has an effect on welfare that is not present in classic models of income taxation (e.g., Mirrlees 1971). When investment in human capital is depressed by higher taxes and productivity falls, employers become less optimistic, and pay workers a lower wage in equilibrium given the *same information* about their productivity.[3] Individual workers do not take this into account. This is in addition to the usual fiscal externality, which arises because workers ignore the effects of their decisions on government revenue. Since the externality introduced by imperfect employer inference adds to the cost of taxation, taking it into account pushes toward lower marginal tax rates.

The core insights of my model apply more generally. For example, asymmetric employer learning leads to monopsony power for firms, which gives them an incentive to invest in their workers (Acemoglu and Pischke 1998); but imperfect employer information still leads to underinvestment in skills.[4] Similarly, introducing a motive for employers to screen their workers using contracts specifying both labor supply and a wage (Stantcheva 2014) causes utility compression rather than wage compression, but nonetheless undermines the incentive for workers to invest in human capital.

If employers can categorize workers based on exogenous characteristics such as race or gender, my model implies that they will statistically discriminate in any situation in which the equilibrium productivity distribution varies by group.[5] Discrimination may in turn motivate the planner to set group-specific marginal tax rates if differences in the size of the belief externality cause the return to increasing one's productivity to differ across groups: for example, there is some evidence to suggest a lower return to skill for black workers than white workers (Bertrand and Mullainathan 2004, Pinkston 2006).

Using a simple example with linear taxation, I demonstrate how rational employer inference based on imperfect signals causes compression of workers' wages toward the average level of productivity. This flattens the relationship between productivity and re-

[3]In Section 4, I show how investment can also *hurt* others in some cases, although not in simple examples.

[4]Asymmetric firm information may also indirectly undermine worker investment by affecting firms' incentives to promote workers (Milgrom and Oster 1987).

[5]My model is a generalization of classic models of statistical discrimination. Pioneered by Phelps (1972) and Arrow (1973), such models rely on imperfect observability of productivity to explain employers' use of a worker's group identity. Contributions include Aigner and Cain (1977), Coate and Loury (1993), Moro and Norman (2004), Lang and Manove (2006), and Fryer (2007). See Fang and Moro (2011) for a review.

muneration, introducing a wedge between the private and social returns to investment. Relative to a model with perfect employer information, the optimal tax is therefore lower; this correction is larger if employers have less precise information about their workers' productivity, or if productivity is more responsive to taxation. In the special case in which all agents receive equal social welfare weight, the optimal tax is always negative, reflecting only the efficiency motive for intervention.

When I generalize to non-linear taxation, imperfect employer information introduces a novel effect of a change to the tax schedule, which I refer to as the *belief externality*: every worker who changes her investment decision also shifts employers' beliefs, which in turn affects the wages and welfare of others. Less accurate employer information makes this externality larger, and pushes toward lower taxes. This is in addition to the two classic effects of income taxation: the *mechanical effect* from the transfer of consumption from high income workers to low income workers; and the *fiscal externality*, which arises because individuals ignore the impact on government revenue of re-optimization of their human capital investment and labor supply decisions.

The welfare impact of the belief externality is greatest at intermediate incomes, which contributes to a "U" shape of the optimal marginal tax schedule.[6] There are two steps to understand this result. First, a given spillover in wages has a larger effect on consumption for higher-income workers, because they supply more labor. Second, as incomes rise even further, social welfare weights decline toward zero. In turn, this means that a given change in consumption has little effect on social welfare at the highest incomes.

My results also highlight how the belief externality can be decomposed into two components of opposite sign and different incidence. When a worker invests more, her higher productivity raises the wages of workers who send signals most similar to her own. However, she hurts workers whose signal distributions are concentrated in regions where her own distribution *changes* the most. The reason for this negative effect is that she becomes more likely to send high signals where her productivity lowers the average, and less likely to send low signals where she had previously raised the average.

To quantify the importance of the belief externality, I calibrate my model to match the United States wage and income distributions, evidence on the gap between the private and social returns to productivity, and estimates of the elasticities of wages and labor supply (e.g., Blomquist and Selin 2010). I calibrate the belief externality in two steps.

[6] The standard trade-off between equality and efficiency already produces a "U" shape, given the shape of the income distribution typically estimated (Diamond 1998). Raising the marginal tax rate at a given income transfers resources from those above to those below that level, but distorts decisions locally. The "U" shape arises because the efficiency cost increases at low incomes as the density of income rises; it then decreases at high incomes, as the density falls. This shape is amplified by the forces in my model.

First, I infer its overall size from existing estimates of the speed of employer learning (Lange 2007, Kahn and Lange 2014). Second, I use data from the National Longitudinal Survey of Youth (NLSY79) to show that there is stronger evidence of learning among low-productivity than high-productivity workers. As a proxy for worker productivity in my empirical work, I use scores on the Armed Forces Qualification Test (AFQT) from before each worker entered the labor market.

Taking into account the belief externality significantly reduces optimal marginal tax rates for most workers. Moreover, the welfare gain from adopting the optimal tax schedule is notable – equivalent to increasing every worker's consumption by around one percent. As predicted by my theoretical results, the downward adjustment to taxes is concentrated at moderate-to-high levels of income, with little change to the marginal tax rates faced by workers with the lowest and highest incomes.

Finally, I extend the model to allow inherently more able workers to have systematically higher or lower investment costs à la Spence (1973). Investment then serves two roles: increasing human capital, and revealing innate ability. Residual uncertainty about a worker's productivity continues to produce wage compression due to employer belief formation, but there is also a signaling component of the private return that may be either positive or negative. In the extreme case in which productivity is entirely innate, the net externality from investment is negative. More generally, however, less accurate employer information reinforces the positive component of the externality that arises from wage compression, but dampens the signaling externality.

Connections in the Income Taxation Literature

This paper builds on a rich literature studying optimal income taxation, the modern analysis of which began with Mirrlees (1971). In these models, a social planner seeks to redistribute resources from high skill to low skill workers. A trade-off between equity and efficiency arises because workers' skill levels are not directly observable by the planner. Redistribution must therefore occur via a tax on earnings, which distorts labor supply choices. Subsequent work (Diamond 1998, Saez 2001) has enriched our understanding of Mirrlees' original results, and has extended them to incorporate extensive margin labor supply responses (Saez 2002), lifecycle concerns (Albanesi and Sleet 2006, Farhi and Werning 2013, Golosov, Troshkin and Tsyvinski 2016), rent-seeking effects (Piketty, Saez and Stantcheva 2014), occupational choice (Gomes, Lozachmeur and Pavan 2018), and migration (Simula and Trannoy 2010, Lehmann, Simula and Trannoy 2014).[7]

My model most closely relates to the strand of this literature in which workers' skills are attained through investment in human capital (Bovenberg and Jacobs 2005, Jacobs

[7] See Piketty and Saez (2013) for a review of the literature on optimal labor income taxation.

2005, Jacobs 2007, Boháček and Kapička 2008, Maldonado 2008, Kapička 2015), which includes models with risky human capital (da Costa and Maestri 2007, Stantcheva 2017, Findeisen and Sachs 2016), overlapping generations (Krueger and Ludwig 2016) and ongoing learning (Best and Kleven 2013, Makris and Pavan 2017). In fact, when employer information becomes arbitrarily accurate in my model, it becomes isomorphic to the version of Bovenberg and Jacobs (2005) in which the social planner cannot observe human capital. Away from this limit, the externality introduced in my model by imperfect employer information suggests that marginal taxes should be lower, all else being equal.

There has been less attention paid to taxation with human capital investment that is imperfectly observable by employers. Where this has occurred, it has been limited to the case of purely unproductive signaling. For example, Andersson (1996) analyzes taxation in a two-type pure signaling model. Similarly, Spence (1974) discusses the case of pure signaling with perfectly inelastic labor supply. In related work, Stantcheva (2014) analyzes the two-level screening problem that arises when labor disutility is directly related to a worker's productivity, so that willingness to work long hours signals high ability. However, productivity is immutable in her model. Most similar in spirit, Hedlund (2018) analyzes bequest taxation in a model that features a similar belief externality to mine, but in which investment is binary and there is no redistributive motive for taxation.

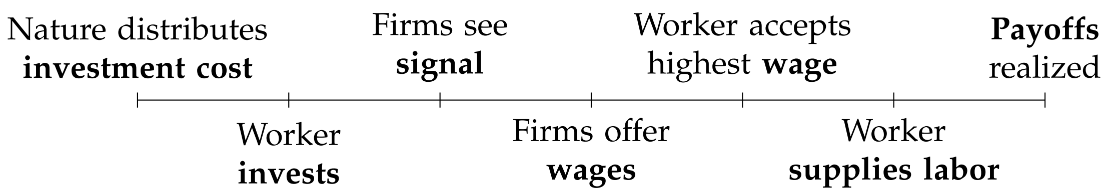
The paper also connects to the literature on optimal income taxation with general equilibrium externalities (e.g., Stiglitz 1982, Rothschild and Scheuer 2013, Rothschild and Scheuer 2016, Sachs, Tsyvinski and Werquin 2016, Lockwood, Nathanson and Weyl 2017), and to the broader literature on human capital externalities in production (Moretti 2004, Kline and Moretti 2014). However, the aggregate production function remains linear in my model. More importantly, the belief externality that is the focus of this paper tends to have local incidence: rather than complementarities in production between dissimilar types, the spillovers here arise because each worker's investment decision changes perceptions by employers about others who are observably similar. This distinction is important in determining the shape of the optimal tax schedule.

2 A Model of Optimal Taxation with Employer Learning

A. Building Blocks

Let there be a fixed tax schedule T. This induces a game between a single worker and several identical firms, indexed by $j \in J$ with $|J| \geq 2$. The timeline is shown in Figure 1. Nature first distributes a cost of investment $k \in K \subseteq \mathbb{R}_{++}$ to the worker, with cumulative distribution $G(k)$. After observing k, the worker invests $x \in \mathbb{R}_+$ at utility cost kx, yielding productivity $q = Q(x)$ where $Q'(x) > 0$, $Q''(x) < 0$, $Q(0) = 0$ and $\lim_{x \to 0} Q'(x) = \infty$.[8]

FIGURE 1: TIMELINE OF THE GAME



Nature then distributes a signal of productivity to the worker and all firms. Specifically, let $\theta \in \Theta \subseteq \mathbb{R}_+$ be a non-contractible signal with conditional density $f(\theta|q)$, which is twice continuously differentiable in q, and has full support for every q. Define $\overline{\theta} = \sup(\Theta)$ and let $f(\theta)$ be the marginal distribution of θ. Assume that $f(\theta|q)$ is differentiable with respect to θ, and that it satisfies the monotone likelihood ratio property: i.e., $\frac{\partial}{\partial \theta}\left(\frac{f(\theta|q_H)}{f(\theta|q_L)}\right) > 0$ for all $q_H > q_L$. Based on θ and a prior $\pi(q)$, each firm forms a posterior belief about the worker's productivity.

Next, each of $|J| \geq 2$ firms simultaneously offers a wage $w_j \in \mathbb{R}_+$ to the worker.[9] The worker accepts her preferred offer, choosing firm $j \in J$, and supplies labor $l \in \mathbb{R}_+$. Of her pre-tax income z, the worker consumes $c = z - T(z)$ where the function $T \in \mathcal{T} \subseteq \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ is the exogenous tax system set by the social planner.[10]

B. WORKER AND FIRM PAYOFFS

The worker receives utility $u(z - T(z), l) - kx$, where: $u_c > 0$, $u_l < 0$, $u_{cc} \leq 0$ and $u_{ll} < 0$. I further assume that u_c is finite for all $c > 0$ and that $\lim_{l \to \infty} u_l = -\infty$ and $\lim_{l \to 0} u_l = 0$. Firms are risk neutral and obtain benefit q per unit of supplied labor.

C. WORKER AND FIRM STRATEGIES

I focus on pure strategy equilibria. The worker's strategy is a set of three functions – an investment decision, an acceptance rule and a labor supply decision. These can be written as: $x : K \times \mathcal{T} \to \mathbb{R}_+$; $A : K \times \mathcal{T} \times \Theta \times \mathbb{R}_+^{|J|} \to J$; and $L : K \times \mathcal{T} \times \Theta \times \mathbb{R}_+^{|J|} \to \mathbb{R}_+$. Each employer's strategy maps signals and tax systems to wage offers $O_j : \Theta \times \mathcal{T} \to \mathbb{R}_+$.

[8] The existence of a utility cost of investment is supported by the findings of Heckman, Lochner and Todd (2006b) and Heckman, Stixrud and Urzua (2006a), and is appropriate to model unobservable investment.

[9] Appendix A shows that this is isomorphic to a model with contracts specifying labor supply and income, because the marginal rate of substitution between labor supply and consumption is independent of labor *quality*. If this is relaxed, employers may be able to use contracts to screen workers (Stantcheva 2014).

[10] I use $\mathcal{C}(A, B)$ to denote the space of continuous functions mapping from A to B.

D. EQUILIBRIUM

An equilibrium of the game induced by a given tax schedule is a Perfect Bayesian Equilibrium (PBE). This requires that firms' beliefs are rationally formed using Bayes rule whenever it applies, and that all strategies satisfy sequential rationality.

E. OPTIMAL FIRM AND WORKER BEHAVIOR

Each firm chooses the wage w_j to maximize its expected profit $\overline{P}_j$. Letting $\Pr(A_j = 1|w_j)$ be the probability that the worker accepts firm j's offer, its expected profit is:

$$\overline{P}_{j,\theta} = E[P_j|\theta, \pi, w_j] = \Pr(A_j = 1|w_j) \times (E[q|\theta, \pi, A_j = 1] - w_j) \times l$$

where l is the quantity of labor supplied by the worker.

Given the assumptions above, every firm earns zero expected profit, and each worker receives a wage $w(\theta|\pi)$ equal to her expected marginal product $E[q|\theta, \pi]$ given the signal θ and the equilibrium distribution of productivity.

Lemma 1. Fix a value of θ and assume $E[q|\theta, \pi]$ is strictly positive and finite given beliefs $\pi(q)$. In any pure-strategy equilibrium, all firms $j \in J$ earn zero expected profit, and the wage offered to each worker by each firm is her expected marginal product $E[q|\theta, \pi]$.

All technical proofs are presented in Appendix G.

After accepting a wage offer, the worker supplies labor $l(\theta|\pi, T)$ as follows.[11]

$$l(\theta|\pi, T) \in L^* = \underset{\tilde{l}_j \in \mathbb{R}_+}{\operatorname{argmax}} \ u\left(w(\theta|\pi)\tilde{l} - T\left(w(\theta|\pi)\tilde{l}\right), \tilde{l}\right) \tag{1}$$

In turn, this implies that her income is $z(\theta|\pi, T) = w(\theta|\pi)\, l(\theta|\pi, T)$. Knowing this, the worker can calculate her expected utility $v(\theta|\pi, T)$ for any signal realization.

$$v(\theta|\pi, T) = u\left(z(\theta|\pi, T) - T(z(\theta|\pi, T)), \left(\frac{z(\theta|\pi, T)}{w(\theta|\pi)}\right)\right) \tag{2}$$

Evaluating the expectation of $v(\theta|\pi, T)$ by integrating over θ, investing x leads to expected utility $V(Q(x)|\pi, T) = E_\theta[v(\theta|\pi, T)|Q(x)] - kx$. At the investment stage, a worker with cost k takes the function $V(q|\pi, T)$ as given, and optimally invests $x(k|\pi, T)$, which solves problem 3. This yields productivity $q(k|\pi, T)$.

$$x(k|\pi, T) \in X^* = \underset{\tilde{x} \in \mathbb{R}_+}{\operatorname{argmax}} \int_\Theta v(\theta|\pi, T)\, f(\theta|Q(\tilde{x}))\, d\theta - k\tilde{x} \tag{3}$$

[11] Throughout the paper, optimal choices of labor supply and investment will be unique.

In turn, these investment decisions collectively suffice to characterize the expected marginal product, and thus the wage, of an individual with signal realization θ.

$$w(\theta|\pi) = \frac{\int_K q(k|\pi, T) f(\theta|q(k|\pi, T)) \, dG(k)}{\int_K f(\theta|q(k|\pi, T)) \, dG(k)} \tag{4}$$

The monotone likelihood ratio property ensures that the equilibrium wage is strictly increasing in θ, and that $V(q|\pi, T)$ increases with q.

F. Characterizing and Selecting Equilibria

Equations 1, 3 and 4 comprise a fixed point at which worker investment decisions and employer beliefs are consistent. Each employer has a correct prior belief $\pi(q)$, and rationally updates it upon observing a signal. Perfect competition ensures that every firm offers the worker a wage equal to her expected marginal product. Combined with the signal distribution, this wage schedule then pins down the worker's expected utility at each productivity level. Finally, the worker's choices of productivity levels induce a productivity distribution that must coincide with every employer's prior belief in equilibrium.

For any tax schedule T, there is a set of equilibria $\boldsymbol{E}(T)$. I consider a selection of these equilibria, defined by choosing one equilibrium $E^{\dagger}(T) \in \boldsymbol{E}(T)$ for each T.[12] The expected utility of a worker with investment cost k is then defined as her expected utility given the tax schedule and this selection: $\overline{V}(k, T) = V\left(k, E^{\dagger}(T), T\right)$. For example, one possibility is to assume that agents always coordinate on one of the social planner's preferred equilibria. I assume that this is the case when presenting the results for non-linear taxation in Section 4. However, my approach is equally valid for other selections.

Note 1. The game here is described as one between a single worker and a set of firms, with the worker's type k drawn from $G(k)$. An alternative interpretation is that there is a continuum of workers whose investment costs have distribution $G(k)$ in the population. I adopt this terminology throughout much of the paper.

G. The Social Planner

Having established the nature of an equilibrium, I now introduce the social planner. The role of this planner is to choose a tax schedule T to maximize social welfare $\boldsymbol{\mathcal{W}}(T)$, which

[12] An alternative is simply to assume that the initial equilibrium is *stable*. This ensures that the economy does not switch equilibria in response to a small change in T. I define stability in Appendix B.

is defined as the average across types of the worker's expected utility levels after they have been transformed by a social welfare function W.[13]

$$\max_{T\in\mathcal{T}} \boldsymbol{\mathcal{W}}\left(T\right) = \int_K W\left(\overline{V}\left(k,T\right)\right) dG\left(k\right)$$

The social welfare function W is assumed to be increasing, concave and differentiable.

The choice of T must satisfy two constraints. First, it can be a direct function only of realized income z. Second, the planner must raise enough tax revenue to cover an exogenously fixed revenue requirement, R. In some examples, I further restrict $T\left(z\right)$ to be linear in z (Section 3, Appendix C and Appendix D).

The planner's problem can be written as a choice of a tax system to maximize welfare, subject to the resource constraint, individual optimization and rational belief formation.

$$\max_{T\in\mathcal{T}} \boldsymbol{\mathcal{W}}\left(T\right) = W\left(\overline{V}\left(k,T\right)\right) dG\left(k\right) \tag{5}$$

where:

$$\overline{V}\left(k,T\right) = \int_\Theta \left(v\left(\theta|\pi,T\right) - kx\left(k,\pi,T\right)\right) f\left(\theta, q\left(k|\pi,T\right)\right) d\theta$$

subject to:

$$x\left(k|\pi,T\right) \in \underset{\tilde{x}\in\mathbb{R}_+}{\text{argmax}} \int_\Theta v\left(\theta|\pi,T\right) f\left(\theta|Q\left(\tilde{x}\right)\right) d\theta - k\tilde{x} \tag{6}$$

$$l\left(\theta|\pi,T\right) \in \underset{\tilde{l}\in\mathbb{R}_+}{\text{argmax}}\ u\left(w\left(\theta|\pi\right)\tilde{l} - T\left(w\left(\theta|\pi\right)\tilde{l}\right), \tilde{l}\right) \tag{7}$$

$$w\left(\theta|\pi\right) = \frac{\int_K q\left(k|\pi,T\right) f\left(\theta|q\left(k|\pi,T\right)\right) dG\left(k\right)}{\int_K f\left(\theta|q\left(k|\pi,T\right)\right) dG\left(k\right)} \tag{8}$$

$$R = \int_\Theta T\left(z\left(\theta|\pi,T\right)\right) f\left(\theta\right) d\theta \tag{9}$$

In summary, the planner's choice of a tax system T alters the set of equilibria in the economy. Given a selection from this equilibrium correspondence – for example, the planner's preferred equilibrium for each tax schedule – the planner maximizes welfare. Changes in the tax schedule shift the worker's incentives to invest and her willingness to supply labor. Due to imperfect employer information, the worker's investment decisions also affect equilibrium wages – an effect she ignores when she invests.[14]

[13] I omit profits from social welfare because they are zero in expectation.

[14] This can be thought of as a problem with inner and outer components à la Rothschild and Scheuer (2013), with rational belief formation serving as the consistency constraint. A difference is that Rothschild and Scheuer (2013) re-write the social planner's problem as a direct choice over allocations.

3 A Simple Example with Linear Taxation

I begin with an example in which the planner is restricted to choosing a linear tax, τ.[15] Each worker's consumption is then an average of her own income, z, and the mean income, $\overline{z}$: $c = (1-\tau)\,z + \tau\overline{z}$. This example highlights an effect that is not present in classic models of income taxation: when productivity falls in response to higher taxes, employers become less optimistic, and pay workers lower wages given the same information about their productivity. Taking this effect into account leads to a lower optimal tax.

A. Additional Assumptions

For convenience, I assume that workers have quasilinear isoelastic utility and that the production function for investment is also isoelastic.[16]

$$u = c - l^{1+\frac{1}{\varepsilon_l}} \Big/ \left(1 + \frac{1}{\varepsilon_l}\right) \qquad\qquad q = x^{\beta}$$

I also make assumptions about the cost and signal distributions, which jointly yield a tractable signal extraction problem for employers. First, the relationship between the signal θ and productivity q is: $\ln\theta = \ln q + \ln\xi$, where $\ln\xi \sim N(0, \sigma_\xi^2)$. Secondly, investment costs k are distributed log-normally: $k \sim LN(\ln\mu_k - \sigma_k^2/2, \sigma_k^2)$.

B. Equilibrium

Given any linear tax rate τ, there is an equilibrium in which productivity and income are both log-normally distributed; and in this equilibrium, a worker's wage is a weighted geometric average of her own productivity q, average productivity μ_q, and idiosyncratic noise. The weight on a worker's own productivity is the fraction of the variance of the signal that arises due to variation in productivity rather than noise, $s = \sigma_q^2/(\sigma_q^2 + \sigma_\xi^2)$. Intuitively, the signal is only useful to employers to the extent that variation in it reflects differences in productivity rather than noise.

Proposition 1. *For any fixed tax rate τ, there exists an equilibrium in which productivity and income are both log-normally distributed.*

$$\ln q \sim N\left(\ln\mu_q - \frac{\sigma_q^2}{2}, \sigma_q^2\right)$$

A worker's wage is $w = q^s \mu_q^{1-s} \xi^s$ where $s = \frac{\sigma_q^2}{\sigma_q^2 + \sigma_\xi^2} \in (0,1)$.

[15] For simplicity, I also assume that the government's revenue requirement, R, is zero.

[16] I assume $\beta\,(1+\varepsilon_l) < 1$ so that investment returns are concave and choices finite (see Appendix G).

The weight on a worker's own productivity is a measure of the wedge between the private and social returns to investment. If a worker of a given cost type were to unilaterally increase her productivity by one percent, her expected wage would increase by $s < 1$ percent. If most of the variance in the signal θ comes from noise (σ_ξ^2 large), s is close to zero, and employers place little weight on the signal when setting a worker's wage. There is then little private return to investment. Alternatively, if σ_ξ^2 is small, then s is close to one, and the private return to investment is close to the social return.

The simplicity of this example stems from the fact that the elasticities of investment and income with respect the retention rate, $1-\tau$, are constant and independent of s. This may seem surprising, since more noise (lower s) flattens the relationship between a worker's log productivity and her log wage. However, there is a second effect of lower s: employers place more weight on *average* productivity, which strengthens a social multiplier in the model. In response to a fall in τ, workers invest more, and μ_q rises; this further increases investment returns, and amplifies the response of productivity to taxation. These two effects of noise cancel out in this example, leaving the elasticities unaffected.

Lemma 2. Assume that the log-normal equilibrium from Proposition 1 is played. The elasticities of productivity (ε_q) and income (ε_z) with respect to the retention rate $1-\tau$ are:

$$\varepsilon_q = \frac{\beta\left(1+\varepsilon_l\right)}{1-\beta\left(1+\varepsilon_l\right)} \qquad \varepsilon_z = \frac{\varepsilon_l + \beta\left(1+\varepsilon_l\right)}{1-\beta\left(1+\varepsilon_l\right)}$$

C. Optimal Taxation

Building on Lemma 2, Proposition 2 provides a formula for the optimal linear tax. First, let $\psi_k = W'\left(\overline{V}\left(k,\tau\right)\right)$ be the marginal social welfare weight placed on an individual with cost k, and let $\overline{\psi}$ be the average welfare weight. Similarly, define $\overline{z}_k$ as the average income for individuals with cost k, and let $\overline{z}$ be the population average income. The first-order condition for the optimal tax is given by equation 10.

Proposition 2. *Assume that the log-normal equilibrium described in Proposition 1 is played. The first-order condition for the optimal linear tax τ^* is:*

$$\frac{\tau^*}{1-\tau^*} = \underbrace{\frac{1-\gamma}{\varepsilon_z}}_{\textit{Standard}} - \underbrace{\frac{\gamma\left(1-s\right)\varepsilon_q}{\varepsilon_z}}_{\textit{New}} \tag{10}$$

where: $\gamma = E_k\left\{\frac{\psi_k}{\overline{\psi}}\frac{\overline{z}_k}{\overline{z}}\right\} \geq 0.$

Equation 10 is similar to the optimal tax formula in the standard case with perfect employer information. Indeed, the first term captures the usual trade-off between redistribution and distortion. The second term is new, and captures the intuition that workers who become more productive impose a positive externality on others by making employers more optimistic, and raising the wage paid for any given signal realization.

The formula in Proposition 2 can be derived by combining the three effects of slightly raising the tax rate. First, there is a mechanical effect (**ME**), which is the welfare gain from taking money from high-income individuals and redistributing it.

$$\mathbf{ME} = \overline{\psi}\overline{z} - E_k\left(\overline{\psi}_k \overline{z}_k\right)$$

This transfer raises social welfare to the extent that workers with high income have lower-than-average welfare weight: $E_k\left(\overline{\psi}_k \overline{z}_k\right) < \overline{\psi}\overline{z}$. If welfare weights decline rapidly with income, γ is close to zero and τ^* is high. Conversely, a social planner with only a weak preference for redistribution has γ close to one, which implies a low value of τ^*.

The second traditional effect of taxation is the fiscal externality (**FE**), which captures the impact of changes in labor supply and investment decisions on the government budget.

$$\mathbf{FE} = -\tau\overline{\psi}\varepsilon_z \frac{\overline{z}}{1-\tau}$$

When workers re-optimize in response to a change in τ, the effect of this on their own welfare is second-order (by the envelope theorem). However, there is a first-order effect on government revenue, which is returned to workers. In classic income taxation models, the fiscal externality is a sufficient statistic for the cost of taxation (Feldstein 1999).

With imperfect employer information, there is a new effect which I refer to as the *belief externality* (**BE**). When workers increase their investment, they do not take into account their effects on employer beliefs, which translate into changes in the equilibrium wage paid for each signal realization. This constitutes a second externality.

$$\mathbf{BE} = -E_k\left(\psi_k \overline{z}_k\right)(1-s)\,\varepsilon_q$$

The new effect pushes toward a lower tax rate. Its magnitude depends on three factors. First, it rises with the size of the wedge between private and social returns, $1-s$. Second, it scales with ε_q, because the externality arises from workers becoming more productive. Third, the welfare impact scales with $E_k\left(\psi_k \overline{z}_k\right)$, because higher income individuals – who supply more labor – are more affected by a given change in their wage.

D. Graphical Demonstration

The effects of a small reduction in the linear tax rate are shown in Figure 2. Panel (a) shows the change in wages at each level of productivity, and decomposes it into the direct effect from a worker's own re-optimization and the indirect effect via employer beliefs. I assume that $s = 0.75$, which is a value that aligns with the evidence on employer learning (see Section 5). This implies that 25 percent of the change in the average wage is not internalized by the workers who respond to the tax change.

Panel (b) shows the utility impacts of the mechanical effect, fiscal externality and belief externality. The effects are weighted by the density of productivity so that the area between each curve and the zero line is the average utility impact. Since the tax rate has been reduced, there is a mechanical transfer of utility from low productivity to high productivity workers. Second, there is a positive fiscal externality, as incomes rise and the government collects more revenue. Finally, there is a positive belief externality, as employers become more optimistic and pay workers a higher wage given any signal realization.

E. Special Cases

It is instructive to consider three special cases of the optimal tax formula. First, if employers perfectly observe productivity ($s = 1$), equation 10 collapses to the standard case.

$$\left.\frac{\tau^*}{1-\tau^*}\right|_{s=1} = \frac{1-\gamma}{\varepsilon_z} \tag{11}$$

While it is critical that ε_z incorporates the long-run response of human capital in any calibration, this equation is otherwise the same as that which arises in a model with fixed productivity types and perfect employer information.[17]

In general, however, there is an efficiency motive to intervene, and this is reflected by the formula that arises when the planner has no redistributive motive (i.e., $\psi_k = 1 \forall k$).

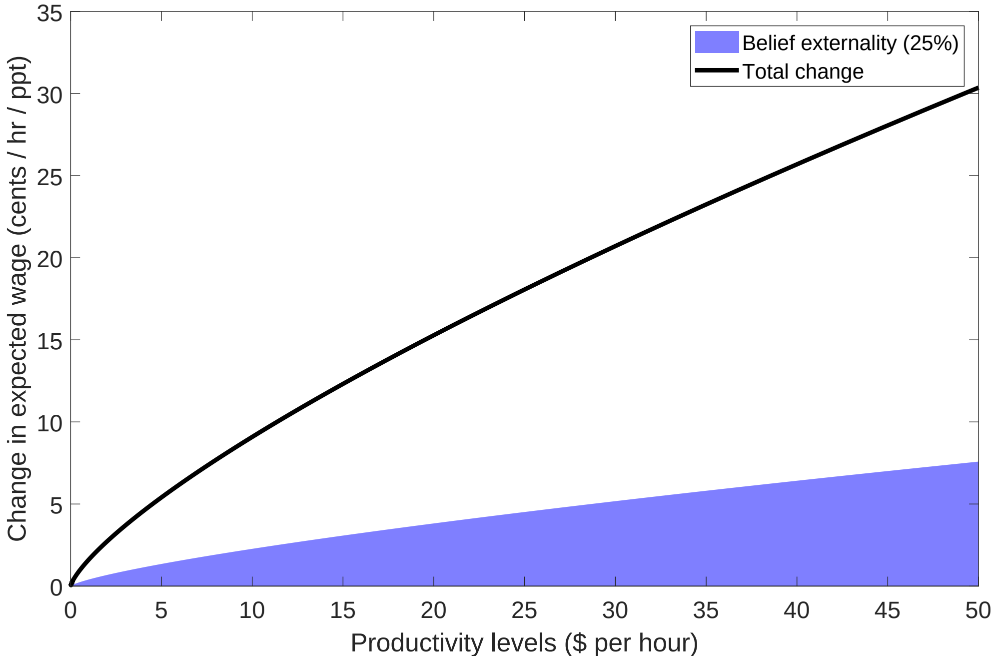
$$\left.\frac{\tau^*}{1-\tau^*}\right|_{\psi_k=1\forall k} = -\frac{(1-s)\,\varepsilon_q}{\varepsilon_z}$$

In this case, the planner simply aims to align private and social returns. This can be contrasted with a Rawlsian social planner, who cares only about the highest-cost worker. The Rawlsian tax rate, $\left.\frac{\tau}{1-\tau}\right|_r = \frac{1}{\sigma_z}$, maximizes government revenue, and is unchanged from the standard case because the highest-cost worker is unaffected by the belief externality.

[17] In this case, both the condition for optimal taxes and the elasticity ε_z are identical to the results of Bovenberg and Jacobs (2005) under the assumption that the planner cannot directly subsidize investment.

FIGURE 2: CHANGES IN WAGES AND UTILITY IN RESPONSE TO LOWER τ

(a) Wage Impact



(b) Utility Impact

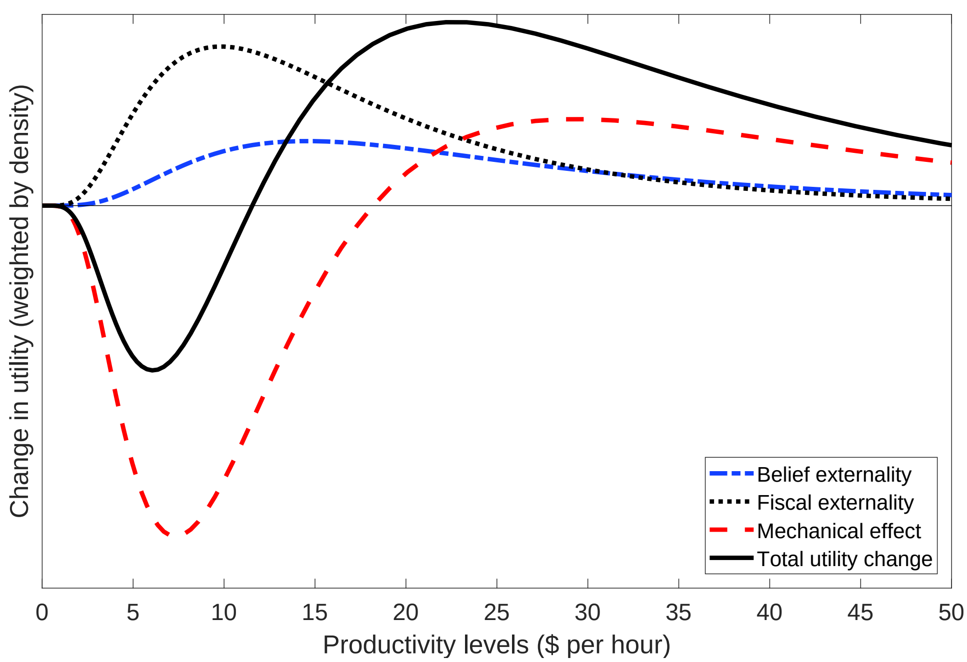


Figure notes. These figures show the effects of a reduction in τ on wages and utility at each productivity level in the linear taxation example, calibrated to achieve $s = 0.75$ in equilibrium and match the United States wage distribution (see Appendix H for details). The utility impacts in panel (b) are scaled by the productivity density so that the area under each curve is proportional to the average impact.

F. TAGGING AND STATISTICAL DISCRIMINATION

If employers can observe exogenous characteristics about a worker such as race, gender or disability status, this model implies that they will statistically discriminate in any situation in which groups of workers differ in their equilibrium productivity distributions. The logic is simple: if the productivity distributions differ, then employers rationally make different assessments of a worker's productivity given the same signal.

For example, suppose there is an advantaged group A and a disadvantaged group D, which are identical except that group D's costs are proportionally higher than group A's ($\mu_k^D > \mu_k^A$). This implies that the equilibrium wage and income distributions of D workers are shifted down relative to those of A workers. An audit study in this economy would reveal a positive wage gap between A and D workers with identical signals.

$$\ln\left(\frac{w\left(\theta|\pi_A\right)}{w\left(\theta|\pi_D\right)}\right) = (1-s)\ln\left(\frac{\mu_k^A}{\mu_k^D}\right) \tag{12}$$

Specifically, with $s = 0.75$, discrimination would appear to "account for" around one quarter of the overall wage gap between the two groups.

This raises the question of whether discrimination motivates the planner to set different tax rates for each group. There are two traditional reasons for doing so. First, the elasticity of taxable income, ε_z, may differ between groups; and second, the covariance between incomes and welfare weights, γ, may differ.[18] In this model, there is an additional tagging motive: the size of the belief externality may differ across groups.

A key result is that the size of the externality depends on the dispersion but not the level of investment. As a result, a cost disadvantage as in the example above does not affect the externality, and does not provide a motive to differentiate between groups. In this sense, statistical discrimination does not necessarily imply that a group-specific subsidy is optimal. This contrasts with subsidy programs suggested based on classic models of purely "self-fulfilling" statistical discrimination (Coate and Loury 1993).[19]

Corollary 1. The standard deviation of log wages (σ_q^2) and signal-to-noise ratio $s = \frac{\sigma_q^2}{\sigma_q^2+\sigma_\xi^2}$

[18]For discussions of "tagging", see Akerlof (1978), Kaplow (2007), and Mankiw and Weinzierl (2010). See also Fryer and Loury (2013), who analyze policies designed to improve the outcomes a disadvantaged group in a job assignment model with endogenous investment in skills.

[19]In simple models of self-fulfilling statistical discrimination, investment is binary and the social benefit is constant. Any equilibrium with less investment must then feature a larger externality. The question these papers ask is also different: they focus on equilibrium *selection*, which I discuss in Appendix D.

are pinned down by the following condition.

$$\sigma_q^2 = \left(\frac{\beta}{\beta s\left(1+\varepsilon_l\right)}\right)^2 \sigma_k^2$$

The optimal tax rate is thus independent of the level of costs μ_k, and does not directly depend on the level of log productivity μ_q.

The underlying reason for statistical discrimination is critical, however. For example, suppose that the cost distributions of groups A and D are identical but that the signal is noisier for group D; this has been posited as a reason for statistical discrimination (see Phelps 1972, Aigner and Cain 1977, Lundberg and Startz 1983). In this case, the belief externality is larger for the disadvantaged group, and s is lower. The social planner therefore has a motive to set $\tau_D < \tau_A$. Alternatively, statistical discrimination may be self-fulfilling in the sense of Arrow (1973)[20], which is an issue that I take up in Appendix D.

If the reason for a disparity is unknown, the right question to ask from the point of view of the social planner is whether the return to increasing one's productivity differs across groups – i.e., whether the belief externality is larger. For example, there is some evidence that the return to skill is either lower for black workers than white workers (Bertrand and Mullainathan 2004, Pinkston 2006), or roughly equal (Neal and Johnson 1996). Combined with assumptions about group differences in ε_z, ε_q and γ, such evidence provides guidance as to whether group-specific tax rates are theoretically warranted.

4 Non-linear Taxation

A. Perturbation of the Tax Schedule

I now relax the restrictive assumptions of Section 3, and derive a necessary condition for optimal non-linear taxation by studying a small perturbation to the tax schedule. Specifically, I consider raising the marginal tax by $d\tau$ over a small range of incomes between z and $z + dz$, where $d\tau$ is second-order compared to dz.[21] This is accompanied by a change in the intercept of the tax schedule – a uniform increase in the consumption of all workers – to ensure that the resource constraint still holds with equality.

[20] See also Coate and Loury (1993), Moro and Norman (2004), and Craig and Fryer (2017).

[21] Requiring $d\tau$ to be small abstracts from bunching and gaps from introducing a kink in the tax schedule.

FIGURE 3: A LOCAL PERTURBATION TO T

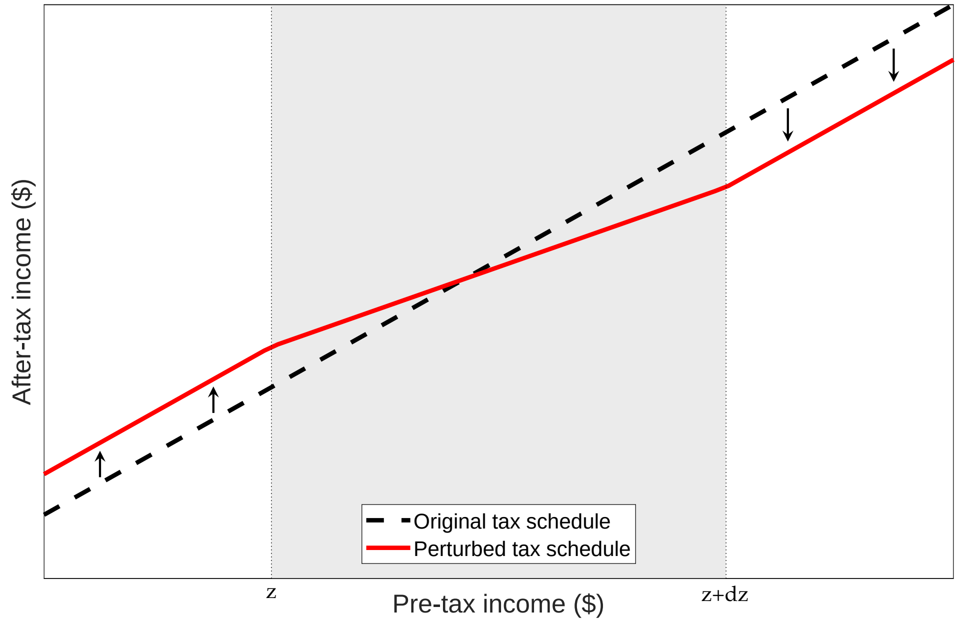


Figure notes. This figure shows the effect of a stylized perturbation. The hypothetical marginal tax change applies in the shaded region, lowering the slope of the relationship between after-tax and before-tax income.

An example of such an experiment is shown in Figure 3. Studying the effects of this perturbation leads to a tax formula that bears a conceptually close relationship to the standard one that arises when workers simply receive their marginal product (Mirrlees 1971, Diamond 1998, Saez 2001). As in the example above, there are three effects: a mechanical effect (**ME**), a fiscal externality (**FE**) and – new to this model – a belief externality (**BE**).

B. REGULARITY ASSUMPTIONS

In deriving a condition that characterizes the optimal tax, I take a continuously differentiable tax schedule T and the social planner's preferred equilibrium given that tax schedule, $E^* (T)$.[22] I adopt regularity assumptions, which jointly ensure that a worker's income responds smoothly to small changes in her wage or the tax schedule around this initial point, and that there is – generically, for an arbitrarily chosen tax schedule – a locally unique Fréchet differentiable function mapping tax schedules to investments.[23]

First, I make the standard single crossing assumption, which is that the marginal rate of substitution between income and consumption is decreasing in the wage (Assumption 1).

[22] Although I assume for concreteness that the planner can implement her preferred equilibrium, my approach is equally valid for any other locally continuous selection of equilibria.

[23] I discuss the existence of such a unique selection of equilibria in Appendix B. Appendix C then discusses why the planner may in some cases choose to locate at a singularity where these conditions break down.

Second, I assume that individuals' second-order conditions for labor supply hold strictly (Assumption 2). As discussed by Saez (2001), this requires that $1 - T'(z) + \varepsilon_z^c z T''(z) > 0$, where ε_z^c is the compensated elasticity of taxable income with respect to her wage. Assumption 2 can be viewed as a restriction on the curvature of the tax schedule. It always holds in my simulations, and must hold if $T''(z) \geq 0$.[24]

Assumption 1 (Single Crossing)**.** *The marginal rate of substitution between income and consumption,* $-\frac{u_l\left(c, \frac{z}{w}\right)}{w u_c\left(c, \frac{z}{w}\right)}$ *is decreasing in* w*.*

Assumption 2 (Labor Supply SOC)**.** *The second derivative of the tax schedule* $T''(z)$ *is bounded strictly below by* $-\frac{1}{\varepsilon_z^c z}\left[1 - T'(z)\right]$*.*

Third, I assume that investment returns are strictly concave, which implies that workers' second-order conditions for investment hold strictly (Assumption 3). This is a joint restriction on the tax schedule, cost distribution $G(k)$ and investment technology $Q(x)$. For any income, wage and productivity distributions, and any tax schedule, there exist underlying cost distributions and investment technologies such that Condition 13 holds. It can also be relaxed, with the key requirement being that workers are not indifferent between two investment levels. With finitely many cost types, this is satisfied generically for equilibrium investment choices; and with a continuum of cost types, the analysis is unchanged if it is violated for countably many types.

Assumption 3 (Investment SOC)**.** *Investment returns are strictly concave for all* x*.*

$$-\frac{Q''(x)}{Q'(x)^2} > \frac{\int_\Theta v(\theta|\pi, T) \left.\frac{\partial^2 f(\theta|q)}{\partial q^2}\right|_{q=Q(x)} d\theta}{\int_\Theta v(\theta|\pi, T) \left.\frac{\partial f(\theta|q)}{\partial q}\right|_{q=Q(x)} d\theta} \tag{13}$$

C. Mechanical Effect

Subject to these regularity assumptions, there are three effects of this perturbation. I begin with the mechanical effect (**ME**), which is isomorphic to the standard model in which workers are paid their marginal product. Raising the marginal tax at income z collects revenue from workers with income greater than z and redistributes it equally to all workers by raising the intercept of the tax schedule.

[24]Failure of Assumption 2 implies bunching of workers with different wages at the same level of income. Accounting for bunching is conceptually straightforward, but unnecessarily complicates the exposition.

Assumptions 1 and 2 jointly ensure that income is strictly increasing in θ. As a result, $z(\theta|\pi, T)$ can be inverted to obtain $\theta(z|\pi, T)$. Defining $G(k|\theta)$ as the distribution of k conditional on θ, and letting $\psi(k) = W'\left(\overline{V}(k, T)\right)$, the mechanical gain in welfare is:

$$\underbrace{\int_{\Theta} u_c(\theta) \int_K \psi(k) dG(k|\theta)\, d\theta \times \int_{\theta(z|\pi,T)}^{\overline{\theta}} f(\theta)\, d\theta}_{\textbf{Value of transfer to average worker}} - \underbrace{\int_{\theta(z|\pi,T)}^{\overline{\theta}} u_c(\theta) \int_K \psi(k) dG(k|\theta)\, f(\theta)\, d\theta}_{\textbf{Loss due to transfer from high income workers}}$$

To simplify this expression, let $H(z) = \int_0^z h(v)\, dv$ be the CDF of income. Secondly, let $\psi_z(z)$ be the normalized marginal social welfare weight of a worker with income z.

$$\psi_z(z) = \frac{u_c(\theta(z|\pi, T)) \int_K \psi(k) dG(k|\theta(z|\pi, T))}{\int_{\Theta} u_c(\theta) \int_K \psi(k) dG(k|\theta)\, f(\theta)\, d\theta}$$

Finally, define $\Psi(z) = \int_0^z \psi_z(v)\, h(v)\, dv$ as the cumulative welfare weight of workers with income less than z. Using these definitions, the mechanical gain can be written as:

$$\mathbf{ME}(z) = d\tau dz \times \{\Psi(z) - H(z)\}. \tag{14}$$

Since $\psi_z(z)$ is decreasing in z, the welfare weight below any finite level of income is higher than the population weight.[25] This in turn implies that $\mathbf{ME}(z) > 0$. Intuitively, transferring income from relatively rich individuals to the broader population of workers mechanically raises social welfare for a planner with a taste for redistribution.

D. Fiscal Externality

The second effect of the perturbation is a fiscal externality, which arises because workers ignore the effects of their decisions on government revenue. Although this effect appears in every income taxation model, it is more complicated with employer belief formation. Not only do all agents respond directly, but each response in turn changes the investment incentives of other workers. The fiscal externality is thus governed by the evolution of a fixed point at which workers' investment decisions are optimal given employers' beliefs and employers' beliefs are rational given workers' investment decisions.

The total fiscal externality is given by equation 15, and is comprised of two effects: changes in the level of income corresponding to each signal realization, and changes in the marginal density of the signal.[26] In turn, the income response for a given signal realization

[25] Since $\psi(k)$ is increasing in k, the assumptions on $f(\theta|q)$ guarantee that $\int_K \psi(k)\, dG(k|\theta)$ is decreasing in θ. Finally, $u_c(\theta)$ is weakly decreasing and $z(\theta|\pi, T)$ strictly increasing, so $\psi_z(z)$ is strictly decreasing.

[26] The derivatives in equation 15 are causal responses to this perturbation to the tax schedule. They cannot therefore be directly related to properties of the utility function.

captures two types of reaction: (i) direct responses of labor supply to the policy change; and (ii) changes in both wages and labor supply due to shifts in employer beliefs. Changes in the marginal density of the signal, $f(\theta)$, capture workers' investment responses.

$$\mathbf{FE}\left(z\right)=-d\tau dz\int_{\Theta}\Bigg\{\underbrace{T'(z(\theta|\pi,T))\left(\frac{dz(\theta|\pi,T)}{d\left[1-T'\left(z\right)\right]}\right)f(\theta)}_{\textbf{Response for given }\tilde{\theta}\textbf{ realization}}+\underbrace{T(z(\theta|\pi,T))\frac{df(\theta)}{d\left[1-T'\left(z\right)\right]}}_{\textbf{Investment response}}\Bigg\}d\theta \tag{15}$$

E. Belief Externality

The final effect of the perturbation is new to this model. When individuals re-optimize their investment decisions, they disregard their effects on the equilibrium wage paid for a given signal realization, $w(\theta|\pi)$. Taking any signal realization $\tilde{\theta}$, this wage externality is comprised of two components, corresponding to the two effects of investment: an increase in productivity, and a shift in the signal distribution.

$$\begin{aligned}\frac{dw(\tilde{\theta}|\pi)}{d\left[1-T'(z)\right]}f(\tilde{\theta})=\int_{K}\left(\frac{dq\left(k|\pi,T\right)}{d\left(1-T'\left(z\right)\right)}\right)\Bigg[&\overbrace{f(\tilde{\theta}|q(k|\pi,T))}^{\textbf{Productivity effect}} \\ &+\underbrace{\left[q\left(k|\pi,T\right)-E(q|\tilde{\theta},\pi)\right]\left(\frac{\partial f(\tilde{\theta}|q)}{\partial q}\bigg|_{q=q(k|\pi,T)}\right)}_{\textbf{Redistributive effect}}\Bigg]dG\left(k\right)\end{aligned} \tag{16}$$

The first component of equation 16 is the *productivity* effect. A worker who invests more shifts employers beliefs upward, and causes the wage paid to individuals with signal $\tilde{\theta}$ to rise despite their qualification decisions being unchanged.

The second component is the *redistributive* effect, and is generally negative. It arises because a worker who raises her investment has higher productivity than the group she leaves, but lower productivity than the group she joins. This manifests in the effect of investment on the distribution of signals observed by employers. If $f(\tilde{\theta}|q)$ increases, $w(\tilde{\theta}|\pi)$ rises if $q > E(q|\tilde{\theta},\pi)$ and falls if $q < E(q|\tilde{\theta},\pi)$. The opposite occurs if $f(\tilde{\theta}|q)$ decreases. The reason the effect tends to be negative is that an increase in q increases $f(\tilde{\theta}|q)$ at high values of $\tilde{\theta}$ where $q < E(q|\tilde{\theta},\pi)$ and decreases $f(\tilde{\theta}|q)$ at low values of $\tilde{\theta}$ where $q > E(q|\tilde{\theta},\pi)$.

The productivity and redistributive effects differ in both sign and incidence, as shown in Figure 5. When workers re-optimize, the productivity effect raises the equilibrium wages of workers whose signal distributions overlap most with those who increased their productivity. In contrast, the redistributive effect reduces the wages of workers who send signals in regions where the signal distribution *changes* most.

FIGURE 4: PRODUCTIVITY AND REDISTRIBUTIVE EFFECTS

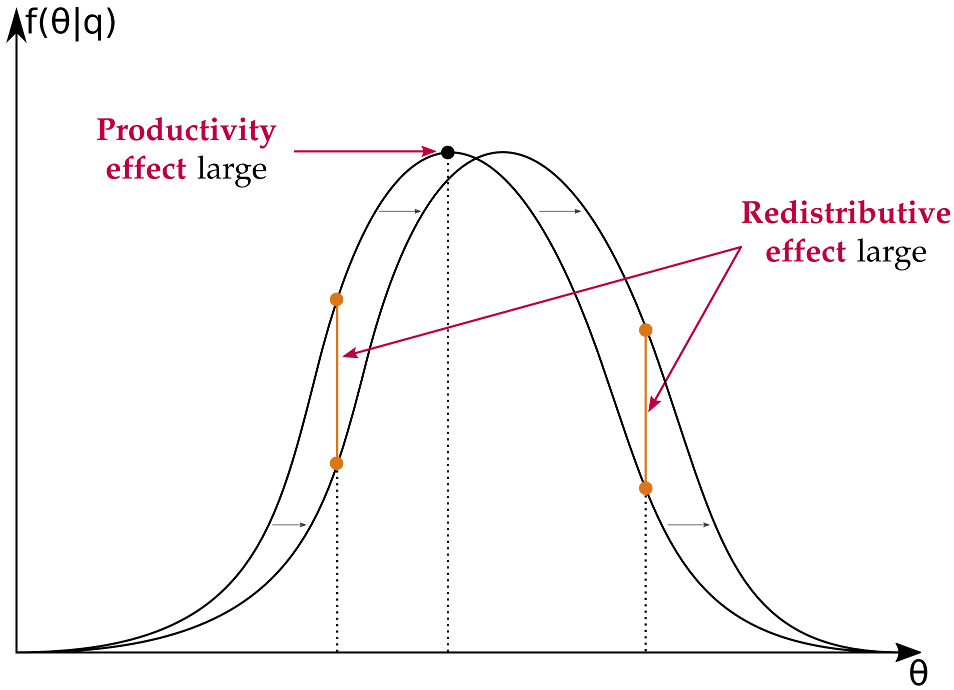


Figure notes. This figure demonstrates the areas in which the productivity effect and redistributive effects are largest. The productivity effect is proportional to the height of $f(\tilde{\theta}|q)$, while the redistributive effect is proportional to the change in $f(\tilde{\theta}|q)$.

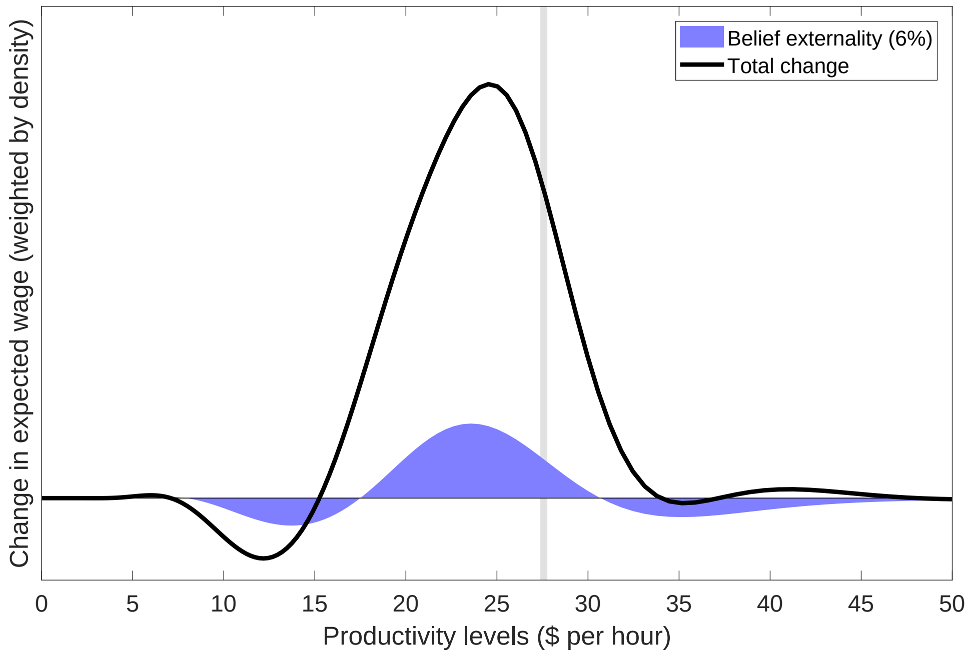
The importance of these differences in incidence are apparent in Figure 6, which starts from the linear tax example in Section 3 and shows the simulated effects on wages of a reduction in the marginal tax rate on income between \$60,000 and \$61,000. When the marginal tax rate falls and productivity rises, panel (a) shows that there is a large positive externality on workers around the epicenter of the productivity response, but also a negative effect on workers who are further way. If the overall externality is larger as in panel (b), the effects are dispersed more widely, and the positive productivity effect outweighs the negative redistributive effect over nearly all of the distribution.

The aggregate belief externality is calculated as follows. First, the effect on consumption is obtained by scaling the wage effect by labor supply and the retention rate, $1 - T'(z)$. Second, the effect on social welfare is obtained by multiplying by the welfare weight, ψ_z. Third, the total impact can be calculated by integrating over the signal distribution.

$$\mathbf{BE}(z) = -d\tau dz \left\{ \int_{\Theta} \psi_z(z(\tilde{\theta}|\pi, T)) \left[1 - T'(z(\tilde{\theta}|\pi, T))\right] l(\tilde{\theta}|\pi, T) \left(\frac{dw(\tilde{\theta}|\pi)}{d\left[1 - T'(z)\right]} \right) f(\tilde{\theta}) d\theta \right\}$$

FIGURE 5: RESPONSE TO A MARGINAL TAX RATE CHANGE

(a) Small Belief Externality ($s = 0.95$)



(b) Larger Belief Externality ($s = 0.75$)

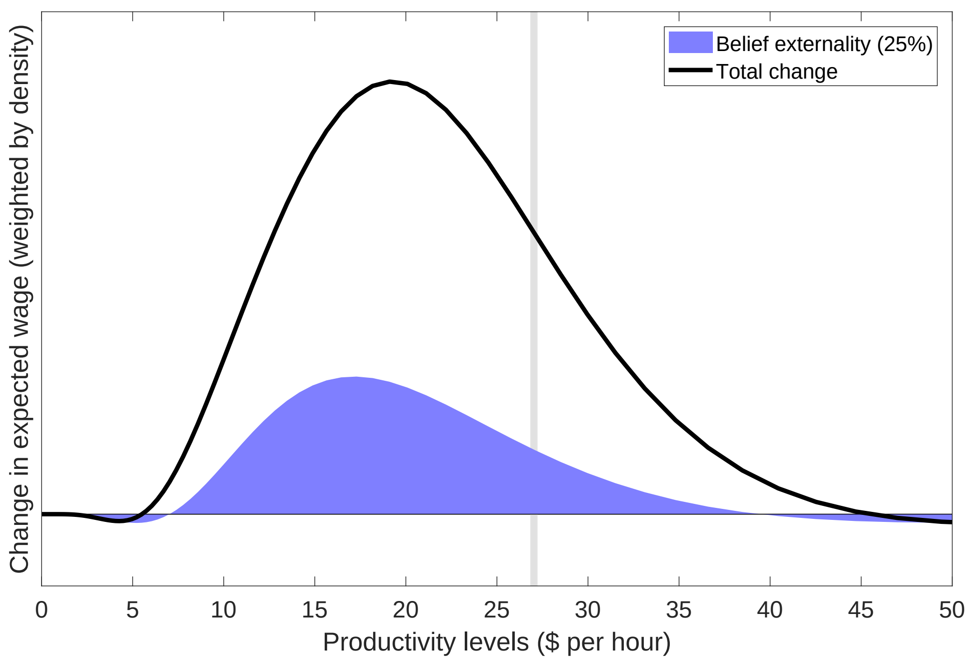


Figure notes. These figures show the effects on wages of a reduction in the marginal tax rate on income between \$60,000 and \$61,000. The baseline economy is the linear taxation example with $s = 0.75$, calibrated to match the United States wage distribution. The effect at each productivity level is scaled by the productivity density so that the area under each curve is proportional to the average wage change accounted for by that component. The gray-shaded bar shows the wage range that is *directly* affected by this perturbation. The productivity and income distributions are the same in both panels.

E. "U" SHAPED TAX SCHEDULES AND THE IMPORTANCE OF INCIDENCE

An intuitive special case arises when the tax rate is initially flat, there is no redistributive motive for taxation and labor supply is perfectly inelastic. In this case, incidence is irrelevant and the belief externality is proportional to the difference between the average productivity increase and the average private gain from investment.

$$-\mathbf{BE}\left(z\right) \propto \underbrace{\frac{d\overline{q}}{d\left(1-T'\left(z\right)\right)}}_{\textbf{Social benefit}} - \underbrace{\int_K \left[\int_\Theta \left(\left.\frac{df\left(\theta|q\right)}{d\left(1-T'\left(z\right)\right)}\right|_{q=q(k|\pi,T)}\right) w(\theta|\pi)d\theta\right] dG\left(k\right)}_{\textbf{Average private gain of responders}}$$

In the general case, wage changes due to the belief externality are re-weighted in a way that is important in driving the shape of the optimal tax schedule. The weights are given by: $\Omega\left(\theta\right) = \psi_z(z(\theta|\pi,T))\left[1-T'(z(\theta|\pi,T))\right]l(\theta|\pi,T)$. For two reasons, this pushes toward a "U" shape of the tax schedule. First, a given wage change is more important if it affects workers who supply a large amount of labor, and who receive significant social welfare weight. This implies larger weights at intermediate incomes. Second, the weights are proportional to the retention rate, which amplifies other forces in the model; in particular, this compounds the milder "U" shape that already arises from the usual trade-off between the mechanical effect of taxation and the fiscal externality (Diamond 1998).

F. A NECESSARY CONDITION FOR OPTIMALITY

Bringing everything together, the perturbation leads to three effects: **ME**(z), **FE**(z) and **BE**(z). Figure 6 shows these effects graphically. Just as in the example in Section 3, the mechanical effect from a reduction in the marginal tax rate transfers utility from workers with low productivity to those with high productivity, and the fiscal externality raises the utility of all workers. For most workers, the belief externality is also positive.

If T is optimal, the sum of the three effects must be equal to zero for all z. Otherwise, there exists a change to the tax schedule that raises welfare. This is summarized in Proposition 3, which also rewrites the belief externality in terms of the income distribution.

Proposition 3. *Consider an arbitrarily small perturbation that raises the marginal tax rate by $d\tau$ between income z and $z+dz$, with $d\tau$ second order compared to dz. The effect on social welfare is:*

$$\mathbf{ME}(z) + \mathbf{FE}(z) \underbrace{-d\tau dz \int_Z \tilde{z}\psi_z\left(\tilde{z}\right)\left(\frac{1-T'\left(\tilde{z}\right)}{1-T'\left(z\right)}\right)\varepsilon^{\pi}_{w(\tilde{z}),1-T'(z)} dH\left(\tilde{z}\right)}_{\mathbf{BE}(z)} \tag{17}$$

FIGURE 6: EFFECT ON UTILITY OF A MARGINAL TAX RATE CHANGE

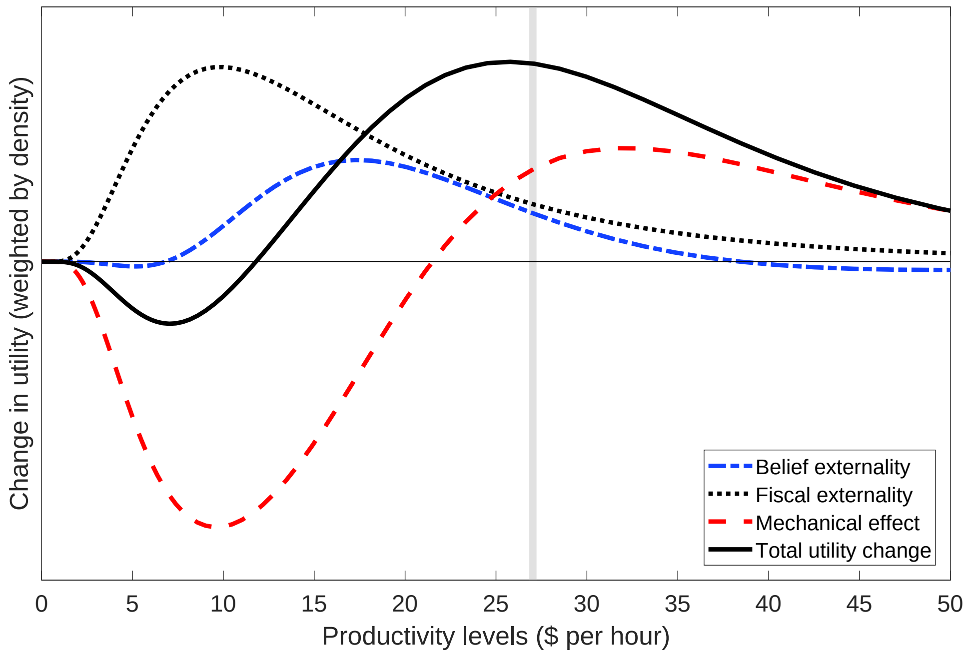


Figure notes. This figure shows the effects on utility of a rise in the marginal tax rate on income between $60,000 and $61,000. The baseline economy is the linear taxation example, calibrated to achieve $s = 0.75$ in equilibrium and match the United States wage distribution. The effect at each productivity level is scaled by the productivity density so that the area under each curve is proportional to the aggregate impact. The gray-shaded bar shows the wage range that is *directly* affected by this perturbation.

where:

$$\varepsilon^{\pi}_{w(\tilde{z}),1-T'(z)} = \frac{dw(\theta\left(\tilde{z}\right)|\pi)}{d\left[1-T'(z)\right]} \times \frac{1-T'(z)}{w(\theta\left(\tilde{z}\right)|\pi)}.$$

Except at a discontinuity, $\mathbf{ME}(z) + \mathbf{FE}(z) + \mathbf{BE}(z) = 0$ *for all* z *if* T *is optimal.*

If condition 17 is zero, there is no first-order gain from perturbing the tax schedule and moving to an equilibrium near the status quo. An alternative way of writing this, following Hendren (2016), is as a requirement that the cost and benefit of a policy change be equated.

$$\underbrace{\frac{1-\Psi\left(z\right)}{1-H\left(z\right)}}_{\text{Benefit per dollar}} = \underbrace{1+\frac{\mathbf{BE}\left(z\right)+\mathbf{FE}\left(z\right)}{1-H\left(z\right)}}_{\text{Cost per dollar}} \tag{18}$$

Here, a one dollar reduction in the tax rate at income z mechanically provides one dollar of consumption to workers with income greater than z. The direct benefit of this change is $1+\Psi\left(z\right) > 1-H\left(z\right)$ per dollar of mechanical expenditure, while the cost in addition to the mechanical expenditure is the sum of the two externalities.

A caveat to Proposition 3 is that equation 17 is necessary for optimality at an equilibrium around which there exists a continuous selection from the equilibrium correspondence, but not at points of discontinuity. Although Assumptions 1 to 3 ensure continuity for a generic tax schedule, one can construct examples in which the planner chooses to locate at such a discontinuous point if one exists. I take up this issue in Appendix C.

5 Quantitative Analysis

A. Evidence on Employer Learning

My next step is to quantitatively assess the importance of the belief externality using new and existing empirical evidence. First, I look to the literature on employer learning for evidence on the accuracy of employer beliefs. The dominant approach – pioneered by Farber and Gibbons (1996) and Altonji and Pierret (2001) – posits that the econometrician has access to a correlate of productivity from before workers entered the labor market, which employers cannot directly observe. In most cases, this is a pre-market score on the Armed Forces Qualification Test (AFQT).

These studies involve estimation of a version of the following regression using data from the National Longitudinal Survey of Youth (NLSY).

$$\begin{aligned}\ln w = \alpha_0 &+ \rho_0 \text{AFQT} + \rho_1 \text{AFQT} \times \text{Experience} \\ &+ \gamma_0 \text{Education} + \gamma_1 \text{Education} \times \text{Experience} \\ &+ \lambda_0 \text{Experience} + \lambda_1 \text{Experience}^2 + \lambda_2 \text{Experience}^3 + X'\beta + \varepsilon\end{aligned} \tag{19}$$

The typical finding is that ρ_1 is strictly positive. This is interpreted as evidence that employers do not initially reward workers fully for their productivity, but that the reward increases over time. A simultaneous finding that $\gamma_1 < 0$ further supports this conclusion, with the argument being that employers initially use education to gauge unobservable productivity; but that over time, they obtain more direct information about productivity and reduce their reliance on pre-existing correlates such as education.

Building on this approach, Lange (2007) estimates the speed of employer learning. He finds that employers' expectation errors take three years to decline to half their original values and five years to reach 36 percent. It then takes 26 years to reduce the remaining errors to less than 10 percent of their initial values. There is thus a long delay before a worker is fully rewarded for her productivity, as reflected by her AFQT score. In turn, this implies a substantial wedge between the private and social returns to improving it.[27]

A limitation of Lange's (2007) approach is that the evidence is confounded if productivity evolves heterogeneously over the lifecycle, since this would itself explain why the weight on AFQT increases with experience. Recognizing this, Kahn and Lange (2014) measure employer learning using a different method. Their key insight is that employer learning predicts that innovations in pay correlate more with past than future innovations in performance, because firms rely on *past* information to set pay.

Using a structural model and a panel dataset with information about both wages and performance reviews, Kahn and Lange (2014) find that workers capture between 60 and 90 percent of the social return to an innovation in their productivity during the first 15 years of their careers – although they capture a much smaller fraction in the later years.[28] This implies that 10 to 40 percent of the social return accrues to others, which is exactly the type of statistic required to calibrate my model.

Many other studies also suggest that employers imperfectly observe worker productivity. For example, MacLeod, Riehl, Saavedra and Urquiola (2017) study the introduction of college exit exams in Colombia. They show that when more information about productivity becomes available, wages begin to more strongly reflect individual ability rather than college reputation. In a similar vein, there is evidence from online marketplaces that information is imperfect (Stanton and Thomas 2016), and that additional information can improve outcomes (Pallais 2014, Pallais and Sands 2016).

Furthermore, numerous studies uncover evidence of statistical discrimination, which implies imperfect information. For example, Blair and Chung (2018) find that occupational licensing reduces reliance on race and gender; and drug testing is shown by Wozniak (2015) to positively impact black employment. Conversely, Agan and Starr (2018) and Doleac and Hansen (2016) show that racial discrimination increases when employers are banned from asking about criminal histories; and Shoag and Clifford (2016) find that banning the use of credit checks leads to relative increases in employment in low credit score census tracts, and more demand for other information about productivity.

B. New Evidence on Heterogeneity in Learning

The results above help calibrate the relationship in my model between productivity q and expected wages $E(w)$, one minus the slope of which is the external wage effect of investment. However, there is only limited evidence on how employer learning varies with

[27] The evidence also suggests that education has a causal impact on AFQT scores (Neal and Johnson 1996, Hansen, Heckman and Mullen 2004), implying that they do not simply measure *innate* ability.

[28] These data come from a firm in the United States, first analyzed by Baker, Gibbs and Holmstrom (1994).

productivity. For example, Arcidiacono, Bayer and Hizmo (2010) find faster learning for college graduates than other workers, which suggests a larger externality at the low end; and Lindqvist and Westman (2011) show that non-cognitive skills – likely the hardest for employers to learn – are most important at low levels of income.

Here, I provide more direct evidence on how learning varies over the productivity distribution. Taking AFQT as a proxy for productivity, I adapt equation 19 by interacting the variables of interest with indicators $\mathbf{I}_A = \mathbb{1}\,(\text{AFQT} > m)$ and $\mathbf{I}_B = \mathbb{1}\,(\text{AFQT} \leq m)$ for whether a worker's AFQT score is above or below the median, m.

$$\begin{aligned}\ln w = \sum_{j\in\{A,B\}} \Big\{ & \rho_{0,j}\text{AFQT} + \rho_{1,j}\text{AFQT} \times \text{Experience} \\ & + \gamma_{0,j}\text{Education} + \gamma_{1,j}\text{Education} \times \text{Experience} \\ & + \lambda_{0,j} + \lambda_{1,j}\text{Exper.} + \lambda_2\text{Exper.}^2 + \lambda_3\text{Exper.}^3 \Big\} \times \mathbf{I}_j + X'\beta + \varepsilon\end{aligned} \tag{20}$$

I then estimate equations 19 and 20 using data from the NLSY79 survey.

The sample follows Arcidiacono et al. (2010).[29] It restricts to black and white men who are employed, have wages between one and one hundred dollars, and at least eight years of education. Following Altonji and Pierret (2001), I also limit the analysis to workers with fewer than 13 years of experience – measured as the number of years a worker has spent outside of school.[30] Employment in the military, at home or without pay is excluded.

Table 1 shows the results. The dependent variable is the log of each worker's real hourly wage, multiplied by 100; and AFQT scores are standardized to have mean zero and unit standard deviation for each age at which the test was taken. The coefficient on AFQT is therefore approximately the percentage wage gain associated with a one standard deviation higher AFQT score. The coefficient on the interaction of AFQT with experience is the number of percentage points that this increases by with each year of experience.

Below the median, there is strong evidence of learning: the weight on AFQT rises steeply with experience, and the weight on education falls comparatively quickly. There is less evidence of learning above the median, where the coefficient on the interaction between AFQT and experience is very close to – and statistically indistinguishable from – zero. The large direct effect of AFQT in the upper half of the distribution suggests that these results are not simply driven by AFQT being unimportant at the high end; and the less negative interaction between education and experience supports the conclusion that learning is driving the results. All of these conclusions are robust to restricting the sample to workers who have exactly twelve or sixteen years of education.

[29]Summary statistics for workers with high and low AFQT scores are available in Table 3 of Appendix H.
[30]The relationship between log wages, AFQT and experience is approximately linear in this region.

TABLE 1: HETEROGENEITY IN EMPLOYER LEARNING

	12 or 16 Years Education		Full Sample	
Whole sample				
AFQT	2.63*		2.90**	
	(1.49)		(1.32)	
AFQT × Experience	0.94***		0.87***	
	(0.19)		(0.16)	
Education	11.09***		8.02***	
	(0.87)		(0.64)	
Education × Experience	−0.30**		−0.21**	
	(0.12)		(0.08)	
Below median AFQT				
AFQT		5.14**		2.51
		(2.56)		(2.12)
AFQT × Experience		1.11***		1.21***
		(0.32)		(0.27)
Education		10.10***		7.21***
		(1.42)		(0.87)
Education × Experience		−0.35*		−0.25**
		(0.20)		(0.11)
Above median AFQT				
AFQT		6.55		5.89*
		(4.02)		(3.51)
AFQT × Experience		−0.05		0.49
		(0.54)		(0.45)
Education		11.33***		8.18***
		(0.95)		(0.75)
Education × Experience		−0.27**		−0.17*
		(0.13)		(0.10)
Observations	15884	15884	25659	25659
Clusters	2553	2553	3673	3673
R^2	0.30	0.30	0.30	0.30

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table notes. Dependent variable is the worker's log hourly wage multiplied by 100. AFQT is a worker's score on the armed forces qualification test, standardized by age to have zero mean and unit standard deviation. Education and experience are measured in years. Standard errors, shown in parentheses, are clustered at the worker level. All regressions include an indicator for urban vs. rural, race, race×experience, and region and year fixed effects. Data are from the National Longitudinal Survey of Youth (NLSY79). The sample is restricted to working black and white men who have wages between one and one hundred dollars, at least eight years of schooling and fewer than 13 years of experience. NLSY sample weights are used.

C. Evidence on the Response of Productivity to Returns

The second piece of evidence I require is an estimate of the relative responsiveness of productivity, compared to taxable income. Although there is less evidence available on the productivity response than there is for labor supply, a precise short-run estimate is provided by Blomquist and Selin (2010) using a difference-in-difference approach applied to a tax reform in Sweden. Blomquist and Selin's (2010) results suggest that around three quarters of the response of taxable income comes through wages.[31] This is consistent with a calibration by Trostel (1993) which suggests that 60 to 80 percent of the long run response of income to taxation comes from changes in labor productivity.

There is also qualitative evidence that longer-run human capital investments respond. First, Abramitzky and Lavy (2010) study the reduction in effective marginal tax rates that occurred when Israeli *kibbutzim* shifted from equal-sharing to productivity-based wages.[32] They find that the reform led to sharply higher graduation rates and test scores. Second, Kuka, Shenhav and Shih (2018) study the introduction of the Deferred Action for Childhood Arrivals (DACA) program, which increased returns to human capital investment. They show that high school graduation and college attendance rates increased markedly for eligible individuals. Finally, studies have demonstrated that human capital investments increase when students are simply informed about returns (e.g., Jensen 2010).

MacLeod et al.'s (2017) study of college exit exams also provides interesting evidence. As more information is provided to employers, and wages begin to more closely track individual ability, *average* wages rise by seven percent given the same formal educational investments. This rise in wages is consistent with a response of human capital investment to the higher return to individual ability, although it could also be explained by improved matching between workers and tasks.

D. Calibration to the United States Economy

I now calibrate the model to match both the evidence above and the empirical United States wage and income distributions. Table 2 summarizes the assumptions needed and my choices for them. Results with alternative calibrations are available in Appendix H.

The wage schedule that I target is the Pareto log-normal approximation provided by Mankiw, Weinzierl and Yagan (2009) using March CPS data. However, a wage schedule cannot be assumed directly. Rather, equilibrium wages are jointly implied by productivity and signal distributions. The approach I take is to posit a conditional signal distribution,

[31] This may be conservative since Blomquist and Selin cannot capture long-run human capital responses.
[32] Kibbutzim are small collective communities in Israel.

TABLE 2: CALIBRATED AND IMPLIED OBJECTS

Assumed object	Assumption	Implied object	Implied value
Social welfare function	$\log(E(U))$	Income elasticity	$\varepsilon_z^{LR} = 1.1$
Noise distribution	LN, $\text{var}(\theta\|q) = 7q$	Wage elasticity	$\varepsilon_w^{LR} = 0.7$
Labor supply elasticity	$\varepsilon_l = 0.25$	Cost distribution	Champernowne: $\lambda = 1.3$,
Production concavity	$\beta = 0.3$		$\alpha = 2.46$, $y_0 = 2.7$
Wage distribution	Pareto LN: $a = 2$,	External fraction	
	$\sigma_q^2 = 0.56$, $\mu_q = 2.76$	of return	0.15 (average)

Table notes. This table summarizes the key assumptions underlying the simulation described in Section 5. Objects in the left column are calibrated directly, while the target objects in the right column are implied. See text and Appendix H for further details and simulations with alternative calibrations.

$f(\theta|q)$, and then find a distribution for productivity that yields a wage distribution as close as possible to the target.[33] As panel (a) of Figure 7 shows, this exercise is successful. Notably, the Pareto right tail is closely replicated in addition to the overall shape.

Next, I choose a signal distribution so that, on average, a worker who increases her productivity by one dollar receives an 85 cent higher expected wage. This is at the upper end of the estimates provided by Kahn and Lange (2014). In line with my empirical results, I also ensure that there is a flatter slope at the low end of the income distribution. I achieve these aims by assuming a conditionally log-normal signal distribution with $E(\theta|q) = q$, and $\text{var}(\theta|q)$ linearly increasing in q. Panel (b) of Figure 7 displays the results.

I choose the remaining parameters to target income, wage and labor supply elasticities. First, I set $\varepsilon_l = 0.25$, which is approximately in line with estimates of the intensive-margin labor supply elasticity (e.g., Chetty 2012). Second, I calibrate the long run elasticities of each variable to the retention rate.[34] Although these cannot be directly assumed, they are closely connected to the concavity of the production function, β. I choose a value for β that produces long-run elasticities of wages and labor supply of 0.7 and 0.4 respectively.[35] This implies that around 60 percent of the long run response of taxable income comes from changes in labor productivity, which is at the low end of the estimates above. The same estimates imply an overall elasticity of taxable income of 1.1.[36]

[33] Specifically, I parameterize a Champernowne (1952) distribution – a family of bell-shaped distributions designed to fit empirical income distributions – to minimize the Kullback-Leibler divergence between the equilibrium and target distributions under the 20 percent linear tax from which the simulation starts.

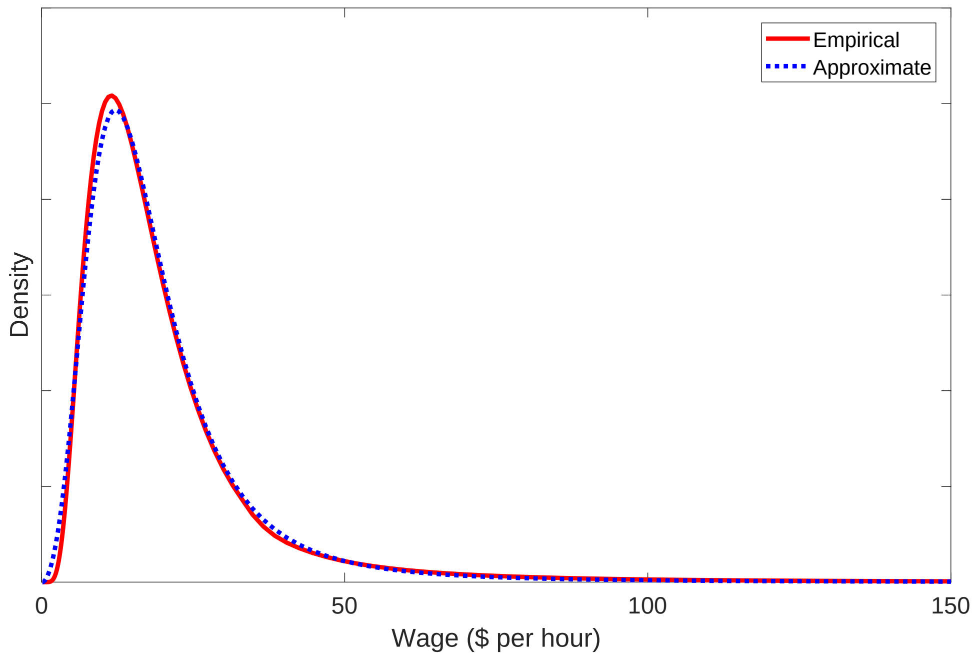
[34] These long-run equilibrium responses take into account the multiplier effect from human capital investment: more productive workers work more, but working more further raises the return to investment.

[35] Long-run elasticities vary with the tax system and over the income distribution. The statistics quoted here are based on responses of aggregate wages and labor supply to a small change to a flat 20 percent tax.

[36] This is in line with Mertens and Montiel-Olea (2018), whose estimates could be viewed as conservative since they do not take into account long-run human capital responses. I provide simulations with different parameter values in Appendix H. Holding $\varepsilon_q^{LR}/\varepsilon_z^{LR}$ constant, the level of ε_z^{LR} is not important to my results.

FIGURE 7: EQUILIBRIUM RELATIONSHIPS IMPLIED BY THE CALIBRATION

(a) Wage Density



(b) Expected Wage as a Function of Productivity

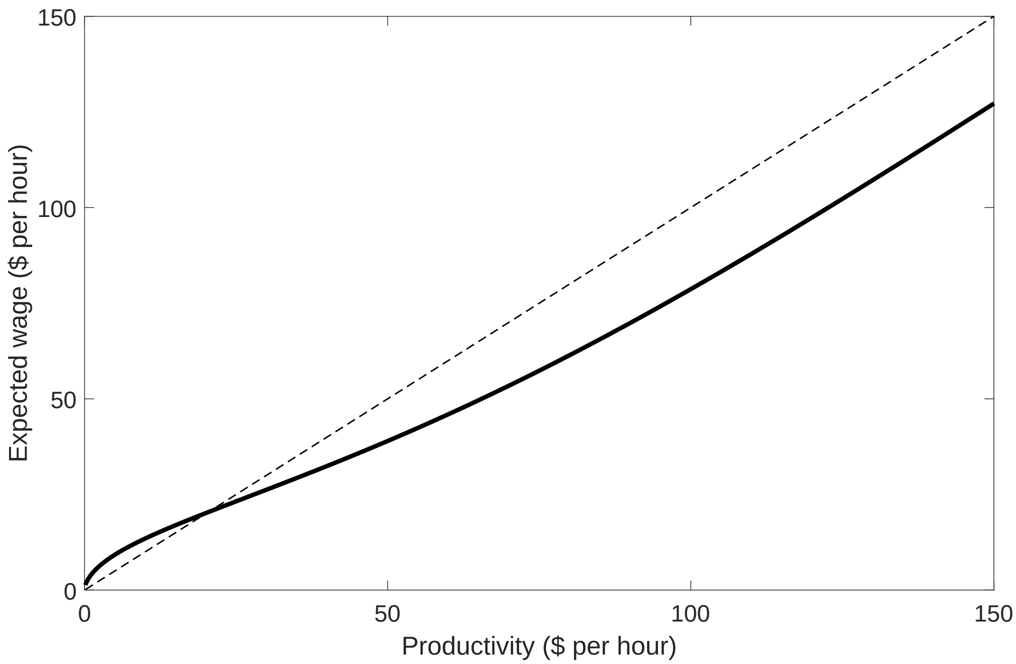


Figure notes. These figures show the implications of the calibration procedure for the simulation described in Section 5. Panel (a) compares the empirical (target) and approximate (simulated) wage distributions. Panel (b) shows the relationship between expected wages and productivity in the baseline economy.

E. Solving for Optimal Taxes

To simulate the model, I start with an initial tax schedule T_0 and a known equilibrium. I then consider adopting an alternative tax schedule, T_1, under which the marginal tax rate is raised or lowered by $\Delta T'$ over a range of incomes from $\underline{z}$ to $\overline{z}$.

$$T_1'(z) = \begin{cases} T_0'(z) + \Delta T' & \text{if } z \in (\underline{z}, \overline{z}) \\ T_0'(z) & \text{otherwise} \end{cases}$$

Given T_1, I re-calculate the expected utility of workers with each level of productivity, and let workers adjust their human capital investments. Next, I re-solve for employer beliefs, and wages, given the new productivity distribution. From here, I repeatedly re-optimize human capital decisions and re-calculate beliefs until a fixed point is obtained. At this fixed point, employers' beliefs and workers' investment decisions are mutually consistent. Finally, I calculate expected utility for each individual, weight using the social welfare function and adopt the new tax schedule if the welfare gain is positive.

This is the procedure that underlies Figures 2, 5 and 6. It can be continued repeatedly, starting with large perturbations and ending with smaller ones until the gain to the marginal perturbation is approximately zero. At this point, condition 17 of Proposition 3 approximately holds. I refer to this final tax schedule as *optimal*. Further details of this process are available in Appendix H.

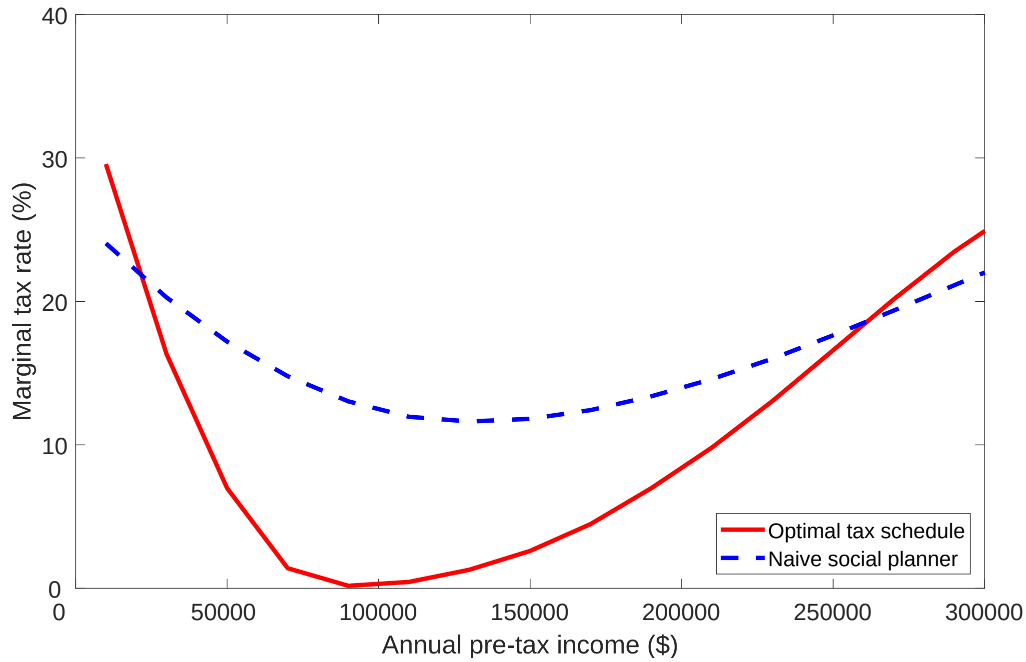
F. A Naïve Benchmark for Comparison

As a benchmark against which to compare the optimal tax schedule, I imagine a naïve planner who neglects to take into account the fact that part of the response of wages to taxation arises due to an externality, and is thus not internalized by investors. This means that she neglects the novel effect of a perturbation of the tax schedule, **BE** (z). Instead, she simply equates the fiscal externality and the mechanical effect, as would be the correct approach in a model with perfect employer information.

Comparison of the naïve and optimal tax schedules therefore facilitates an assessment of the quantitative importance of the belief externality. This is similar in spirit to Rothschild and Scheuer's (2013) concept of a *self-confirming policy equilibrium* (SCPE). However, the planner is more sophisticated here in that she is aware when measuring the fiscal externality that she must take into account both wage and labor supply responses. Similarly, she knows that the mapping from productivity to wages is stochastic. However, she is unaware that part of the change in equilibrium wages arises due to a spillover that workers ignore when re-optimizing.

FIGURE 8: OPTIMAL NON-LINEAR TAXATION

(a) Non-linear Taxation



(b) Decomposition of a Marginal Tax Cut in Each Tax Bracket

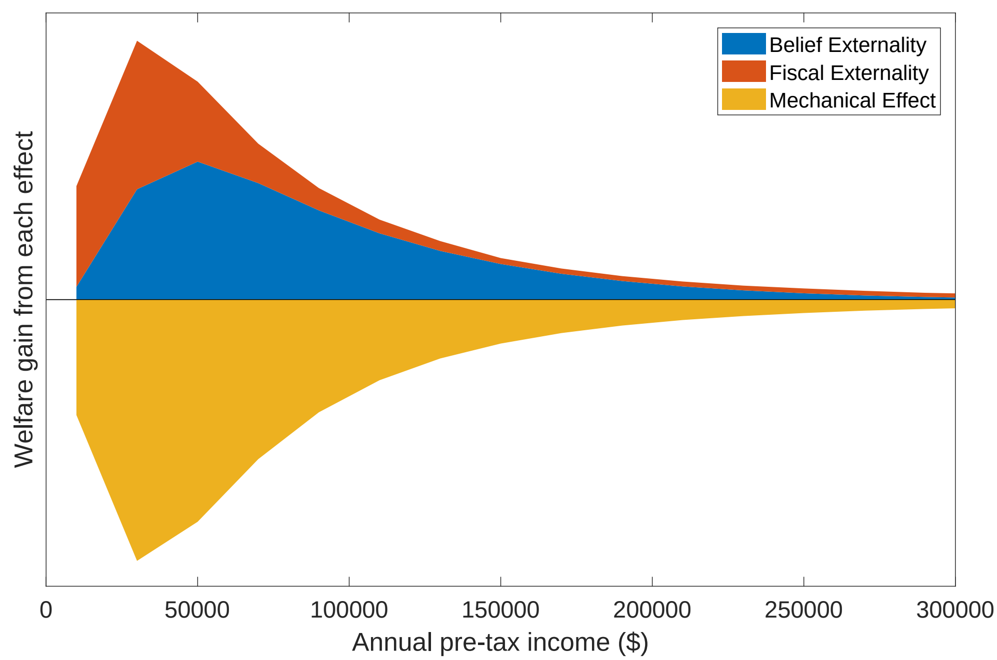


Figure notes. This figure shows the results of the simulation described in Section 5. The solid red line in panel (a) shows the optimal tax schedule, while the dashed blue line shows a tax schedule that would be accepted by a naïve social planner who sets the sum of the mechanical effect and fiscal externality equal to zero. Panel (b) shows a decomposition of small marginal tax cuts in each tax bracket. The tax function in this simulation is discretized into $20,000 brackets. Details of the procedure are available in Appendix H.

Figure 8 shows the tax schedules produced by this procedure. The red line shows a tax schedule that satisfies equation 17, so that there is no first order gain from a small perturbation in any tax bracket. The blue line would satisfy a naïve social planner, because the mechanical effect and the fiscal externality sum to zero.[37] Marginal tax rates are generally substantially lower under the optimal than the naïve schedule, reflecting the planner's additional incentive to encourage investment by lowering taxes.

Both tax schedules have the familiar "U" shape, which comes from the trade-off between the mechanical effect and the fiscal externality when the income distribution has a Pareto right tail (Diamond 1998). This shape is accentuated under optimal taxation because the belief externality is more important at intermediate incomes (see Figure 8). In part, this is because a given wage impact from the belief externality is less important at high incomes where social welfare weight is low, and at low incomes where little labor is supplied; and in part the shape is due to variation in the wage impact of the externality.[38]

At very high incomes, the optimal tax schedule is above the naïve tax schedule. This is for two reasons. First, as income rises, the belief externality becomes arbitrarily small so that the planner simply trades off the mechanical effect and the fiscal externality. Second, changes in marginal tax rates at high incomes shift investment incentives throughout the productivity distribution; and most of those who respond now face lower tax rates most of the time – implying a smaller fiscal externality from their re-optimization.[39]

At very low incomes, optimal tax rates are also higher. In this case, the main reason is that the downward adjustment to taxes throughout the distribution raises expected utility for most workers, but not those at the bottom (see Figure 9). As a result, welfare weights rise at the lowest incomes. This increases the mechanical gain from raising marginal taxes at low incomes and collecting infra-marginal income all but the lowest-income workers.

The welfare gain from transitioning to the optimal tax schedule is equivalent to raising the consumption of all workers by one percent, holding labor supply and investment decisions fixed. However, this is not a Pareto improvement. As Figure 9 shows, individuals of moderate-to-high productivity experience large gains in utility. But workers with very low productivity are worse off because the government collects less revenue. This means that the transfer to the lowest-income household is five percent smaller. Workers with the highest productivity levels are also hurt, due to higher tax rates at top incomes.

[37] The algorithm to find the naïve schedule is conceptually identical to that used for the optimal schedule.

[38] The shape is further amplified because the belief externality scales with with the retention rate, $1 - T'(z)$.

[39] Comparisons of the mechanical effect, fiscal externality and belief externality for marginal perturbations under the naïve and optimal tax schedules are available in Appendix H.

FIGURE 9: UTILITY GAIN FROM OPTIMAL TAXATION

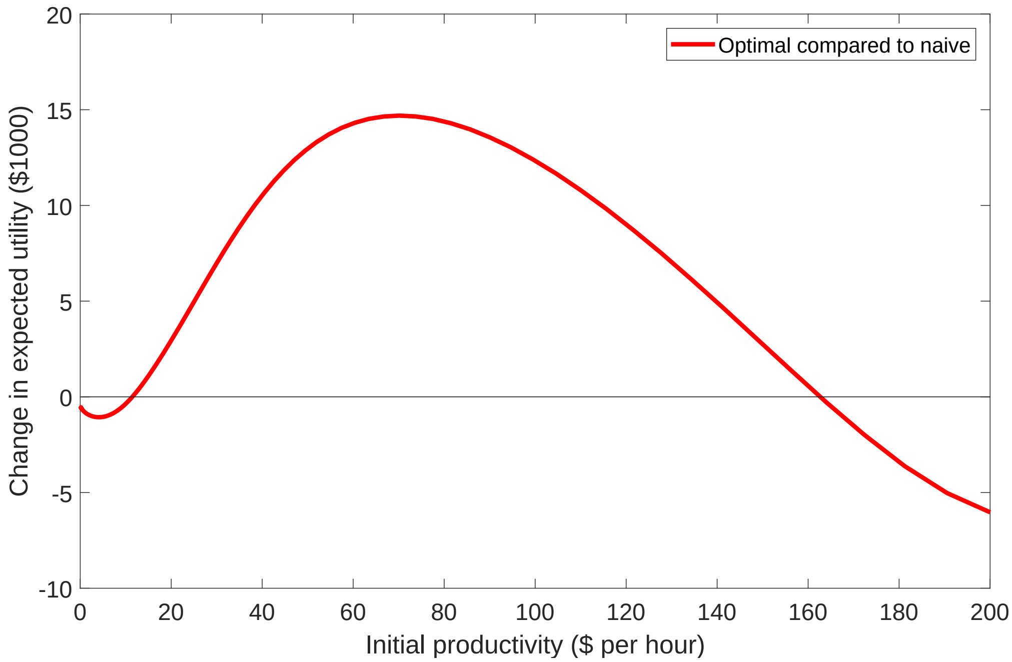


Figure notes. This figure compares the utility levels of agents at each productivity level under naïve and optimal taxation in the simulation described in Section 5. Further details are included in Appendix H.

H. APPROXIMATELY OPTIMAL TAXATION

Although the optimal tax formula cannot be written in terms of sufficient statistics, 60 percent of the gain from optimal taxation can be obtained using a simple approximation based on two principles. First, assume that a change in $T'(z)$ causes workers with income close to z to respond. Second, assume that the incidence of the externality falls on workers with welfare weight, labor supply and tax rate similar to those with income z. In Appendix E, I show that these principles yield:

$$\mathbf{FE}(z) + \mathbf{ME}(z) - (1 - s(z))\,\psi_z(z)\,l(z)\left[1 - T'(z)\right]\frac{d\overline{w}}{d\left[1 - T'(z)\right]} = 0 \qquad (21)$$

where $l(z)$ is the labor supply of workers with income z, $d\overline{w}/[1 - T'(z)]$ is the response of average wages, and $s(z)$ is the share of the wage change that is not internalized.

An advantage of equation 21 is that it facilitates assumptions about how the belief externality varies with income, without having to find distributional assumptions which produce that profile. As in Section 4, the correction term is larger if investment is more responsive, investing workers capture little of the return to investment; or if workers supply a large amount of labor, face a low tax rate and receive significant welfare weight.

6 Unproductive Signaling

I have so far assumed that investment is productive in the sense that its sole effect is to increase productivity. It is also possible for investment to play a 'pure' signaling role in the sense of Spence (1973).[40] Investment then reflects both human capital accumulation and a worker's immutable ability. In this case, the externality from investment may, in general, be more positive or more negative than in my baseline model.

To allow for this type of signaling, I replace the production function with $q = Q(x,k)$ so that productivity is a direct function of the worker's type. Secondly, I assume that employers observe a signal of investment rather than productivity. Specifically, $\theta \in \Theta \subseteq \mathbb{R}_+$ has conditional density $f(\theta|x)$ twice differentiable in x, and full support for all x. As before, it satisfies the monotone likelihood ratio property: $\frac{\partial}{\partial\theta}\left(\frac{f(\theta|x_H)}{f(\theta|x_L)}\right) > 0$ for all $x_H > x_L$. Otherwise, I adopt all the assumptions from Section 2.

A. Unproductive Signaling: Example with Linear Taxation

I begin by adapting the example in Section 3. Specifically, let productivity $q = n^{\alpha}h^{1-\alpha}$ be a function of both human capital $h = x^{\beta}$, and inherent ability n, with ability negatively related to the worker's investment cost: $n = 1/k$. As before, assume that the distribution of ability and the conditional signal distribution are log-normally distributed.

$$n \sim LN\left(\ln \mu_n - \frac{\sigma_n^2}{2}, \sigma_n^2\right) \qquad \ln\theta = \ln x + \ln\xi \qquad \ln\xi \sim N\left(0, \sigma_\xi^2\right)$$

With these assumptions, there again exists an equilibrium in which income and productivity are log-normally distributed. Similar to the original example, the elasticities of productivity and income depend only on the labor supply elasticity ε_l, the concavity of the production function β and, in this case, the relative importance of ability, α.

Proposition 4. *For any tax rate τ, there is an equilibrium in which productivity and income are log-normally distributed. Assuming this equilibrium is played, the elasticities of productivity and investment with respect $1-\tau$ are as follows.*

$$\varepsilon_q = \frac{\beta(1-\alpha)(1+\varepsilon_l)}{1-\beta(1-\alpha)(1+\varepsilon_l)} \qquad \varepsilon_z = \frac{\varepsilon_l + \beta(1-\alpha)(1+\varepsilon_l)}{1-\beta(1-\alpha)(1+\varepsilon_l)}$$

[40] It is hard to assess the contribution of unproductive signaling to the return to education. Evidence from large-scale school reforms demonstrate large productive effects of education (Meghir and Palme 2005, Aakvik, Salvanes and Vaage 2010, Oreopoulos 2006) but there is also some evidence to suggest a role for unproductive signaling (Lang and Kropp 1986, Bedard 2001). See Lange and Topel (2006) for a discussion.

This nests the original example in Section 3 in which investment is purely productive: when $\alpha = 0$, the two elasticities ε_q and ε_z collapse to that case, and equation 22 collapses to equation 10. Alternatively, when $\alpha = 1$, productivity does not respond to taxation, and the income elasticity collapses to the elasticity of labor supply.

The first-order condition for the optimal tax is given by Proposition 5. It features a second externality correction, $1 + s\alpha(1 + \varepsilon_l)$, which pushes toward *higher* rather than lower taxes. Intuitively, there is no social benefit from the component of the private return to investment that comes from signaling innate ability, which in turn implies that this return comes at the expense of others. The logic is similar to the redistributive effect in Section 4: a worker who invests more negatively affects other workers, because she has higher productivity than the group she leaves, but lower productivity than the group she joins.

Proposition 5. *Assume that the log-normal equilibrium described in Proposition 4 is played. Then the first-order condition for the optimal linear tax τ^* is:*

$$\frac{\tau^*}{1-\tau^*} = \frac{1 - \gamma\left[\frac{1+(1-s)\varepsilon_q}{1+s\alpha(1+\varepsilon_l)}\right]}{\varepsilon_z} \tag{22}$$

where $s = \frac{\sigma_x^2}{\sigma_x^2+\sigma_\varepsilon^2}$ and $\gamma = E_n\left(\frac{\psi_n}{\psi}\frac{\overline{z}_n}{\overline{z}}\right)$.

Since imperfect employer information now generates two opposite-signed externalities, there exist combinations of α and s that cause them to perfectly offset each other.

$$\underbrace{\beta(1-\alpha)}_{\textbf{Social benefit}} = \underbrace{s\left[\frac{\alpha+\beta(1-\alpha)}{1+s\alpha(1+\varepsilon_l)}\right]}_{\textbf{Private benefit}} \qquad \Longleftrightarrow \qquad \underbrace{s\alpha(1+\varepsilon_l)}_{\textbf{Signaling}} = \underbrace{(1-s)\varepsilon_q}_{\textbf{Learning}}$$

The first condition states that the social and private benefits of investment are aligned. The second states that the unproductive component of the private return is equal in magnitude to the part of the productive component that is not captured by the individual. If these conditions hold, then condition 22 collapses to the standard optimal tax formula. However, any other parameter values imply an efficiency role for intervention.

An implication of the equations above is that less accurate employer information (lower s) implies a smaller private benefit of investment with an unchanged social benefit. Stated equivalently, lower s means that the signaling externality is smaller and the learning externality is larger. In this sense, evidence of residual employer uncertainty (Lange 2007, Kahn and Lange 2014) suggests a more positive externality, and lower taxes.

B. Non-Linear Taxation: Signaling with Observable Investment

To further build intuition, I now relax the parametric assumptions of the example and consider non-linear taxation, but in a special case of the model in which employers perfectly observe investment. There is then a deterministic equilibrium mapping from investment to wages, $w(x)$. Taking this as given, the worker's investment problem is:

$$\max_{x \in \mathbb{R}_+} \quad v(w(x)|T) - kx \tag{23}$$

where:

$$v(w(x)|T) = \max_{l \in \mathbb{R}_+} u(w(x)l - T(w(x)l), l). \tag{24}$$

The solutions to problem 23 for each cost type jointly define a second mapping, $x(k)$, from costs to investment levels.

To simplify the analysis, I assume $w(x)$ is one-to-one. Then, given this assumption, I provide conditions in Appendix F to guarantee that $x(k)$ and $w(x)$ are differentiable, which ensures that the investment choice for a worker with cost k is characterized by:

$$u_c(z(k) - T(z(k)), l(k)) \left[1 - T'(z(k))\right] l(k) w'(x(k)) = k \tag{25}$$

where $l(k)$ is the level of labor supply that solves problem 24, and $z(k) = w(x(k)) l(k)$ is the equilibrium income of a worker with cost k.

As is well known, the equilibrium relationship between innate ability and investment drives a wedge between the private and social returns, which I refer to as the unproductive component.

$$\underbrace{\frac{Q_k(x(k),k)}{x'(k)}}_{\textbf{Unproductive}} = \underbrace{w'(x(k))}_{\textbf{Private}} - \underbrace{Q_x(x(k),k)}_{\textbf{Productive (social)}} \tag{26}$$

If $Q_k(x(k),k) < 0$ so that costs are positively related to ability, there is a positive externality from investment: an individual who invests more makes others look better because she has higher productivity than those who invest at that level in equilibrium. Conversely, if $Q_k(x(k),k) < 0$, there is a negative externality from investment.

These results provide a foundation for policy analysis that mirrors Section 4. Specifically, consider again a perturbation that raises the marginal tax rate by $d\tau$ on income between z and $z + dz$, while raising the intercept of the tax schedule to ensure that the resource constraint still holds. A different but related form of belief externality arises.

$$\textbf{BE}(z) = -d\tau dz \int_K \psi(k) \left[1 - T'(z(k))\right] l(k) \frac{dx(k)}{d\left[1 - T'(z)\right]} \left[w'(x(k)) - Q_x(x(k),k)\right] dG(k)$$

This equation for **BE** (z) can again be written in terms of the observable income distribution, and combined with the fiscal externality and mechanical effect to obtain a necessary condition for optimality of the tax system:

$$\mathbf{FE}(z) + \mathbf{ME}(z) + \int_Z \tilde{z}\psi(\tilde{z}) \left(\frac{1-T'(\tilde{z})}{1-T'(z)}\right) \varepsilon_{\tilde{x}(\tilde{z}),1-T'(z)} \left[\varepsilon^{\text{Private}}_{\tilde{w}(\tilde{z}),\tilde{x}(\tilde{z})} - \varepsilon^{\text{Social}}_{\tilde{w}(\tilde{z}),\tilde{x}(\tilde{z})}\right] dH(\tilde{z}) = 0 \tag{27}$$

where $\tilde{w}(\tilde{z})$ and $\tilde{x}(\tilde{z})$ are the wages and investment levels of a worker with income $\tilde{z}$, and the elasticities are defined as follows.

$$\varepsilon^{\text{Private}}_{\tilde{w}(\tilde{z}),\tilde{x}(\tilde{z})} = w'(x(k))\frac{x(k)}{w(k)} \qquad \varepsilon_{\tilde{x}(\tilde{z}),1-T'(z)} = \frac{dx(\tilde{z})}{d[1-T'(z)]}\frac{1-T'(z)}{x(\tilde{z})}$$

$$\varepsilon^{\text{Social}}_{\tilde{w}(\tilde{z}),\tilde{x}(\tilde{z})} = Q_x(x(k),k)\frac{x(k)}{w(k)}$$

Note the similarity between conditions 17 and 27. This is not coincidental: just as before, employer inference causes misalignment between the private and social returns to investment, and the resulting externality enters social welfare in the same way.

C. Non-Linear Taxation: Imperfectly Observable Investment

My final step is to return to the general model with both unproductive signaling and imperfectly observable investment. For any given set of wage externalities, the equation for the belief externality, **BE** (z), remains very similar to Section 4, and there remain distinct *productivity* and *redistributive* effects.

$$\frac{dw(\tilde{\theta}|\pi)}{d[1-T'(z)]}f(\tilde{\theta}) = \int_K \left(\frac{dx(k|\pi,T)}{d[1-T'(z)]}\right)\Bigg[\overbrace{Q_x(x(k|\pi,T),k)\,f(\tilde{\theta}|x(k|\pi,T))}^{\textbf{Productivity effect}} \tag{28}$$

$$+\underbrace{[Q(x(k|\pi,T),k) - E(q|\tilde{\theta},\pi)]\left(\left.\frac{\partial f(\tilde{\theta}|x)}{\partial x}\right|_{x=x(k|\pi,T)}\right)}_{\textbf{Redistributive effect}}\Bigg]dG(k)$$

However, there are important differences in the interpretation of these two effects. First, the productivity effect may be small or even entirely absent if investment costs are negatively correlated with ability. For example, an extreme possibility is that $q = Q(k)$ so that productivity is unaffected by investment. In this case, the productivity effect is zero and investment returns must come entirely from unproductive signaling of one's ability. The private gain from investment is thus fully offset by negative impacts on the wages of other workers. In this extreme case, the planner would set *higher* rather than lower optimal taxes, given the same mechanical effect and fiscal externality.

A second possibility is that investment costs are positively rather than negatively related to ability, which is possible providing that investment also raises productivity. The redistributive effect then becomes less negative, and may even be positive, since a worker who considers increasing her investment has *higher* productivity than those who invest at that new level in equilibrium. In this case, the "unproductive" component of the return reinforces rather than offsets the positive learning externality, and provides still further motivation to lower marginal tax rates and encourage investment.

7 Conclusion

A substantial body of evidence suggests that employers have imperfect information about the productivity of their workers. This paper provides a framework to study optimal income taxation in this environment. In the model I develop, employers observe an imperfect signal of workers' human capital investments. I show how moral hazard caused by Bayesian inference introduces an externality: workers who invest more raise their own wage but also affect employers' perceptions – and thus the wages – of other workers.

My quantitative results suggest that this new externality is of first-order importance. Taking it into account leads to marginal tax rates that are substantially lower on average. This downward adjustment to tax rates is concentrated at intermediate incomes, leading to an amplification of the classic "U" shape of the optimal tax schedule. There is a notable welfare gain from moving to optimal taxation.

My model provides a framework that could be extended to analyze the implications of many other features of the labor market. This could include asymmetric employer learning (Acemoglu and Pischke 1998), screening by employers (e.g., Stantcheva 2014), an extensive margin of labor supply (Saez 2002), and richer labor market structures including tournaments or other dynamic contracts (Lazear and Rosen 1981, Prendergast 1993). These extensions would preserve the conclusion that wages or utility are compressed, lowering the private return to investment relative to the social return, but they will also lead to other insights. Some (e.g., asymmetric learning) will also feature multiple equilibria, which I provide a way of dealing with in an optimal taxation framework.

More broadly, the core insight of this paper is general: inference based on imperfect information disconnects the private and social returns to engaging in positive behavior. For example, police officers interpret the actions of suspects based on their experiences with previous individuals; thus, compliance by one individual may reduce the likelihood that an officer uses force against a similar suspect in the future. Likewise, buyers form beliefs about the imperfectly observable qualities of goods and services based on past purchases; investment in quality by one seller may therefore raise a consumer's willingness to pay for other similar products. In the future, the approach of this paper may be expanded to provide new insights into these contexts and many others.

References

Aakvik, Arild, Kjell G. Salvanes, and Kjell Vaage, "Measuring Heterogeneity in the Returns to Education using an Education Reform," *European Economic Review*, 2010, *54* (4), 483–500.

Abramitzky, Ran and Victor Lavy, "How Responsive is Investment in Schooling to Changes in Returns? Evidence from an Unusual Pay Reform in Israel's Kibbutzim," *Econometrica*, 2010, *82* (4), 1241–1272.

Acemoglu, Daron and Jorn-Steffen Pischke, "Why Do Firms Train? Theory and Evidence," *The Quarterly Journal of Economics*, 1998, *113* (1), 79–119.

Agan, Amanda and Sonja Starr, "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

Aigner, Dennis J. and Glen G. Cain, "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 1977, *30* (2), 175–187.

Akerlof, George A., "The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *American Economic Review*, 1978, *68* (1), 8–19.

Albanesi, Stefania and Christopher Sleet, "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 2006, *733* (1), 1–30.

Altonji, Joseph G. and C. Pierret, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 2001, *116* (1), 313–50.

Andersson, Fredrik, "Income Taxation and Job-market Signaling," *Journal of Public Economics*, 1996, *59* (2), 277–298.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo, "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2010, *2* (4), 76–104.

Arrow, Kenneth, "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, New Jersey: Princeton University Press, 1973, pp. 3–33.

Baker, George P., Michael Gibbs, and Bengt Holmstrom, "The Internal Economics of the Firm-Evidence From Personnel Data," *The Quarterly Journal of Economics*, 1994, *109* (4), 881–919.

Bedard, Kelly, "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy*, 2001, *109* (4), 749–775.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Best, Michael Carlos and Henrik Jacobsen Kleven, "Optimal Income Taxation with Career Effects of Work," *Working Paper*, 2013.

Blair, Peter Q and Bobby W Chung, "Job Market Signaling Through Occupational Licensing," *NBER Working Paper No. 24791*, 2018.

Blomquist, Sören and Håkan Selin, "Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates," *Journal of Public Economics*, 2010, *94* (11-12), 878–889.

Boháček, Radim and Marek Kapička, "Optimal Human Capital Policies," *Journal of Monetary Economics*, 2008, *55* (1), 1–16.

Bovenberg, A. Lans and Bas Jacobs, "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 2005, *89* (11-12), 2005–2035.

Champernowne, D. G., "The Graduation of Income Distributions," *Econometrica*, 1952, *20* (4), 591–615.

Chetty, Raj, "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (3), 969–1018.

Coate, Stephen and Glenn C. Loury, "Antidiscrimination Enforcement and the Problem of Patronization," *American Economic Review*, 1993, *83* (2), 92–98.

Craig, Ashley C. and Roland G. Fryer, "Complementary Bias: A Model of Two-Sided Statistical Discrimination," *NBER Working Paper No. 23811*, 2017.

da Costa, Carlos E. and Lucas J. Maestri, "The Risk Properties of Human Capital and the Design of Government Policies," *European Economic Review*, 2007, *51* (3), 695–713.

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

Doleac, Jennifer L. and Benjamin Hansen, "The Unintended Consequences of 'Ban the Box': Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden," *SSRN Working Paper*, 2016.

Fang, Hanming and Andrea Moro, "Theories of Statistical Discrimination and Affirmative Action: A Survey," in Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds., *Handbook of Social Economics*, Elsevier, 2011, chapter 5, pp. 133–200.

Farber, Henry S and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

Farhi, Emmanuel and Iván Werning, "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 2013, *80* (2), 596–635.

Feldstein, Martin, "Tax Avoidance and the Deadweight Loss of the Income Tax," *Review of Economics and Statistics*, 1999, *81* (4), 674–680.

Findeisen, Sebastian and Dominik Sachs, "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Journal of Public Economics*, 2016, *138*, 1–21.

Fryer, Roland G., "Belief Flipping in a Dynamic Model of Statistical Discrimination," *Journal of Public Economics*, 2007, *91* (5-6), 1151–1166.

—— **and Glenn C. Loury**, "Valuing Diversity," *Journal of Political Economy*, 2013, *121* (4), 747–774.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski, "Redistribution and Social Insurance," *American Economic Review*, 2016, *106* (2), 359–386.

Gomes, Renato, Jean Marie Lozachmeur, and Alessandro Pavan, "Differential Taxation and Occupational Choice," *Review of Economic Studies*, 2018, *85* (1), 511–557.

Hansen, Karsten T., James J. Heckman, and Kathleen J. Mullen, "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 2004, *121* (1-2), 39–98.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Non-cognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), 411–482.

——, **Lance J. Lochner, and Petra E. Todd**, "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in Eric A. Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Elsevier, 2006, chapter 7, pp. 307–458.

Hedlund, Aaron, "Estate Taxation and Human Capital with Information Externalities," *Macroeconomic Dynamics*, 2018, pp. 1–33.

Hendren, Nathaniel, "The Policy Elasticity," *NBER Working Paper No. 19177*, 2016.

Jacobs, Bas, "Optimal Income Taxation with Endogenous Human Capital," *Journal of Public Economic Theory*, 2005, *7* (2), 295–315.

——, "Optimal Tax and Education Policies and Investments in Human Capital," in Joop Hartog and Henriëtte Maassen van den Brink, eds., *Human Capital: Moving the Frontier*, Cambridge: Cambridge University Press, 2007, pp. 212–232.

Jensen, Robert T., "The (Perceived) Returns to Education and the Demand for Schooling," *Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

Kahn, Lisa B. and Fabian Lange, "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *Review of Economic Studies*, 2014, *81* (4), 1575–1613.

Kapička, Marek, "Optimal Mirrleesean Taxation in a Ben-Porath Economy," *American Economic Journal: Macroeconomics*, 2015, *7* (2), 219–248.

Kaplow, Louis, "Optimal Income Transfers," *International Tax and Public Finance*, 2007, *14* (3), 295–325.

Kline, Patrick and Enrico Moretti, "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 2014, *129* (1), 275–331.

Krueger, Dirk and Alexander Ludwig, "On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium," *Journal of Monetary Economics*, 2016, *77*, 72–98.

Kuka, Elira, Na'ama Shenhav, and Kevin Shih, "Do Human Capital Decisions Respond to the Returns to Education? Evidence from DACA," *NBER Working Paper No. 24315*, 2018.

Lang, Kevin and David Kropp, "Human Capital Versus Sorting: The Effects of Compulsory Attendance Laws," *The Quarterly Journal of Economics*, 1986, *101* (3), 609–624.

____ **and Michael Manove**, "Education and Labor-market Discrimination," *American Economic Review*, 2006, *101* (4), 1467–1496.

Lange, Fabian, "The Speed of Employer Learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

____ **and Robert Topel**, "The Social Value of Education and Human Capital," in Eric A. Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 1, Elsevier, 2006, chapter 8, pp. 459–509.

Lazear, Edward P. and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

Lehmann, Etienne, Laurent Simula, and Alain Trannoy, "Tax Me if You Can! Optimal Nonlinear Income Tax Between Competing Governments," *Quarterly Journal of Economics*, 2014, *129* (4), 1995–2030.

Lindqvist, Erik and Roine Westman, "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 2011, *3* (January), 101–128.

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl, "Taxation and the Allocation of Talent," *Journal of Political Economy*, 2017, *125* (5), 1–49.

Lundberg, Shelly J. and Richard Startz, "Private Discrimination and Social Intervention in Competitive Labor Market," *American Economic Review*, 1983, *73* (3), 340–347.

MacLeod, W. Bentley, Evan Riehl, Juan E. Saavedra, and Miguel Urquiola, "The Big Sort: College Reputation and Labor Market Outcomes," *American Economic Journal: Applied Economics*, 2017, *9* (3), 223–261.

Mailath, George J. and Ernst Ludwig von Thadden, "Incentive Compatibility and Differentiability: New Results and Classic Applications," *Journal of Economic Theory*, 2013, *148* (5), 1841–1861.

Makris, Miltiadis and Alessandro Pavan, "Taxation under Learning-by-Doing," *Working Paper*, 2017.

Maldonado, Darío, "Education Policies and Optimal Taxation," *International Tax and Public Finance*, 2008, *15* (2), 131–143.

Mankiw, N. Gregory and Matthew Weinzierl, "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2010, *2* (1), 155–176.

———, ———, and Danny Yagan, "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 2009, *23* (4), 147–174.

Meghir, Costas and Marten Palme, "Educational Reform, Ability, and Family Background," *American Economic Review*, 2005, *95* (1), 414–424.

Mertens, Karel and José L. M. Montiel-Olea, "Marginal Tax Rates and Income: New Time Series Evidence," *Quarterly Journal of Economics*, 2018, *4* (1), 1803–1884.

Milgrom, Paul and Sharon Oster, "Job Discrimination, Market Forces, and the Invisibility Hypothesis," *Quarterly Journal of Economics*, 1987, *102* (3), 453–476.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Moretti, Enrico, "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-sectional Data," *Journal of Econometrics*, 2004, *121* (1-2), 175–212.

Moro, Andrea and Peter Norman, "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory*, 2004, *114* (1), 1–30.

Neal, Derek and William R. Johnson, "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 1996, *104* (5), 869–895.

Oreopoulos, Philip, "The Compelling Effects of Compulsory Schooling: Evidence from Canada," *Canadian Journal of Economics*, 2006, *39* (1), 22–52.

Pallais, Amanda, "Inefficient Hiring in Entry-level Labor Markets," *American Economic Review*, 2014, *104* (11), 3565–3599.

——— and Emily Glassberg Sands, "Why the Referential Treatment? Evidence from a Field Experiment on Referrals," *Journal of Political Economy*, 2016, *124* (6), 1793–1828.

Phelps, Edmund S, "The Statistical theory of Racism and Sexism," *American Economic Review*, 1972, *62* (4), 659–661.

Piketty, Thomas and Emmanuel Saez, "Optimal Labor Income Taxation," in Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, eds., *Handbook of Public Economics*, Elsevier, 2013, chapter 7, pp. 391–474.

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

Pinkston, Joshua C., "A Test of Screening Discrimination with Employer Learning," *Industrial and Labor Relations Review*, 2006, *59* (2), 267–284.

Prendergast, Canice, "The Role of Promotion in Inducing Specific Human Capital Acquisition," *The Quarterly Journal of Economics*, 1993, *108* (2), 523–534.

Rothschild, Casey and Florian Scheuer, "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 2013, *128* (2), 623–668.

____ and ____, "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, 2016, *83* (1), 1225–1262.

Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin, "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *SSRN Working Paper*, 2016.

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

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

Sarsons, Heather, "Interpreting Signals in the Labor Market: Evidence from Medical Referrals," *Working Paper*, 2018.

Shoag, Daniel and Robert Clifford, "'No More Credit Score': Employer Credit Check Bans and Signal Substitution," *SSRN Working Paper*, 2016.

Simula, Laurent and Alain Trannoy, "Optimal Income Tax Under the Threat of Migration by Top-Income Earners," *Journal of Public Economics*, 2010.

Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

———, "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution," *Journal of Economic Theory*, 1974, *7* (3), 296–332.

Stantcheva, Stefanie, "Optimal Income Taxation with Adverse Selection in the Labour Market," *Review of Economic Studies*, 2014, *81* (3), 1296–1329.

———, "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, 2017, *125* (6), 1931–1990.

Stanton, Christopher T. and Catherine Thomas, "Landing the First Job: The Value of Intermediaries in Online Hiring," *Review of Economic Studies*, 2016, *83* (2), 810–854.

Stiglitz, Joseph E., "Self-selection and Pareto Efficient Taxation," *Journal of Public Economics*, 1982, *17* (2), 213–240.

Trostel, Philip A., "The Effect of Taxation on Human Capital," *Journal of Political Economy*, 1993, *101* (2), 327–350.

Wozniak, Abigail, "Discrimination and the Effects of Drug Testing on Black Employment," *Review of Economics and Statistics*, 2015, *97* (3), 548–566.

A Generalized Contracts

The model outlined in Section 2 assumes that employers offer a wage to workers, as opposed to offering a general contract that specifies both a wage and labor supply. In this appendix, I show that this is not restrictive. To do so, I adopt all the assumptions of the baseline model except that I allow each employer to offer a contract $C_j = \{z_j, l_j\} \in \mathbb{R}_+ \times \mathbb{R}_+ = \mathbb{C}$ to the worker. Each contract specifies a salary $z_j \in \mathbb{R}_+$ and a quantity of labor $l_j \in \mathbb{R}_+$, which jointly imply a price per unit (wage) $w_j = z_j/l_j$. As before, the worker accepts her preferred offer, supplies labor and consumes $c = z - T(z)$.

The worker's strategy is now a set of two functions – an investment decision and an acceptance rule – which can be written as: $x : K \times \mathcal{T} \to \mathbb{R}_+$; and $A : K \times \mathcal{T} \times \Theta \times \mathbb{C}^{|J|} \to J$. Each employer's strategy is a a function that maps signals and tax systems to *contract* offers $O_j : \Theta \times \mathcal{T} \to \mathbb{C}$. Despite the increased complexity, it remains true that every firm earns zero expected profit. Moreover, contracts can always be equivalently characterized as an offer of a wage $w_j = w(\theta|\pi)$ equal to the worker's expected marginal product given the signal θ, with the worker freely choosing how much labor to supply. In this sense, nothing substantive is changed from the baseline model.

Lemma 3. Fix a realized value of θ and assume that $E[q|\theta,\pi]$ is strictly positive and finite given equilibrium beliefs $\pi(q)$. In any pure-strategy equilibrium: all firms $j \in J$ earn zero expected profit; the wage $w_j = z_j/l_j$ implied by every contract offered to the worker is equal to her expected marginal product $E[q|\theta,\pi]$; and the worker's labor supply l_j satisfies $l_j \in L_j^* = \operatorname{argmax}_{\tilde{l}_{j\in\mathbb{R}_+}} u\left(w_j\tilde{l}_j - T\left(w_j\tilde{l}_j\right), \tilde{l}_j\right)$.

B Continuity and Stability

A. Continuity of Investment Responses

In this appendix, I discuss conditions under which equilibrium indeterminacy is avoided, and a given equilibrium shifts continuously in response to the perturbations that I consider. I assume throughout that there is a finite number of cost types. Let $i = 1, \ldots, |K|$ index these types, let $\boldsymbol{x}$ be the vector of investment decisions, and define $q_i = Q(x_i)$.

For each i, Assumption 3 ensures that the following binding first-order condition characterizes the optimal investment decision.

$$f_i(\boldsymbol{x}, T) = Q'(x_i) \int_\Theta v(\theta|\pi, T) \left. \frac{f(\theta|q)}{\partial q} \right|_{q=q_i} d\theta - k_i = 0 \tag{29}$$

Differentiating $f_i(\boldsymbol{x}, T)$ with respect to x_j, we obtain the effect of higher investment by type j on the investment returns of type i. There are two cases:

$$\frac{\partial f_i}{\partial x_j}(\boldsymbol{x}) = \begin{cases} f_{ii}^q + f_{ij}^w & \text{if } i = j \\ f_{ij}^w & \text{if } i \neq j \end{cases} \tag{30}$$

where f_{ii}^q is type k's second-order condition, and f_{ij}^w is the effect via employer beliefs.

$$f_{ii}^q = Q''(x_i) \int_\Theta v(\theta|\pi, T) \left.\frac{\partial f(\theta|q)}{\partial q}\right|_{q=q_i} d\theta + Q'(x_i)^2 \int_\Theta v(\theta|\pi, T) \left.\frac{\partial^2 f(\theta|q)}{\partial q^2}\right|_{q=q_i} d\theta \tag{31}$$

$$f_{ij}^w = Q'(x_j) \int_\Theta u_c(\theta) \left[1 - T'(z(\theta|\pi, T))\right] l(\theta|\pi, T) \frac{\partial w(\theta|\pi)}{\partial q_j} f(\theta|q_i)\, d\theta \tag{32}$$

Letting $p(k_j)$ be the probability of drawing type k_j, the equation for $\frac{\partial w(\theta|\pi)}{\partial q_j}$ is as follows.

$$\frac{\partial w(\theta|\pi)}{\partial q_j} = \left(f(\theta|q_j) + [q_j - w(\theta|\pi)] \left.\frac{\partial f(\theta|q)}{\partial q}\right|_{q=q_j} \right) p(k_j) \tag{33}$$

The partial derivatives (equation 30) can be arranged to form the Jacobian $\mathcal{J}_{f,\boldsymbol{x}}$.

$$\mathcal{J}_{f,\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_{|K|}}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{|K|}}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_{|K|}}{\partial x_{|K|}}(\boldsymbol{x}) \end{bmatrix} \tag{34}$$

Next, let $dc(\theta|\pi, T) = -dT(z(\theta|\pi, T))$ be the Fréchet derivative with respect to T of consumption by a worker with signal θ. The Fréchet derivative of $v(\theta|\pi, T)$ is then:

$$dv(\theta|\pi, T) = u'(z(\theta|\pi, T) - T(z(\theta|\pi, T))) \times dc(\theta|\pi, T)$$

And in turn, the Fréchet derivative of $f_i(\boldsymbol{x}, T)$ is given by $df_i(\boldsymbol{x}, T)$.

$$df_i(\boldsymbol{x}, T) = Q'(x_i) \int_\Theta dv(\theta|\pi, T) \left.\frac{f(\theta|q)}{\partial q}\right|_{q=q_i} d\theta \tag{35}$$

These derivatives can be stacked into a $|K| \times 1$ vector $df(\boldsymbol{x}, T)$.

Providing that $\mathcal{J}_{f,\boldsymbol{x}}$ invertible, the Implicit Function Theorem implies that there is a neighborhood around $\boldsymbol{x}$ and T in which there is a unique Fréchet differentiable function mapping T to $\boldsymbol{x}$, and the response of investments is given by $-\mathcal{J}_{f,\boldsymbol{x}}^{-1} \times df(\boldsymbol{x}, T)$. As I argue below, invertibility of $\mathcal{J}_{f,\boldsymbol{x}}$ is the generic case.

B. INVERTIBILITY OF $\mathcal{J}_{f,\boldsymbol{x}}$

I next show that, if $\mathcal{J}_{f,\boldsymbol{x}}$ is not invertible, it can be rendered invertible by an arbitrarily small perturbation to the investment technology $Q(x)$, which preserves both the key properties of that technology and the existing equilibrium. Moreover, starting with any equilibrium in which $\mathcal{J}_{f,\boldsymbol{x}}$ is invertible, this clearly remains the case after a similarly small perturbation. In these two senses, invertibility of $\mathcal{J}_{f,\boldsymbol{x}}$ is generic.

First, I construct a parameterized family of functions, $\tilde{Q}(x|\boldsymbol{c})$, where $\boldsymbol{c}$ is a vector of strictly negative real numbers $c_1, \ldots c_{|K|}$. Each function in this family retains the key properties of $Q(x)$, but the second derivative of $\tilde{Q}(x|\boldsymbol{c})$ evaluated at x_j is c_j.

1. Take each x_j and define a narrow domain $x_j \pm r$ where $r > 0$ is arbitrarily small. On this domain, define a function $B_j(x|c_j) = Q(x_j) + Q'(x_j)(x - x_j) + \frac{1}{2}c_j(x - x_j)^2$. $B_j(x|c_j)$ has the same level and derivative as $Q(x)$ at x_j, but $B_j''(x_j|c_j) = c_j$.

2. Link the functions $B_j(x|c_j)$ to form any twice-differentiable function $\tilde{Q}(x|\boldsymbol{c})$ with $\tilde{Q}(0|\boldsymbol{c}) = 0$, $\tilde{Q}'(x|\boldsymbol{c}) > 0$, $\tilde{Q}''(x|\boldsymbol{c}) > 0$ and $\lim_{x\to 0}\tilde{Q}'(x|\boldsymbol{c}) = \infty$. This is always possible, since r is small and Q strictly concave.

3. Let $\hat{Q}(x|\boldsymbol{c},\alpha) = \alpha\tilde{Q}(x|\boldsymbol{c}) + (1-\alpha)Q(x)$ with $\alpha \in (0,1)$.

Next, I replace $Q(x)$ with $\hat{Q}(x|\boldsymbol{c},\alpha)$ in the economy described in Section 2. For any $\boldsymbol{c}$, there remains an equilibrium with the same investment decisions. However, the diagonal elements of the Jacobian $\mathcal{J}_{f,\boldsymbol{x}}$ are replaced by:

$$f_{ii}^q = c_i \int_\Theta v(\theta|\pi,T) \left.\frac{\partial f(\theta|q)}{\partial q}\right|_{q=Q(x_i)} d\theta + Q'(x_i)^2 \int_\Theta v(\theta|\pi,T) \left.\frac{\partial^2 f(\theta|q)}{\partial q^2}\right|_{q=Q(x_i)} d\theta.$$

Moreover, f_{ii}^q scales with c_i since $\int_\Theta v(\theta|\pi,T) \left.\frac{\partial f(\theta|q)}{\partial q}\right|_{q=q_i} d\theta > 0$. Non-diagonal elements of $\mathcal{J}_{f,\boldsymbol{x}}$ are unchanged.

Finally, let $c_j = Q''(x_j) + \varepsilon_j < 0$ where ε_j are distinct real numbers with $\varepsilon_j < -Q''(x_j)$. For small enough α, $\hat{Q}(x|\boldsymbol{c},\alpha)$ is an arbitrarily close approximation of $Q(x)$. However, the Jacobian $\mathcal{J}_{f,\boldsymbol{x}}$ of the new economy is invertible. Specifically, any two rows that were collinear are no longer collinear; and, since α, is small, no two rows are newly collinear.

C. STABILITY OF EQUILIBRIA

Restricting the set of equilibria to those that are stable is one way to ensure that the economy does not switch equilibria in response to a perturbation such as that described in

Section 5. To define such a notion of stability, suppose that the economy evolves according to the following backward-looking dynamic adjustment process:

$$x_{k,t+1} \in X^*_{k,t+1} = \underset{\tilde{x} \in \mathbb{R}_+}{\operatorname{argmax}} \quad \int_\Theta v\left(\theta|\pi_t, T\right) f\left(\theta|Q\left(\tilde{x}\right)\right) d\theta - k\tilde{x} \tag{36}$$

where:

$$\begin{aligned} v\left(\theta|\pi_t, T\right) &= w\left(\theta|\pi_t, T\right) l\left(\theta|\pi_t, T\right) \\ l\left(\theta|\pi_t, T\right) \in L^*\left(\theta|\pi_t, T\right) &= \underset{\tilde{l}_j \in \mathbb{R}_+}{\operatorname{argmax}} \; u\left(w\left(\theta|\pi_t\right)\tilde{l} - T\left(w\left(\theta|\pi_t\right)\tilde{l}\right), \tilde{l}\right) \\ w\left(\theta|\pi_t\right) &= \frac{\int_K Q(x_{k,t}) f\left(\theta|Q(x_{k,t})\right) dG\left(k\right)}{\int_K f\left(\theta|Q(x_{k,t})\right) dG\left(k\right)} \end{aligned}$$

In general, this does not necessarily define a unique path for the economy. However, Assumptions 1 to 3 ensure that this is true *locally* because both $x_{k,t+1}$ and $l\left(\theta|\pi_t, T\right)$ are both uniquely pinned down and vary continuously with other agents' investment decisions.

Thus, letting $\overline{\boldsymbol{x}}(T)$ be a set of equilibrium investment decisions, the dynamic adjustment process above can be approximated locally around $\overline{\boldsymbol{x}}(T)$ by a first-order linear system $\boldsymbol{x}_{t+1}(T) = \boldsymbol{A}\boldsymbol{x}_t(T)$. If all the eigenvalues of the matrix $\boldsymbol{A}$ have moduli strictly less than one, then the equilibrium is locally asymptotically stable. Providing that $\mathcal{J}_{f,\boldsymbol{x}}$ is invertible (see part A above) so that there is a locally unique Fréchet differentiable function mapping T to $\boldsymbol{x}$, local asymptotic stability in turn ensures that the economy does not switch equilibria in response to a small change in the tax schedule.

C Beyond the First Order Approach

Proposition 3 provides the derivative of social welfare with respect to a perturbation in the tax schedule, providing that there is a locally continuous selection around the initial point, $(E(T), T)$. I adopted assumptions that ensure this is true for an arbitrary tax system. The proposition also states a condition that holds at an optimum, providing that the planner does not systematically locate at a point where the regularity conditions break down.

In this appendix, I discuss complications that arise when the planner does in fact have a reason to locate at a discontinuity, in which case the derivatives in Proposition 3 are not defined. I also discuss reasons why the planner's first-order condition is not *sufficient* for optimality. For expositional clarity, I focus on a particularly simple case of the general model, in which the planner is restricted to a linear tax, labor supply is perfectly inelastic, and investment decisions are binary.[41] This greatly simplifies the analysis of this subset of issues, while providing insights that are conceptually general.

A. Special Case of the Model with Binary Investment

In this special case of the model, investment is dichotomous. A worker decides to become qualified (q) at cost k, or remain unqualified (u) at no cost. A qualified worker who is hired produces a fixed payoff $\omega > 0$ for the firm who hires her, while an unqualified worker produces zero. As before, the cost distribution $G(k)$ is the probability that a worker has investment cost no greater than k; here, I additionally assume that $G(0) = 0$ and that $G(k)$ is continuously differentiable, with density $g(k)$.

With binary investment, an employer's prior belief is summarized by the fraction of workers it believes have invested. In addition, employers see a common non-contractible signal $\theta \in [0, 1]$, which in this case has CDF $F_i(\theta)$ and PDF $f_i(\theta)$ where $i \in \{q, u\}$ and $f_u(\theta) / f_q(\theta)$ is strictly decreasing in θ. In equilibrium, firms' prior beliefs coincide with the true equilibrium probability π that a worker invests; and each firm offers to pay the worker a wage $w(\theta|\pi)$ equal to her expected marginal product.

$$w(\theta|\pi) = \omega \times \frac{\pi f_q(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)}$$

The worker accepts her best offer, supplies a unit of labor and receives that wage. If she invested, she obtains utility $v(\theta|\pi, \tau) - k = u((1-\tau) w(\theta|\pi) + \tau \overline{w}) - k$, where τ is a linear income tax, and $\overline{w} = \pi\omega$ is the average wage. If she did not invest, she receives $v(\theta|\pi, \tau) = u((1-\tau) w(\theta|\pi) + \tau \overline{w})$. I assume that $u(c)$ is strictly increasing, strictly concave and satisfies Inada conditions: $\lim_{c \to 0} u'(c) = \infty$ and $\lim_{c \to \infty} u'(c) = 0$.

Integrating over θ, the expected utilities for an investor ($\overline{v}_q$) and non-investor ($\overline{v}_u$) are given by equations 37 and 38.

$$\overline{v}_q(\pi|\tau) = \int_0^1 v(\theta|\pi, \tau)\, dF_q(\theta) - k \tag{37}$$

$$\overline{v}_u(\pi|\tau) = \int_0^1 v(\theta|\pi, \tau)\, dF_u(\theta) \tag{38}$$

Since workers invest if their expected return is greater than their cost, this implies an investment rate of $G(\beta(\pi|\tau))$ where $\beta(\pi|\tau) = \overline{v}_q(\pi|\tau) - \overline{v}_u(\pi|\tau)$.

The final requirement of equilibrium is that workers invest at a rate that coincides with employers' beliefs. This is embodied in equation 39, which states that the fraction of

[41] The model with binary investment is similar to Moro and Norman (2004).

investors must be equal to the fraction of workers that employers believe are qualified.

$$\pi = G\left(\beta\left(\pi|\tau\right)\right) \tag{39}$$

For a given tax rate τ, equation 39 defines a fixed point as shown in Figure 10. An employer belief π, combined with the tax τ, pins down the investment return and an investment rate, $G\left(\beta\left(\pi|\tau\right)\right)$.

FIGURE 10: EQUILIBRIA AND TAXATION

(a) Determination of Equilibria τ (b) Equililibrium Set for Each Value of τ

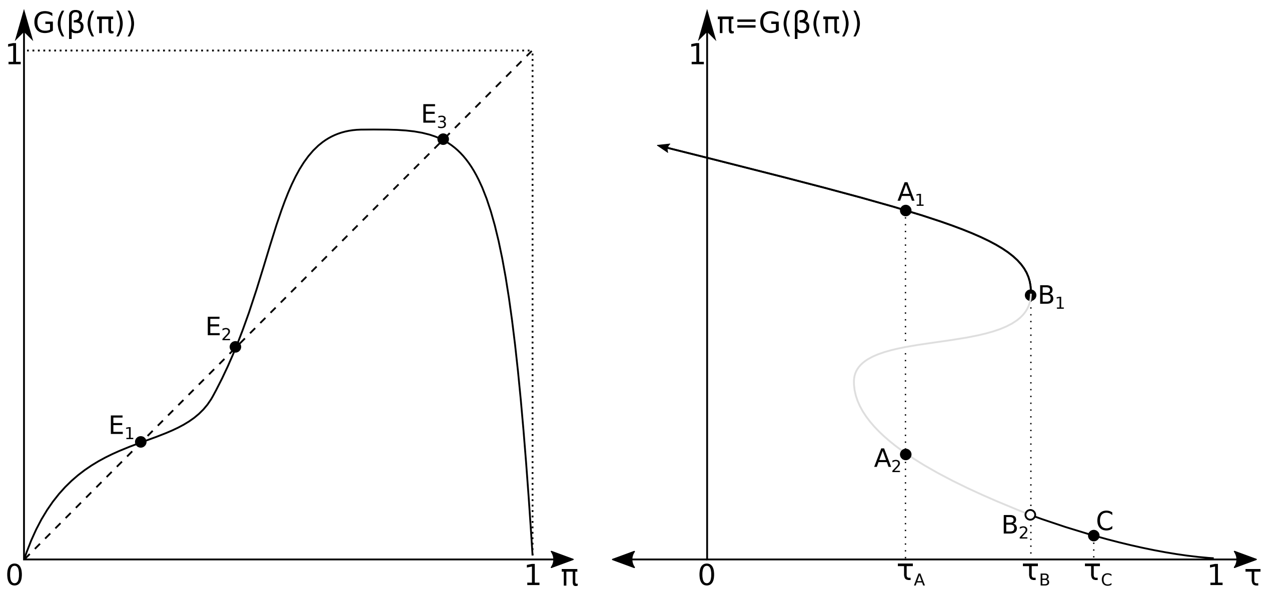


Figure notes. This figure shows an example economy with binary investment. In panel (a), the aggregate rate of investment implied by worker and firm optimization, $G\left(\beta\left(\pi\right)\right)$, is plotted against the employer prior, π. Any intersection between this line and the 45 degree line is an equilibrium. The arrows show the direction in which each equilibrium moves as τ rises. Panel (b) shows the set of equilibria over a range of values of τ. Pareto dominant equilibria are shown by the black line segments.

Any point on the 45 degree line constitutes an equilibrium, since employers' beliefs are confirmed. At the extremes, either $\pi = 0$ or $\pi = 1$ ensure that there is no return to investment, since employers who are certain of a worker's decision place no weight on the signal. There is thus always an equilibrium in which no workers invest, and all workers receive a zero wage. Proposition 6 provides sufficient conditions for there to be others. For example, the economy in Figure 10 has four equilibria: 0, E_1, E_2 and E_3.

Proposition 6. *Assume that $\phi\left(\theta\right) = f_u\left(\theta\right) / f_q\left(\theta\right)$ is continuous and strictly positive on $[0, 1]$. If there exists π such that $G\left(\beta\left(\pi|\tau\right)\right) > \pi$ then there are multiple solutions to condition 39.*

Intuitively, these conditions are satisfied if the returns to investment are high enough, as ensured by a large value of ω or a low enough tax rate. In turn, this means there is some employer belief π such that the fraction of investors given that belief, $G(\beta(\pi|\tau))$, is *higher* than π. Since $G(\beta(1|\tau)) = 0$, and the regularity assumptions ensure that $G(\beta(\pi|\tau))$ is continuous π, this guarantees that there is a belief $\pi^* > 0$ such that $\pi^* = G(\beta(\pi^*|\tau))$.

B. Optimal Taxation with Binary Investment

Tax policy can be analyzed in the same way as in the general model. Raising the linear tax τ causes $G(\beta(\pi|\tau))$ to shift down for every employer belief π. As a result, the location of an equilibrium falls if $G(\beta(\pi|\tau))$ crosses the 45 degree line from above, and rises if it crosses from below, as shown in panel (b) of Figure 10.

For simplicity, I assume that agents play the planner's preferred equilibrium, which ensures that investment and welfare always increase as τ is lowered.[42] The arguments that follow do not depend on this assumption. However, it provides a concrete equilibrium selection criterion that is especially compelling here because equilibria for a given tax rate are Pareto-ranked, with higher investment corresponding to higher welfare. In Figure 10, the black line traces out the Pareto-dominant equilibria.

Proposition 7. *Assume that multiple values of π satisfy equation 39 for a given tax rate τ. Let π_i and π_j be two solutions. Welfare is higher for every worker under π_i than π_j iff $\pi_i > \pi_j$. Moreover, investment in the planner's preferred equilibrium increases as τ is lowered.*

Next, to characterize optimal taxation, define ε_z as the elasticity of average income with respect to the retention rate. Second, let u'_θ be the marginal utility of consumption of an individual who sends signal θ and therefore receives wage $w(\theta|\pi)$. Finally, let $\tilde{u}'_\theta$ be the same individual's marginal utility relative to the average: i.e., $\tilde{u}'_\theta = u'_\theta / \bar{u}'_\theta$. For simplicity, I assume here that the planner's social welfare function is linear, but additional concavity from the social welfare function does not change the analysis.

Proposition 8 provides a necessary condition for the optimality of τ, in the same form as Propositions 2 and 3. As before, there is a trade-off between redistribution from high-wage to low-wage workers, a fiscal externality and a belief externality. Ignoring the belief externality, an optimal τ at which this condition holds would always be strictly positive.

[42] The set of equilibria can alternatively be refined by requiring stability under a simple adjustment process: $\pi_{t+1} = G(\beta(\pi_t|\tau))$. This amounts to a requirement that the absolute value of the slope of $G(\beta(\pi|\tau))$ is less than one (see Appendix D), which in turn implies that investment must fall when τ rises. In Figure 10, both the zero investment equilibrium and E_2 are ruled out by the stability condition.

However, the belief externality $\overline{w}_z$ provides an efficiency motive for intervention and pushes toward lower – and possibly negative – tax rates.

Proposition 8. *Fix a value of τ and an investment rate $\pi^*(\tau) > 0$, which satisfies equation 39. If $g(\beta(\pi^*(\tau)|\tau))\beta'(\pi^*(\tau)|\tau) \neq 1$ and τ is optimal, then the following condition holds:*

$$\frac{\tau}{1-\tau} = \frac{\overline{v}_\tau - \varepsilon_z \overline{w}_z}{\varepsilon_z} \tag{40}$$

where $\overline{v}_\tau = (1-\pi)\int_0^1 \tilde{u}'_\theta [f_u(\theta) - f_q(\theta)]\, d\theta$, ε_z is the elasticity of income to the retention rate $1-\tau$, and $\overline{w}_z = \frac{1}{\omega}\int_0^1 \tilde{u}'_\theta \frac{\partial w(\theta|\pi)}{\partial \pi}[\pi f_q(\theta) + (1-\pi) f_u(\theta)]\, d\theta$ is the belief externality.

Proposition 8 parallels the results from the linear tax example (Proposition 2) and non-linear taxation (Proposition 3). The requirement that $g(\beta(\pi^*(\tau)|\tau))\beta'(\pi^*(\tau)|\tau) \neq 1$ simply suffices to ensure the investment rate varies continuously with τ at the optimum, which is equivalent to invertibility of the Jacobian, $\mathcal{J}_{f,\boldsymbol{x}}$, discussed in Appendix B.

C. Limitations of the First Order Approach

The model with binary investment provides a transparent and flexible platform to discuss complications that could lead to discontinuity at the optimum or prevent my necessary conditions from being sufficient for optimality. The first caveat is that condition 8 may hold at other points. For example, the planner's optimal tax rate may be A_1 in panel (b) of Figure 10, but the first order condition may also hold at C. This a natural limitation of the first-order approach and is easily resolved by examining a second-order condition.

The second caveat is more interesting: in *some* economies, there may be an incentive for the planner to choose a tax rate that places the economy at a discontinuity. For example, consider again panel (b) of Figure 10. By Proposition 8, we know that B_1 dominates B_2. The complication is that it is possible for social welfare to be increasing in τ as we approach τ_B from below and *also* as we approach τ_B from above, so that τ_B is the optimal tax rate. However, equation 40 does not hold at the discontinuity. This is not a violation of Proposition 8, since $g(\beta(\pi|\tau))\beta'(\pi|\tau) = 1$ at B_1. However, it highlights a conceptually important limitation of the first-order approach in this context.

D Multiple Groups and Self-fulfilling Disparities

A possibility with multiple equilibria is that employers have different beliefs about members of distinct groups (e.g., black and white workers). Although this is ruled out if agents always play the planner's preferred equilibrium and the groups are identical, asymmetric equilibria could well arise in reality. This is the classic case of self-fulfilling statistical discrimination, as analyzed by Arrow (1973), Coate and Loury (1993), and others. In this appendix, I discuss the implications of this for optimal taxation.

My first step is to adapt the model in Appendix C by dividing workers into an *advantaged* (A) group and a *disadvantaged* (D) group. Specifically, I assume that a worker is of type A with probability λ_A and of type D with probability $\lambda_D = 1 - \lambda_A$. The two groups are identical in fundamentals. As in Appendix C, the planner is restricted to linear taxation. However, she can set a different tax rate τ_j for each group $j \in \{A, D\}$, and a lump sum transfer $T_{A\to D}$ from As to Ds. These three variables constitute a tax system T.

Definition. *A tax system T is a triple $(\tau_A, \tau_D, T_{A\to D})$, comprised of a marginal tax rate τ_j for each group combined with an intergroup transfer $T_{A\to D}$.*

Equilibrium in the model with two distinct groups can be characterized as follows. First, net of investment costs, a worker of type j with signal θ receives utility $v_j(\theta|\pi_j, T)$.

$$v_A(\theta|\pi_A, T) = u\left[(1-\tau_A)\,\omega\frac{\pi_A f_q(\theta)}{\pi_A f_q(\theta) + (1-\pi_A) f_u(\theta)} + \tau_A \pi_A \omega - \frac{T_{A\to B}}{\lambda_A}\right]$$
$$v_D(\theta|\pi_D, T) = u\left[(1-\tau_D)\,\omega\frac{\pi_D f_q(\theta)}{\pi_D f_q(\theta) + (1-\pi_D) f_u(\theta)} + \tau_D \pi_D \omega + \frac{T_{A\to D}}{\lambda_D}\right]$$

Thus, a worker's expected utility is $\overline{v}_q^j(\pi_j|T)$ if she invests, and $\overline{v}_u^j(\pi_j|T)$ if she does not.

$$\overline{v}_q^j(\pi_j|T) = \int_0^1 v_A(\theta|\pi_j, T)\, dF_q(\theta) - k \qquad \overline{v}_u^j(\pi_j|T) = \int_0^1 v_B(\theta|\pi_j, T)\, dF_u(\theta)$$

The model remains otherwise unchanged from Appendix C. Workers invest if the return, $\beta_j(\pi_j|T) = \overline{v}_q^j(\pi_j|T) - \overline{v}_u^j(\pi_j|T)$, is greater than their cost, implying an investment rate of $G(\beta_j(\pi_j|T))$. Equilibrium requires that $\pi_j = G(\beta_j(\pi_j|T))$, $j \in \{A, D\}$.

Unlike Appendix C, I do not assume that agents coordinate on the planner's preferred equilibrium. Instead, I follow the approach of Section 4, which applies given any continuous selection of equilibria. Specifically, for any given tax schedule T, let $\boldsymbol{\pi}(T)$ be the set of pairs (π_A, π_D) such that $\pi_j(T) = G(\beta_j(\pi(T)|T))$ for $j \in \{A, D\}$. The correspondence $\boldsymbol{\pi}(T)$ suffices to characterize the set of equilibria for each tax schedule. I define a selection by choosing one equilibrium pair $\pi^\dagger(T)$ for each tax schedule from this set.

Optimal taxation is then similar to the case with one group. The planner values both groups equally, so welfare is the weighted average $W = \lambda_A W_A + \lambda_D W_D$, where:

$$W_j = \pi_j \overline{v}_q^j (\pi_j|T) + (1-\pi_j)\, \overline{v}_q^j (\pi_j|T) - \int_0^{\overline{v}_q^j(\pi_j|T) - \overline{v}_u^j(\pi_j|T)} k dG_j(k).$$

Within each group, the same perturbation arguments apply and the condition required for τ_j to be optimal is unchanged. The only additional complication is the inter-group transfer, which is set so that the average marginal utility is the same for As and Ds.

Proposition 9. *If $\pi^\dagger(T)$ is locally continuous and T is optimal, the following conditions hold.*

$$\frac{\tau_j}{1-\tau_j} = \frac{\overline{v}_{j,\tau} - \varepsilon_z^j \overline{w}_z^j}{\varepsilon_z^j} \tag{41}$$

$$\int_\theta u'_{A,\theta} dF(\theta) = \int_\theta u'_{B,\theta} dF(\theta) \tag{42}$$

where $\overline{v}_{j,\tau} = (1-\pi_j)\int_0^1 \tilde{u}'_{j,\theta}\left[f_u(\theta) - f_q(\theta)\right] d\theta$, ε_z^j is the income elasticity of group j, and $\overline{w}_z^j = \frac{1}{\omega}\int_0^1 \tilde{u}'_{j,\theta} \frac{\partial w(\theta|\pi_j)}{\partial \pi_j}\left[\pi_j f_q(\theta) + (1-\pi_j) f_u(\theta)\right] d(\theta)$ is the belief externality.

To build intuition, consider the case in which $T_{A\to D}$ is constrained to be zero and $\pi^\dagger(T)$ selects equilibria that are symmetric in the sense that $\pi_A = \pi_B$. This is always possible, because the groups are identical. The planner's choice of τ_j is then isomorphic to the model with a single group, so $\tau_A = \tau_B$ and $\pi_A = \pi_B$. Moreover, if condition 41 holds, equation 42 must as well. Starting from equal treatment ($\tau_A = \tau_B$ and $T_{A\to D} = 0$), there is therefore no first order gain from slightly changing the tax system. This implies that the planner would not want to set $T_{A\to D} \neq 0$, even if she could. Intuitively, if the two groups are identical and equilibria are symmetric, there is no motive for the planner to choose a tax system that favors one group over the other.

In general, however, it is possible that $\pi^\dagger(T)$ includes non-symmetric equilibria, which raises the possibility of "self-fulfilling" differences between groups. In this case, even through groups A and D are *ex ante* identical, it is not generally true that $\pi_A = \pi_B$ even at the planner's optimal choice of T. The optimal T may then involve different marginal tax rates for A and B workers, and an inter-group transfer.

Although Proposition 9 still holds in this non-symmetric case, the potential for self-fulfilling asymmetries raises the question of whether there are policies that can eliminate this problem. One possibility is for the planner to set a tax that conditions on the aggregate level of investment, which would always allow the planner to ensure Pareto efficiency. Alternatively, one could imagine a dynamic policy that transitions the economy from one equilibrium to another. For example, one could temporarily implement a very low tax rate and then ratchet it back up, ensuring convergence to a Pareto efficient equilibrium.

E Approximately Optimal Taxation

In this appendix, I derive a way of approximating the optimal tax schedule given only a few measurable statistics. Two principles underlie the approximation. First, I assume that a change in $T'(z)$ primarily causes individuals with income close to z to respond. Second, I assume that the incidence of the externality falls on workers with similar welfare weight, labor supply and tax rate to those with income z.

Letting $l(z)$ be the labor supply at income z, I first define $\tilde{\Omega}\left(z,\tilde{\theta}\right)$ as the difference between the weight on externalities at income $z(\tilde{\theta}|\pi,T)$ and the weight at income z.

$$\tilde{\Omega}\left(z,\tilde{\theta}\right) = \psi_z(z(\tilde{\theta}|\pi,T))\left[1-T'(z(\tilde{\theta}|\pi,T))\right] l(\tilde{\theta}|\pi) - \psi_z(z)\left[1-T'(z)\right] l(z)$$

The belief externality can then be re-written as an approximation, plus a covariance bias.

$$\mathbf{BE}(z) = -d\tau dz\Bigg\{\psi_z(z)\left[1-T'(z)\right] l(z)\left[\int_\Theta \left(\frac{dw(\tilde{\theta}|\pi)}{d\left[1-T'(z)\right]}\right) f(\tilde{\theta})d\theta\right] + \underbrace{\int_\Theta \tilde{\Omega}\left(z,\tilde{\theta}\right)\left(\frac{dw(\tilde{\theta}|\pi)}{d\left[1-T'(z)\right]}\right) f(\tilde{\theta})d\theta\Bigg\}}_{\textbf{Covariance bias}} \tag{43}$$

Next, without loss of generality, I write the externality as a share of the average wage rise.

$$\int_\Theta \frac{dw(\tilde{\theta}|\pi)}{d\left[1-T'(z)\right]} f(\tilde{\theta})d\tilde{\theta} = (1-s(z))\frac{d\overline{w}}{d\left[1-T'(z)\right]} \tag{44}$$

Bringing everything together, condition 17 can then be approximated by:

$$\mathbf{FE}(z) + \mathbf{ME}(z) - (1-s(z))\,\psi_z(z)\,l(z)\left[1-T'(z)\right]\frac{d\overline{w}}{d\left[1-T'(z)\right]} = 0. \tag{45}$$

Figure 11 shows the results when equation 45 is implemented in the simulated economy, using the values of $s(z)$ implied by the simulation. Starting from the naïve benchmark, 60 percent of the gains from optimal taxation are achieved.

FIGURE 11: APPROXIMATELY OPTIMAL TAXATION

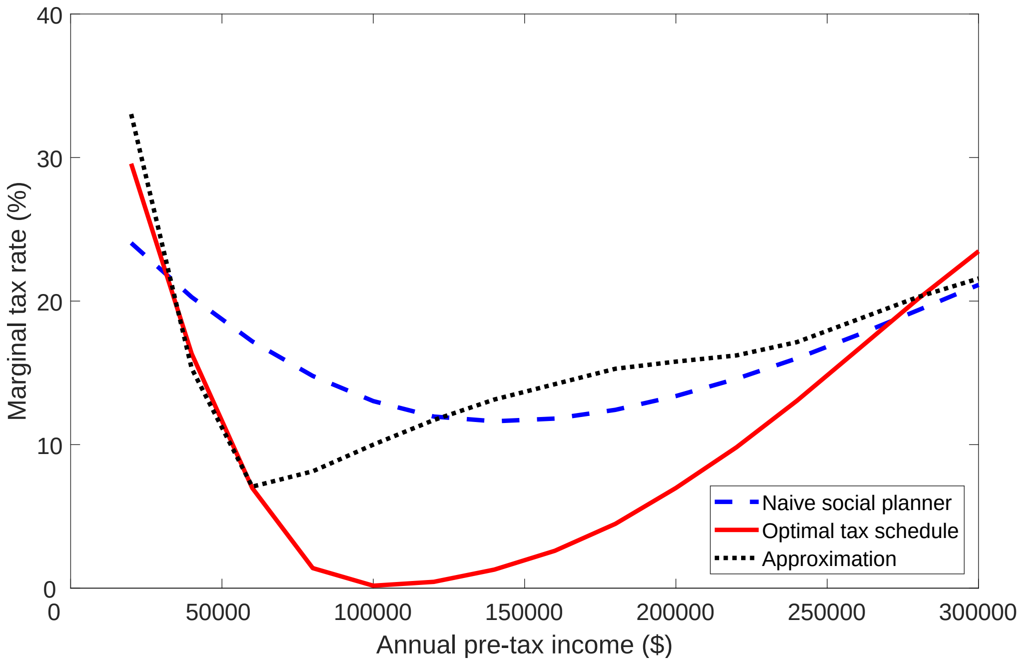


Figure notes. This figure shows the results of the simulation described in Section H. The solid red line shows the optimal tax schedule, the dashed blue line shows the naïve schedule, and the dotted black line shows a schedule what would be accepted by a planner who implemented equation 45.

F Unproductive Signaling with Observable Investment

In this appendix, I provide conditions under which $w(x)$ and $x(k)$ are differentiable. As in Section 4, I assume that problem 24 is strictly concave given a wage $w = w(x)$ so that the labor supply choice can be characterized by a first-order condition (equation 46):

$$wu_c(wl^*(w) - T(wl^*(w)), l^*(w)) \left[1 - T'(wl^*(w))\right] + u_l(wl^*(w) - T(wl^*(w)), l^*(w)) = 0 \tag{46}$$

where $l^*(w) = \text{argmax}_{l \in \mathbb{R}_+} \; u(wl - T(wl), l)$.

Next, I define $\hat{v}(x) = v(w(x)|T)$, and let $x_{FB}(k) = \text{argmax}_x v(Q(x,k)|T) - kx$ be the investment level chosen by an agent with cost k in the equivalent problem with perfect employer information. Using these definitions, I adopt three assumptions regarding problem 23, which can be viewed as restrictions on the investment technology, $Q(x,k)$.

Assumption 4. *The solution to the first best contracting problem, $x_{FB}(k)$, is unique for all k.*

Assumption 5. *For all $k \in K$, $\hat{v}(x)$ is strictly concave around $x_{FB}(k)$.*

Assumption 6. *$\exists \kappa > 0$ such that $\hat{v}''(x) \geq 0 \Rightarrow \hat{v}'(x) > \kappa$ for all $(k,x) \in K \times \mathbb{R}_+$.*

A sufficient condition for assumption 4 to hold is that the first best contracting problem is strictly concave, which is always true given sufficient concavity of the investment technology. Assumption 5 simply states that problem 23 is *locally* strictly concave around the first-best investment choice, while assumption 6 is a *global* equivalent that is weaker than strict concavity but stronger than strict quasi-concavity.

Assumptions 4, 5 and 6 jointly ensure that $x(k)$ is differentiable for all $k \in K$ (see Mailath and von Thadden 2013), which in turn implies that $w(x)$ is differentiable and that the following condition holds for all k:

$$u_c(z(k) - T(z(k)), l(k)) \left[1 - T'(z(k))\right] l(k) w'(x(k)) = k \tag{47}$$

where $l(k) = l^*(w(x(k)))$ and $z(k) = w(x(k)) l(k)$.

G Proofs and Derivations

Proof of Lemma 1. Firm beliefs about the distribution of productivity in the population must be confirmed in equilibrium and identical across firms. Let π denote the equilibrium set of beliefs. Firm j's expectation of the worker's productivity is $E[q|\theta, \pi, A_j = 1] \geq 0$. Next, let $\tilde{u}(w_j) = u(w_j l^*(w_j) - T(w_j l^*(w_j)), l^*(w_j))$ represent the utility that the worker receives from accepting wage w_j and supplying labor optimally.

Suppose that some firm j makes strictly positive expected profits given its wage offer w_j. It must then be the case that $\tilde{u}(w_j) \geq \tilde{u}(w_k)$ for all wages w_k offered by other firms. There are several cases to consider, each of which lead to a contradiction.

Case 1: $\tilde{u}(w_j) > \tilde{u}(w_k)$ for some w_k.

In this case, firm k initially earns zero expected profit, since no workers accept its offer. However, it can offer a wage slightly higher than w_j. It then attracts the worker with probability one and earns strictly positive profits. This is a profitable deviation.

Case 2: $\tilde{u}(w_j) = \tilde{u}(w_k)$ for all w_k, and $\overline{P}_{k,\theta} \leq 0$ for some k.

If any firm makes weakly negative profits, then the same deviation as Case 1 applies.

Case 3: $\tilde{u}(w_j) = \tilde{u}(w_k)$ and $\overline{P}_{k,\theta} > 0$ for all k.

Since the worker always accepts an offer, $E[q|\theta, \pi, A_j = 1]$ is bounded weakly below $E[q|\theta, \pi]$ for at least one firm. This firm's expected profit is bounded below $\overline{P}_{\text{MAX}}$.

$$\overline{P}_{\text{MAX}} = \max_w \left[E[q|\theta, \pi] - w\right] l^*(w) \text{ s.t. } u(wl^*(w) - T(wl^*(w)), l^*(w)) \geq u(T(0), 0)$$

The assumptions on the worker's utility function ensure that this yields finite labor supply for any finite $E[q|\theta,\pi]$. Since w_j is greater than zero and $E[q|\theta,\pi]$ is finite, $\overline{P}_{\text{MAX}}$ is also bounded. Finally, this firm can strictly increase its profit by raising w_j slightly and attracting the worker with probability one.

Since every case in which a firm makes a strictly positive expected profit implies a profitable deviation, and all firms can obtain zero expected profit by offering a zero wage, it must be true that every firm makes zero expected profit. Finally, the wage, w, must be the same at every firm who hires the worker with positive probability. We have therefore established that $[E[q|\theta,\pi]-w]\,l^*(w)=0$, which is only satisfied if $w=E[q|\theta,\pi]$. □

Proof of Proposition 1. Assume – subject to verification – that investment is distributed log-normally as hypothesized.

$$\ln q_i \sim N\left(\ln\mu_q - \frac{\sigma_q^2}{2}, \sigma_q^2\right)$$

Given this, employers face a log-normal signal extraction problem. The expectation of log-productivity is as follows.

$$\begin{aligned} E[\ln q|\theta] &= \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_\xi^2}\right)\ln\theta + \left(\frac{\sigma_\xi^2}{\sigma_q^2+\sigma_\xi^2}\right)\left(\ln\mu_q - \frac{\sigma_q^2}{2}\right) \\ &= \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_\xi^2}\right)\ln q + \left(\frac{\sigma_\xi^2}{\sigma_q^2+\sigma_\xi^2}\right)\left(\ln\mu_q - \frac{\sigma_q^2}{2}\right) + \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_\xi^2}\right)\ln\xi \end{aligned}$$

Since employers offer workers their expected marginal product, the after-tax wage is:

$$\ln[(1-\tau)\,w] = \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_\xi^2}\right)\ln q + \left(\frac{\sigma_\xi^2}{\sigma_q^2+\sigma_\xi^2}\right)\ln\mu_q + \left(\frac{\sigma_q^2}{\sigma_q^2+\sigma_\xi^2}\right)\ln\xi + \ln(1-\tau)\,.$$

Exponentiating, we obtain the level of wages: $w = q^s\mu_q^{1-s}\xi^s$, where $s=\sigma_q^2/(\sigma_q^2+\sigma_\xi^2)$. Given this wage, labor supply is $l=(1-\tau)^{\varepsilon_l}\,w^{\varepsilon_l}$, which implies an after-tax income of:

$$(1-\tau)\,z = (1-\tau)\,wl = (1-\tau)^{1+\varepsilon_l}\,w^{1+\varepsilon_l} = \left[(1-\tau)\,q^s\mu_q^{1-s}\xi^s\right]^{1+\varepsilon_l}.$$

Next, since $q=Q(x)=x^\beta$ and costs are linear, expected utility is as follows.

$$\left((1-\tau)^{1+\varepsilon_l}\,\mu_q^{(1-s)(1+\varepsilon_l)}\right)E\left[\xi^{s(1+\varepsilon_l)}\right]\frac{x^{\beta s(1+\varepsilon_l)}}{1+\varepsilon_l} - kx + \tau\overline{z}$$

Since I assume that $\beta s\left(1+\varepsilon_l\right)<1$, we can differentiate to find the agent's choice of q.

$$q=\left[\frac{\beta s\left(\left(1-\tau\right)^{1+\varepsilon_l}\mu_q^{(1-s)(1+\varepsilon_l)}\right)E\left[\xi^{s(1+\varepsilon_l)}\right]}{k}\right]^{\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}}$$

Then, since $\ln q$ is the sum of two normally distributed variables and a constant term, q is itself log-normally distributed. Specifically, it has the following distribution.

$$\begin{aligned}\ln q\sim N\Bigg(&\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\ln\beta+\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\ln s+\frac{\beta\left(1+\varepsilon_l\right)}{1-\beta s\left(1+\varepsilon_l\right)}\ln\left(1-\tau\right)\\&+\left(1-s\right)\frac{\beta\left(1+\varepsilon_l\right)}{1-\beta s\left(1+\varepsilon_l\right)}\ln\mu_q+\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\ln E\left[\xi^{s(1+\varepsilon_l)}\right]\\&-\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\left(\ln\mu_k-\frac{\sigma_k^2}{2}\right),\left(\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\right)^2\sigma_k^2\Bigg)\end{aligned}$$

Finally, we can obtain expressions for μ_q and σ_q^2 by matching coefficients.

$$\sigma_q^2=\left(\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\right)^2\sigma_k^2 \tag{48}$$

$$\mu_q=\left\{\frac{\beta s\left(1-\tau\right)^{1+\varepsilon_l}E\left[\xi^{s(1+\varepsilon_l)}\right]}{\mu_k}\exp\left[\left(1+\frac{\beta}{1-\beta s\left(1+\varepsilon_l\right)}\right)\frac{\sigma_k^2}{2}\right]\right\}^{\frac{\beta}{1-\beta\left(1+\varepsilon_l\right)}} \tag{49}$$

Equation 48 implicitly pins down σ_q^2 in terms of σ_k^2, β, ε_l and σ_ξ^2. It is independent of μ_k. In turn, equation 49 characterizes μ_q as a function of the same set of parameters plus μ_k. The elasticity of μ_q with respect to μ_k is $-\beta/\left[1-\beta\left(1+\varepsilon_l\right)\right]$. □

Proof of Lemma 2. There are two effects on q of increasing the retention rate $1-\tau$: a direct effect, and an effect via average productivity. Combining these yields the total elasticity.

$$\begin{aligned}\sigma_q=\frac{dq}{d\left(1-\tau\right)}\times\frac{1-\tau}{q}&=\left[\frac{\partial q}{\partial\left(1-\tau\right)}+\frac{\partial q}{\partial\mu_q}\frac{d\mu_q}{d\left(1-\tau\right)}\right]\frac{1-\tau}{q}\\&=\left[\frac{\beta\left(1+\varepsilon_l\right)}{1-\beta s\left(1+\varepsilon_l\right)}+\left(1-s\right)\frac{\beta\left(1+\varepsilon_l\right)}{1-\beta s\left(1+\varepsilon_l\right)}\frac{\beta\left(1+\varepsilon_l\right)}{1-\beta\left(1+\varepsilon_l\right)}\right]\\&=\frac{\beta\left(1+\varepsilon_l\right)}{1-\beta\left(1+\varepsilon_l\right)}\end{aligned}$$

Similarly, we can derive the elasticity of income z to the retention rate.

$$\begin{aligned}\sigma_z = \frac{dz}{d(1-\tau)} \times \frac{1-\tau}{z} &= \left[\frac{\partial z}{\partial(1-\tau)} + \frac{\partial z}{\partial q}\frac{\partial q}{\partial(1-\tau)} + \frac{\partial z_i}{\partial \mu_q}\frac{d\mu_q}{d(1-\tau)}\right]\frac{1-\tau}{z} \\ &= \varepsilon_l + (1+\varepsilon_l)\frac{\beta(1+\varepsilon_l)}{1-\beta(1+\varepsilon_l)} \\ &= \frac{\varepsilon_l + \beta(1+\varepsilon_l)}{1-\beta(1+\varepsilon_l)}\end{aligned}$$

□

Proof of Proposition 2. The utility of a worker with noise realization ξ and cost k is:

$$v = \frac{\left[(1-\tau)\, q^s \mu_q^{1-s} \xi^s\right]^{1+\varepsilon_l}}{1+\varepsilon_l} - kx + \tau\overline{z}$$

where x is chosen optimally according to the following first order condition.

$$k = \beta s \left((1-\tau)^{1+\varepsilon_l} \mu_q^{(1-s)(1+\varepsilon_l)}\right) E\left[\xi^{s(1+\varepsilon_l)}\right] x^{\beta s(1+\varepsilon_l)-1}$$

Taking the expectation over ξ, the expected utility for an individual with cost k is:

$$\left[\frac{1-\beta s(1+\varepsilon_l)}{1+\varepsilon_l}\right]\left((1-\tau)^{1+\varepsilon_l} \mu_q^{(1-s)(1+\varepsilon_l)}\right) E\left[\xi^{s(1+\varepsilon_l)}\right] q^{s(1+\varepsilon_l)} + \tau\overline{z}$$

Then, substituting in the optimal choice of q, and weighting by the worker's welfare weight ψ_k, we get expected welfare in terms of μ_q and ξ.

$$\begin{aligned}E_\xi\left[\psi_k v_{k,\xi}|k\right] = \psi_k &\left[\frac{1-\beta s(1+\varepsilon_l)}{1+\varepsilon_l}\right]\left((1-\tau)^{1+\frac{\varepsilon_l+\beta s(1+\varepsilon_l)}{1-\beta s(1+\varepsilon_l)}} \mu_q^{(1-s)(1+\varepsilon_l)\left[1+\frac{\beta s(1+\varepsilon_l)}{1-\beta s(1+\varepsilon_l)}\right]}\right) \\ &\times \left(\frac{\beta s}{k}\right)^{\frac{\beta s(1+\varepsilon_l)}{1-\beta s(1+\varepsilon_l)}} \left\{E\left[\xi^{s(1+\varepsilon_l)}\right]\right\}^{\frac{1}{1-\beta s(1+\varepsilon_l)}} + \psi_k\tau\overline{z} \\ &= (1-\tau)\,\psi_k\overline{z}_k\left[\frac{1-\beta s(1+\varepsilon_l)}{1+\varepsilon_l}\right] + \psi_k\tau\overline{z}\end{aligned}$$

Finally, we can integrate over cost realizations to obtain average welfare.

$$E\left[\psi_k v_{k,\xi}\right] = (1-\tau)\, E\left[\psi_k\overline{z}_k\right]\left[\frac{1-\beta s(1+\varepsilon_l)}{1+\varepsilon_l}\right] + \tau\overline{\psi}\overline{z}$$

Building on this result, there are three effects from raising the retention rate. First, there is a fiscal externality from the change in average income, $\overline{z}$.

$$\mathbf{FE} = \tau\overline{\psi}\varepsilon_z\frac{\overline{z}}{1-\tau}$$

Second, welfare rises due to the externality via employer beliefs. Specifically, differentiating with respect to μ_q and aggregating over k, the gain in social welfare is as follows.

$$\mathbf{BE} = (1-s)\, E_k\left(\psi_k\overline{z}_k\right)\varepsilon_q$$

Finally, there is a mechanical welfare loss due to the transfer from the average worker to high-income workers:

$$\mathbf{ME} = E_k\left(\psi_k\overline{z}_k\right) - \overline{\psi}\overline{z}$$

Summing the three effects we obtain an expression for the total welfare gain.

$$\mathbf{FE} + \mathbf{ME} + \mathbf{BE} = \frac{\tau}{1-\tau}\varepsilon_q\overline{z} + E_k\left(\psi_k\overline{z}_k\right)\left[1+(1-s)\,\varepsilon_q\right]\overline{z} - \overline{\psi}\overline{z}$$

Then setting this to zero yields the first order condition shown in the proposition. □

Proof of Proposition 3. The objective of the social planner is to maximize welfare $\boldsymbol{\mathcal{W}}(T)$ subject to the four constraints of Problem 5. This problem is restated here for convenience.

$$\max_{T}\ \boldsymbol{\mathcal{W}}(T) = W\left(\overline{V}(k,T)\right)dG(k)$$

where:

$$\overline{V}(k,T) = \int_\Theta \left(v(\theta|\pi,T) - kx(k,\pi,T)\right) f\left(\theta, q(k|\pi,T)\right) d\theta$$

subject to:

$$\begin{aligned}
x(k|\pi,T) &\in \underset{\tilde{x}\in\mathbb{R}_+}{\operatorname{argmax}}\ \int_\Theta v(\theta|\pi,T)\, f\left(\theta|Q(\tilde{x})\right)d\theta - k\tilde{x} \\
l(\theta|\pi,T) &\in \underset{\tilde{l}\in\mathbb{R}_+}{\operatorname{argmax}}\ u\left(w(\theta|\pi)\,\tilde{l} - T\left(w(\theta|\pi)\,\tilde{l}\right), \tilde{l}\right) \\
w(\theta|\pi) &= \frac{\int_K q(k|\pi,T)\, f\left(\theta|q(k|\pi,T)\right)dG(k)}{\int_K f\left(\theta|q(k|\pi,T)\right)dG(k)} \\
R &= \int_\Theta T\left(z(\theta|\pi,T)\right) f(\theta)\, d\theta
\end{aligned}$$

For ease of discussion, it will also be helpful to recall that $v(\theta|\pi,T)$ can be expanded and written as a function of a worker's wage, labor supply and tax liability.

$$v\left(\theta|\pi,T\right)=u\left(w\left(\theta|\pi\right)l\left(\theta|\pi,T\right)-T\left(w\left(\theta|\pi\right)l\left(\theta|\pi,T\right)\right),l\left(\theta|\pi,T\right)\right) \tag{50}$$

A perturbation to T as described has three effects that I will consider in turn. First, there is a welfare loss (**WL**) from taking money from individuals with income higher than z.

$$\mathbf{WL}=-d\tau dz\left\{\int_{\theta(z|\pi,T)}^{\overline{\theta}}u_c(\theta)\int_K\psi(k)dG(k|\theta)f\left(\theta\right)d\theta\right\} \tag{51}$$

Since the revenue raised is returned to all individuals equally via an increase in the intercept of the tax schedule, it is worth λ per dollar in terms of social welfare, where:

$$\lambda=\int_\Theta u_c(\theta)\int_K\psi(k)dG(k|\theta)f\left(\theta\right)d\theta \tag{52}$$

Multiplying by the amount of revenue raised, the welfare gain (**WG**) from this transfer is:

$$\mathbf{WG}=d\tau dz\left\{\int_{\theta(z|\pi,T)}^{\overline{\theta}}f\left(\theta\right)d\theta\right\}\lambda. \tag{53}$$

Summing **WL** and **WG**, then dividing by λ yields the mechanical gain in welfare, **ME** (z).

The second effect to consider is the fiscal externality, **FE** (z), which arises when individuals re-optimize. The value of the fiscal externality can be obtained by differentiating the resource constraint, yielding the impact on government revenue from re-optimization.

Since the focal selection $\left(E\left(T\right),T\right)$ is assumed to be locally continuously differentiable with respect to T, $l\left(\theta|\pi,T\right)$ and $x\left(k|\pi,T\right)$ respond continuously to the perturbation. Next, since $x\left(k|\pi,T\right)$ responds continuously and Q is differentiable, so does $q\left(k|\pi,T\right)=Q\left(x\left(k|\pi,T\right)\right)$. Finally, since $f\left(\theta\right)=\int_K f\left(\theta|q\left(k|\pi,T\right)\right)dG(k)$ is continuous in $q\left(k|\pi,T\right)$, $f\left(\theta\right)$ responds continuously. In turn, this implies that $w\left(\theta|\pi\right)$ responds continuously. The change in income given a signal realization θ can therefore be written as follows.

$$-\frac{dz(\theta|\pi,T)}{d\left[1-T'\left(z\right)\right]}=-w\left(\theta|\pi,T\right)\frac{dl\left(\theta|\pi,T\right)}{d\left[1-T'\left(z\right)\right]}-l\left(\theta|\pi,T\right)\frac{dw\left(\theta|\pi,T\right)}{d\left[1-T'\left(z\right)\right]}$$

These results allow the fiscal externality to be written as a combination of the effects of changes in $z(\theta|\pi,T)$ and $f(\theta)$, capturing the effect on government revenue from both investment and labor supply decisions. After dividing through by λ, the total fiscal externality is as follows.

$$\mathbf{FE}\left(z\right)=-d\tau dz\int_{\Theta}\left\{T'(z(\tilde{\theta}|\pi))\left(\frac{dz(\tilde{\theta}|\pi,T)}{d\left[1-T'\left(z\right)\right]}\right)f(\tilde{\theta})-T(z(\tilde{\theta}|\pi,T))\frac{df(\tilde{\theta})}{d\left[1-T'\left(z\right)\right]}\right\}d\tilde{\theta}$$

The final effect of taxation is the effect on individual utility of changing wages in response to shifts in the distribution of productivity (**BE**). Since individuals take the wage paid given any signal realization as fixed, they ignore this effect. Differentiating the belief consistency constraint, the effect of a rise in individual k's productivity on the wage of a worker with signal realization θ is as follows.

$$\frac{dw(\theta|\pi)}{dq\left(k|\pi,T\right)}=\frac{f(\theta,q(k|\pi,T))}{f(\theta)}+\left(\frac{\left.\frac{\partial f(\theta,q)}{\partial q}\right|_{q=q(k|\pi,T)}}{f(\theta)}\right)\left[q\left(k|\pi,T\right)-E(q|\theta,\pi)\right]$$

Applying the envelope theorem and again dividing by λ, the effect of this wage change on social welfare is simply scaled by the affected worker's labor supply, retention rate and the average welfare weight of an individual with signal realization θ.

$$\frac{dw(\tilde{\theta}|\pi)}{dq\left(k|\pi,T\right)}\psi_z(z(\tilde{\theta}|\pi,T))\left[1-T'(z(\tilde{\theta}|\pi,T))\right]l(\tilde{\theta}|\pi)$$

To obtain the total belief externality shown in the main text, we then integrate over the distributions of θ and k.

These three effects jointly capture the total change in welfare from a perturbation, since the effects of individuals' re-optimization on their own welfare are second-order. Thus, given any continuous selection, if $\mathbf{FE}+\mathbf{BE}+\mathbf{ME}\neq 0$, welfare increases in response either to an arbitrarily small positive perturbation or an equivalent negative perturbation. Except at a discontinuity at which **ME**, **FE** and **BE** are not defined, a necessary condition for an optimum is therefore that the sum of the three effects is zero. □

Proof of Proposition 4. Assume – subject to verification – that productivity and investment are log-normally distributed.

$$q\sim LN\left(\ln\mu_q-\frac{\sigma_q^2}{2},\sigma_q^2\right)$$

Next, suppose the relationship between productivity and investment can be written as:

$$\ln q=\ln A+B\ln x$$

where A and B are scalars that will be found by matching coefficients. This allows the signal to be written as a linear combination of productivity q and noise ξ.

$$\ln \theta = \left(\frac{1}{B}\right) \ln q - \left(\frac{1}{B}\right) \ln A + \ln \xi$$

For convenience, define $\ln \tilde{\xi} = B \ln \xi$ and let $\ln \tilde{\theta}$ be the following linear transformation of the original signal.

$$\ln \tilde{\theta} = B \ln \theta + \ln A = \ln q + B \ln \xi = \ln q + \ln \tilde{\xi}$$

The expected log-marginal product of an individual follows from the fact that the employer faces a standard normal signal extraction problem:

$$E\left[\ln q | \tilde{\theta}\right] = s \ln \tilde{\theta} + (1-s)\left(\ln \mu_q - \frac{\sigma_q^2}{2}\right)$$

where $s = \sigma_q^2/(\sigma_q^2 + \sigma_{\tilde{\xi}}^2) = \sigma_x^2/(\sigma_x^2 + \sigma_{\xi}^2)$. A worker's expected level of productivity is therefore a geometric weighted average of A, x, ξ and μ_q.

$$w = \tilde{\theta}^s \mu_q^{1-s} = A^s x^{sB} \xi^{sB} \mu_q^{1-s}$$

Optimal labor supply is $l = (1-\tau)^{\varepsilon_l} w^{\varepsilon_l}$, which means that after tax income is:

$$\begin{aligned}
(1-\tau) z &= (1-\tau)^{1+\varepsilon_l} w^{1+\varepsilon_l} \\
&= (1-\tau)^{1+\varepsilon_l} \left[A^s x^{sB} \xi^{sB} \mu_q^{1-s}\right]^{1+\varepsilon_l}.
\end{aligned}$$

In turn, this implies a value of expected utility for any investment level.

$$v = \left[A^s (1-\tau) \mu_q^{1-s}\right]^{1+\varepsilon_l} E\left[\xi^{sB(1+\varepsilon_l)}\right] \frac{x^{sB(1+\varepsilon_l)}}{1+\varepsilon_l} - kx + \tau \overline{z}$$

Assuming again that $\beta s(1+\varepsilon_l) < 1$, it will also turn out to be true that $sB(1+\varepsilon_l) < 1$. This in turn ensures that the worker's optimal choice of $\ln x$ is as follows.

$$\begin{aligned}
\ln x = \frac{1}{1 - sB(1+\varepsilon_l)} \Big[&\ln n + \ln(sB) + (1+\varepsilon_l) \ln(1-\tau) + (1-s)(1+\varepsilon_l) \ln \mu_q \\
&+ \ln E\left[\xi^{s(1+\varepsilon_l)}\right] + s(1+\varepsilon_l) \ln A \Big]
\end{aligned}$$

Next, using the fact that $\ln q = \alpha \ln n + \beta(1-\alpha)\ln x$, and matching coefficients, B is:

$$B = \frac{\alpha + \beta(1-\alpha)}{1 + s\alpha(1+\varepsilon_l)}.$$

This can in turn be used to solve for $\ln A$ in terms of x.

$$\ln A = \alpha \ln n - \frac{\alpha - \beta(1-\alpha)s\alpha(1+\varepsilon_l)}{1 + s\alpha(1+\varepsilon_l)} \ln x$$

A can then be eliminated to yield a new expression for $\ln x$.

$$\begin{aligned}\ln x = & \frac{1 + s\alpha(1+\varepsilon_l)}{1 - s\beta(1-\alpha)(1+\varepsilon_l)} \ln(n) + \frac{1}{1 - s\beta(1-\alpha)(1+\varepsilon_l)} \left[\ln s + \ln\left(\frac{\alpha + \beta(1-\alpha)}{1 + s\alpha(1+\varepsilon_l)}\right)\right. \\ & \left. + (1+\varepsilon_l)\ln(1-\tau) + (1-s)(1+\varepsilon_l)\ln\mu_q + \ln E\left[\xi^{s(1+\varepsilon_l)}\right]\right]\end{aligned}$$

Finally, since x inherits the log-normality of n, and $\ln q = \alpha \ln n + (1-\alpha)\beta \ln x$, q is also log-normal. This means that the values of μ_q and σ_q^2 can be found by matching coefficients.

$$\begin{aligned}\sigma_q^2 &= \left[\frac{\alpha + \beta(1-\alpha)}{1 - \beta s(1-\alpha)(1+\varepsilon_l)}\right]^2 \sigma_n^2 \\ \ln\mu_q &= \left[\frac{\alpha + \beta(1-\alpha)}{1 - (1-\alpha)\beta(1+\varepsilon_l)}\right] \ln n + \left[\frac{\beta(1-\alpha)}{1 - (1-\alpha)\beta(1+\varepsilon_l)}\right] \\ & \quad \times \left\{\ln s + \ln\left[\frac{\alpha + \beta(1-\alpha)}{1 + s\alpha(1+\varepsilon_l)}\right] + (1+\varepsilon_l)\ln(1-\tau) + \ln E\left[\tilde{\xi}^{s(1+\varepsilon_l)}\right]\right\} \\ & \quad + \frac{\sigma_n^2}{2}\left[\frac{\alpha + \beta(1-\alpha)}{1 - s(1-\alpha)\beta(1+\varepsilon_l)}\right]\left[\frac{\alpha + \beta(1-\alpha)}{1 - (1-\alpha)\beta(1+\varepsilon_l)}\right]\end{aligned}$$

The elasticity of productivity follows directly.

$$\frac{d\ln\mu_q}{d\ln(1-\tau)} = \left(\frac{\beta(1-\alpha)(1+\varepsilon_l)}{1 - \beta(1-\alpha)(1+\varepsilon_l)}\right)$$

Finally, the elasticity of income can be found as follows.

$$\begin{aligned}\frac{d\ln z}{d\ln(1-\tau)} &= \frac{\partial z}{\partial(1-\tau)} + \frac{\partial \ln z}{\partial \ln q}\frac{\partial \ln q}{\partial \ln(1-\tau)} + \frac{\partial \ln z}{\partial \ln \mu_q}\frac{d\ln\mu_q}{d\ln(1-\tau)} \\ &= \varepsilon_l + (1+\varepsilon_l)\left[\frac{\beta(1-\alpha)(1+\varepsilon_l)}{1 - \beta(1-\alpha)(1+\varepsilon_l)}s + \frac{\beta(1-\alpha)(1+\varepsilon_l)}{1 - \beta(1-\alpha)(1+\varepsilon_l)}(1-s)\right] \\ &= \frac{\varepsilon_l + (1+\varepsilon_l)\beta(1-\alpha)}{1 - \beta(1-\alpha)(1+\varepsilon_l)}\end{aligned}$$

$\square$

Proof of Proposition 5. Using the results from Proposition 4, a worker's expected utility, $\overline{v}_n$, can be derived in the same way as in Proposition 2.

$$\overline{v}_n = n^{\frac{s(1+\varepsilon_l)[\alpha+\beta(1-\alpha)]}{1-\beta s(1-\alpha)(1+\varepsilon_l)}} \left[(1-\tau)^{(1+\varepsilon_l)} \mu_q^{(1-s)(1+\varepsilon_l)} E\left(\varepsilon^{s(1+\varepsilon_l)}\right)\right]^{\frac{1}{1-\beta s(1-\alpha)(1+\varepsilon_l)}} [sB]^{\frac{\beta s(1+\varepsilon_l)(1-\alpha)}{1-\beta s(1-\alpha)(1+\varepsilon_l)}} \times \left[\frac{1-(1+\varepsilon_l)\, sB}{1+\varepsilon_l}\right] + \tau\overline{z}$$

where $B = \frac{\alpha+\beta(1-\alpha)}{1+s\alpha(1+\varepsilon_l)}$. The expected after-tax income for an individual with investment cost n can be derived similarly.

$$(1-\tau)\,\overline{z}_n = n^{\frac{s(1+\varepsilon_l)[\alpha+\beta(1-\alpha)]}{1-\beta s(1-\alpha)(1+\varepsilon_l)}} \left[(1-\tau)\,\mu_q^{(1-s)}\right]^{\frac{1+\varepsilon_l}{1-\beta s(1-\alpha)(1+\varepsilon_l)}} \times E\left[\varepsilon^{s(1+\varepsilon_l)}\right]^{\frac{1}{1-\beta s(1-\alpha)(1+\varepsilon_l)}} [sB]^{\frac{\beta s(1+\varepsilon_l)(1-\alpha)}{1-\beta s(1-\alpha)(1+\varepsilon_l)}}$$

The welfare of workers with ability n can then be re-written in terms of income, and weighted by ψ_n.

$$\psi_n\overline{v}_n = (1-\tau)\,\psi_n\overline{z}_n\left[\frac{1-(1+\varepsilon_l)\,sB}{1+\varepsilon_l}\right] + \tau\psi_n\overline{z}$$

Differentiating $\psi_n\overline{v}_n$ with respect to $1-\tau$, we obtain the effects on welfare of both the mechanical transfer *and* the distortion from the unproductive component of investment, which is built into $\overline{v}_n$. Then taking the expectation over ability types, n, we obtain:

$$\textbf{MEU} = E\left[\overline{z}_n\psi_n\right]\left[\frac{1}{1+s\alpha\,(1+\varepsilon_l)}\right] - \overline{\psi}\overline{z}$$

Next, we can calculate the belief externality. This is again captured by the effect via μ_q. Using the elasticities from Proposition 4 and the expression for $\overline{v}_n$, the effect on the welfare of a worker with ability n is:

$$\frac{(1+\varepsilon_l)\,(1-s)}{1-\beta s\,(1-\alpha)\,(1+\varepsilon_l)}\frac{\overline{v}_n-\tau\overline{z}}{\mu_q}\frac{\beta\,(1-\alpha)\,(1+\varepsilon_l)}{1-\beta\,(1-\alpha)\,(1+\varepsilon_l)}\frac{\mu_q}{1-\tau}.$$

Weighting by ψ_n, using the expression for $\overline{v}_n$ and taking the expectation over ability types, this gives us the total belief externality.

$$\textbf{BE} = (1-s)\,E\left[\overline{z}_n\psi_n\right]\left[\frac{1}{1+s\alpha\,(1+\varepsilon_l)}\right]\frac{\beta\,(1-\alpha)\,(1+\varepsilon_l)}{1-\beta\,(1-\alpha)\,(1+\varepsilon_l)}$$

Finally, the fiscal externality follows from the elasticity of income.

$$\textbf{FE} = \tau \left[\frac{\varepsilon_l + (1+\varepsilon_l)\,\beta\,(1-\alpha)}{1-\beta\,(1-\alpha)\,(1+\varepsilon_l)} \right] \frac{\overline{z}}{1-\tau}$$

By the same argument as Proposition 2, the sum of **BE**, **MEU** and **FE** must be zero for τ to be optimal, which yields the result.

$$\frac{\tau}{1-\tau} = \frac{1 - E_n\left(\frac{\overline{z}_n}{\overline{z}}\frac{\overline{\psi}_n}{\overline{\psi}}\right)\left[\frac{1}{1+s\alpha(1+\varepsilon_l)}\right]\left[1+(1-s)\left(\frac{(1+\varepsilon_l)\beta(1-\alpha)}{1-\beta(1-\alpha)(1+\varepsilon_l)}\right)\right]}{\frac{\varepsilon_l+(1+\varepsilon_l)\beta(1-\alpha)}{1-\beta(1-\alpha)(1+\varepsilon_l)}}$$

□

Proof of Lemma 3. Firm beliefs about the distribution of productivity in the population must be confirmed in equilibrium and identical across firms. Let π denote the equilibrium set of beliefs. Firm j's expectation of the worker's productivity is $E\left[q|\theta,\pi,A_j=1\right] \geq 0$. Finally, let $\tilde{u}\left(C_j\right) = u\left(z_j - T\left(z_j\right), l_j\right)$ represent the utility that the worker receives from accepting offer C_j.

Suppose that some firm j makes strictly positive expected profits given its contract offer C_j. It must then be the case that $\tilde{u}\left(C_j\right) \geq \tilde{u}\left(C_k\right)$ for all contracts C_k offered by other firms. There are several cases to consider, each of which will lead to a contradiction.

Case 1: $\tilde{u}\left(C_j\right) > \tilde{u}\left(C_k\right)$ for some C_k.

Firm k initially earns zero expected profit, since not workers accept its offer. However, it can replicate C_j but slightly reduce l_j. By doing so, it attracts the worker with probability one and earns strictly positive profits. This is a profitable deviation.

Case 2: $\tilde{u}\left(C_j\right) = \tilde{u}\left(C_k\right)$ for all C_k, and $\overline{P}_{k,\theta} \leq 0$ for some k.

If any firm makes weakly negative profits, then the same deviation as Case 1 applies.

Case 3: $\tilde{u}\left(C_j\right) = \tilde{u}\left(C_k\right)$ and $\overline{P}_{k,\theta} > 0$ for all k.

Since the worker always accepts an offer, $E\left[q|\theta,\pi,C_j\right]$ is bounded weakly below $E\left[q|\theta,\pi\right]$ for at least one firm. This firm's expected profit is bounded below $\overline{P}_{\text{MAX}}$.

$$\overline{P}_{\text{MAX}} = \max_{l,z} E\left[q|\theta,\pi\right] l - z \qquad \text{s.t.} \qquad u\left(z - T\left(z\right), l\right) \geq u\left(T\left(0\right), 0\right)$$

The assumptions on the worker's utility function ensure that this yields finite labor supply for any finite $E\left[q|\theta,\pi\right]$. Since z_j is restricted to be greater than zero and

$E[q|\theta,\pi]$ is finite, $\overline{P}_{\text{MAX}}$ is also bounded. Finally, this firm can strictly increase its profit by reducing l_j slightly and attracting the worker with probability one.

Since every case in which a firm makes a strictly positive expected profit implies a profitable deviation, and all firms can obtain at least zero expected profit by offering a contract with $z_j = 0$, it must be true that every firm makes zero expected profit.

Next consider two cases for the worker's effective wage and labor supply.

Case A: One firm hires the worker with probability one.

If one firm j always hires the worker in equilibrium, zero profit implies directly that the worker's wage is her expected marginal product.

$$w_j = \frac{z_j}{l_j} = E[q|\theta,\pi]$$

Next, suppose that C_j specifies a labor supply $l_j \notin L^*$ where:

$$L^* = \underset{\tilde{l}_j}{\operatorname{argmax}}\ u\left(E[q|\theta,\pi]\,\tilde{l}_j - T\left(E[q|\theta,\pi]\,\tilde{l}_j\right), \tilde{l}_j\right).$$

Some other firm k could offer a contract with the same implied wage as C_j but with $l_k \in L^*$. Since $w_j = E[q|\theta,\pi]$, this produces zero profits but the worker's utility is strictly higher. Firm k can now increase l_k slightly, thereby attracting the worker with probability one and earning strictly positive profit. Thus, it must be that $l_j \in L^*$.

Case B: Multiple firms hire the worker with positive probability.

Since each firm earns zero profit, a similar wage condition must hold for firms who hire a worker with positive probability.

$$w_j = \frac{z_j}{l_j} = E[q|\theta,\pi,A_j = 1]\,\forall j$$

Moreover, similar logic to above implies that $l_j \in L_j^*$ where:

$$L_j^* = \underset{\tilde{l}_j}{\operatorname{argmax}}\ u\left(E[q|\theta,\pi,A_j = 1]\,\tilde{l}_j - T\left(E[q|\theta,\pi,A_j = 1]\,\tilde{l}_j\right), \tilde{l}_j\right).$$

Otherwise, firm j could offer a contract with the same implied wage but with $l_j \in L_j^*$, so that $\tilde{u}(C_j)$ is higher than before. It could then slightly increase l_j. The worker would always accept the firm's offer and it earns strictly positive expected profit.

Next, suppose $E[q|\theta, \pi, A_j = 1] > E[q|\theta, \pi, A_k = 1]$ for some firms j and k. For at least one pair, it must be that $E[q|\theta, \pi, A_j = 1] > E[q|\theta, \pi] > E[q|\theta, \pi, A_k = 1]$. Let $l_j^* \in L_j^*$ be the labor supply offered by firm j. By the definition of L_j^* we know that:

$$u\left(w_j l_j^* - T\left(w_j l_j^*\right), l_j^*\right) \geq u\left(w_j l_k^* - T\left(w_j l_k^*\right), l_k^*\right).$$

Suppose now that $u\left(w_j l_k^* - T\left(w_j l_k^*\right), l_k^*\right) \leq u\left(w_k l_k^* - T\left(w_k l_k^*\right), l_k^*\right)$. Then firm j can alter its offer to $z_j = w_k l_k^* < w_j l_k^*$ and set l_j below but arbitrarily close to l_k. Firm j then attracts the worker with probability one. Since $E[q|\theta, \pi] > E[q|\theta, \pi, A_k = 1]$, firm j can make strictly positive profit with this strategy.

Alternatively, suppose that $u(w_j l_k^* - T(w_j l_k^*), l_k^*) > u(w_k l_k^* - T(w_k l_k^*), l_k^*)$, which implies that $u(w_j l_j^* - T(w_j l_j^*), l_j^*) > u(w_k l_k^* - T(w_k l_k^*), l_k^*)$. This is a contradiction since we assumed that *both* firms attract the worker with positive probability. which requires that $u(w_j l_j^* - T(w_j l_j^*), l_j^*) = u\left(w_k l_k^* - T\left(w_k l_k^*\right), l_k^*\right)$.

In conclusion, firms must earn zero expected profit, and $E[q|\theta, \pi, A_j = 1] = E[q|\theta, \pi]$. □

Proof of Proposition 6. I begin by establishing that there is an equilibrium with zero investment. The stated assumptions ensure that $w(\theta|\pi)$ is strictly increasing in π, that $w(\theta|0) = 0$ for all θ and that $w(\theta|1) = \omega$ for all θ. This guarantees that $\overline{v}_q(0|\tau) = \overline{v}_u(0|\tau)$ and $\overline{v}_q(1|\tau) = \overline{v}_u(1|\tau)$, which in turn implies that $G(\beta(0|\tau)) = 0$ and $G(\beta(1|\tau)) = 0$. Thus, there is a solution with no investment and no solution in which all agents invest.

Finally, if $G(\beta(\pi|\tau)) > \pi$ for some π^* then the continuity of $\phi(\theta)$ and G combined with the fact that $G(\beta(1|\tau)) = 0$ implies that $G(\beta(\hat{\pi}|\tau)) = \hat{\pi}$ for some $\hat{\pi} > \pi^*$. There are therefore at least two solutions to equilibrium condition 39. □

Proof of Proposition 7. Social welfare is given by:

$$\pi\overline{v}_q(\pi) + (1-\pi)\overline{v}_u(\pi) - \int_0^{\overline{v}_q(\pi)-\overline{v}_u(\pi)} k dG(k).$$

where:

$$\overline{v}_q(\pi|\tau) = \int_0^1 v(\theta|\pi) dF_q(\theta) - k$$
$$\overline{v}_u(\pi|\tau) = \int_0^1 v(\theta|\pi) dF_u(\theta).$$

By differentiating the equation for the worker's wage, it can be shown that the wage is increasing in π.

$$\frac{\partial w\left(\theta|\pi\right)}{\partial \pi}=\omega\times\frac{f_u\left(\theta\right)f_q\left(\theta\right)}{\left[\pi f_q\left(\theta\right)+\left(1-\pi\right)f_u\left(\theta\right)\right]^2}>0$$

In turn, this means that $v\left(\theta|\pi,\tau\right)=u\left(\left(1-\tau\right)w\left(\theta|\pi\right)+\tau\pi w\right)$ is increasing in π. Thus, holding investment decisions and τ constant, welfare increases with π. The accompanying change in individual investment decisions can only make those marginal individuals better off. Thus, welfare is higher for all workers.

Next, let $\pi^*\left(\tau\right)$ be the investment rate in the planner's preferred equilibrium for each tax rate. The proof that $\pi^*\left(\tau\right)$ rises as with τ falls is simple. First, if $\pi^*\left(\tau\right)=0$, it cannot fall. Alternatively, suppose that $\pi^*\left(\tau_0\right)>0$. Since lowering τ from τ_0 to τ_1 raises $G(\beta(\pi|\tau))$ for any π, it must be true that $G(\beta(\pi^*\left(\tau_0\right)|\tau_1))>\pi^*\left(\tau_0\right)$. Since $G(\beta(\pi|\tau))$ is continuous and $G(\beta(1|\tau))=0$, there must be some higher investment rate $\hat{\pi}$ such that $G(\beta(\hat{\pi}|\tau_1))=\hat{\pi}>\pi^*\left(\tau_0\right)$. Since the equilibrium with the highest investment rate Pareto dominates all others, the planner's preferred equilibrium now features a higher investment rate. □

Proof of Proposition 8. Just as in Sections 3 and 4, there are three effects from a fall in τ. First, there is a mechanical effect. For a worker with signal θ, this is as follows.

$$\begin{aligned}\frac{\partial v\left(\theta|\pi\right)}{\partial\left(1-\tau\right)}&=u'\left[\left(1-\tau\right)\omega\frac{\pi f_q\left(\theta\right)}{\pi f_q\left(\theta\right)+\left(1-\pi\right)f_u\left(\theta\right)}+\tau\pi\omega\right]\left[\frac{\pi f_q\left(\theta\right)}{\pi f_q\left(\theta\right)+\left(1-\pi\right)f_u\left(\theta\right)}-\pi\right]\omega\\&=u'_\theta\pi\left(1-\pi\right)\omega\left[\frac{f_q\left(\theta\right)-f_u\left(\theta\right)}{\pi f_q\left(\theta\right)+\left(1-\pi\right)f_u\left(\theta\right)}\right]\end{aligned}$$

Aggregating up, we obtain the total mechanical effect on social welfare.

$$\mathbf{ME}=\omega\pi\left(1-\pi\right)\int_0^1 u'_\theta\left[f_q\left(\theta\right)-f_u\left(\theta\right)\right]d\theta=-\omega\pi\overline{v}_\tau$$

Next, there is a fiscal externality. Assuming $\pi^\dagger\left(\tau\right)$ is locally continuous, this is given by:

$$\mathbf{FE}=\tau\frac{d\pi}{d\left(1-\tau\right)}\omega\int_0^1 u'_\theta\left[\pi f_q\left(\theta\right)+\left(1-\pi\right)f_u\left(\theta\right)\right]d\theta=\frac{\tau}{1-\tau}\pi\varepsilon_z\omega\overline{u}'_\theta$$

Finally, there is the externality via employer beliefs, which raises wages for all workers but is not taken into account when workers optimize. Using the continuity of $\pi^\dagger\left(\tau\right)$ again:

$$\mathbf{BE}=\left(1-\tau\right)\frac{d\pi}{d\left(1-\tau\right)}\int_0^1 u'_\theta\left[\frac{\partial w\left(\theta|\pi\right)}{\partial\pi}\right]\left[\pi f_q\left(\theta\right)+\left(1-\pi\right)f_u\left(\theta\right)\right]d\theta$$

$$= \varepsilon_z \pi \omega \int_0^1 u'_\theta \left[\frac{f_u(\theta) f_q(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)} \right] d\theta = \varepsilon_z \pi \overline{w}_z$$

Adding the three effects and re-arranging yields the following first-order condition.

$$\begin{aligned} \frac{\tau}{1-\tau} &= \frac{(1-\pi)\int_0^1 u'_\theta \left[f_u(\theta) - f_q(\theta)\right] d\theta - \varepsilon_z \int_0^1 u'_\theta \left[\frac{f_u(\theta) f_q(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)}\right] d\theta}{\varepsilon_z \int_0^1 u'_\theta \left[\pi f_q(\theta) + (1-\pi) f_u(\theta)\right] d\theta} \\ &= \frac{\overline{v}_\tau - \varepsilon_z \overline{w}_z}{\varepsilon_z} \end{aligned}$$

□

Proof of Proposition 9. Fixing a value of $T_{A\to D}$, the proof that condition 41 must hold at the optimum is analogous to the proof of Proposition 8. A similar perturbation argument can be used to establish that condition 42 must hold. An increase in $T_{A\to D}$ leads to the following gain in welfare for type A and D individuals:

$$-\Delta_A = \frac{1}{\lambda_A} \int_0^1 u'_{A,\theta} dF(\theta) \qquad\qquad \Delta_D = \frac{1}{\lambda_D} \int_0^1 u'_{D,\theta} dF(\theta)$$

The welfare gain, $\lambda_D \Delta_D - \lambda_A \Delta_A$, must be zero at interior optima if $\pi^\dagger(T)$ is locally continuous, implying condition 42. □

H Simulation of the Model

This appendix provides detailed information on the methods I use to simulate the full model. The first step is to discretize the signal space into n_θ possible values, and categorize individuals into n_q groups, each with a different productivity decision. I then use the noise and productivity distributions to define an $n_q \times n_\theta$ matrix $\boldsymbol{B}_0$, which maps productivity decisions to distributions of realized signals.

A. Evaluation of a Single Perturbation

Evaluation of a perturbation proceeds as follows. First, define a perturbation that raises the tax rate on income between $\underline{z}$ and $\overline{z}$ by $\Delta T'$. This yields a new tax schedule, T_1.

$$T'_1(z) = \begin{cases} T'_0(z) + \Delta T' & \text{if } z \in [\underline{z}, \overline{z}) \\ T'_0(z) & \text{otherwise} \end{cases}$$

Take the existing wage given each θ but apply T_1 instead of T_0. Re-optimize labor supply decisions and calculate $v\left(w\left(\theta|\pi_0\right)|T_1\right)$ for each θ, yielding a candidate vector of utilities $\boldsymbol{v}_1^{(0)}$. Using $\boldsymbol{v}_1^{(0)}$, calculate $E_\theta\left(v\left(\theta|\pi_0, T_1\right)|q\right)$ and adjust workers' investment decisions toward their preferred choice. This yields a new distribution of productivity, $\delta_1^{(0)}\left(q|\pi_0, T_1\right)$.

In the discretized space, $\delta_1^{(0)}\left(q|\pi_0, T_1\right)$ implies a new candidate vector of productivity choices $\boldsymbol{q}_1^{(1)}$. Use these choices to reconstruct a new candidate $\boldsymbol{B}_1^{(1)}$ matrix. Then solve for employers' rational productivity inferences at each value of θ, yielding a candidate set of employer beliefs $\pi_1^{(1)}(q)$ and a hypothesized vector of wages $\boldsymbol{w}_1^{(1)}$.

$$\boldsymbol{w}_1^{(1)} = \left[\text{diag}\left(\boldsymbol{B}_1^{(1)\prime} \times \delta_1^{(1)}\left(q|\pi_1^{(1)}, T_1\right)\right)\right]^{-1} \times \left[\boldsymbol{B}_1^{(1)\prime} \times \text{diag}\left(\boldsymbol{q}_1^{(1)}\right) \times \delta_1^{(1)}\left(q|\pi_1^{(1)}, T_1\right)\right]$$

Recalculate utilities to obtain $\boldsymbol{v}_1^{(1)}$ and adjust workers' investment decisions again, yielding $\boldsymbol{q}^{(2)}$. Iterate this process until individuals do not want to adjust their investment decisions given the hypothesized employer beliefs: i.e., when $\pi_1^{(k)}(q) \approx \delta_1^{(k)}\left(q|\pi_1^{(k)}(q), T_1\right)$. At this point, the process has converged.

Once this inner fixed point has been obtained, compare the new value of expected utility for each level of costs, weight using the assumed social welfare function, and adopt the perturbation if and only if it produced an increase in average social welfare.

B. Decomposition of a Perturbation

The effect of a perturbation on equilibrium social welfare can be decomposed into its three components: the mechanical effect (**ME**), the fiscal externality (**FE**) and the belief externality (**BE**). To calculate the mechanical effect, simply hold all decisions (wages, labor supply and investment) constant and evaluate the mechanical change in utility. The belief externality can be calculated by comparing the true gain in expected utility to the gain holding fixed the wage paid at each level of θ. Finally, the fiscal externality can be evaluated by subtracting the behavioral effect on tax revenue from all individuals' incomes.

C. Solving for the Optimal Tax Schedule

To solve for the optimal tax schedule, simply consider a series of perturbations as defined above. Define a size for each perturbation, ΔT. Then divide the income distribution into n_b tax brackets. Loop through the tax brackets and calculate the gain in welfare from a perturbation in each direction. Adopt the perturbation that increases welfare, then move to the next bracket. Repeat until there are no perturbations that increase welfare. Optionally, reduce the size of each perturbation and repeat.

D. Recovery of Fundamentals

To back out fundamentals for the simulation described in Section 5, I begin with the Pareto log-normal approximation of the United States wage distribution provided by Mankiw et al. (2009). Next, I use this wage distribution, and the posited log-normal conditional signal distribution, to infer a productivity distribution that produces this wage schedule.

The specific procedure that I follow is to parameterize a Champernowne distribution for log wages with density proportional to:

$$\frac{1}{\frac{1}{2}\exp\left(\alpha\left(z-z_0\right)\right)+\lambda+\frac{1}{2}\exp\left(-\alpha\left(z-z_0\right)\right)}$$

To choose the parameters, I use MATLAB's *fminunc* function to solve for the set of parameters that jointly minimize the Kullback-Leibler divergence between the target wage distribution f_w and the simulated distribution f_w^{sim}.

$$D_{KL}\left(f_w||f_w^{\text{sim}}\right)=\sum_w f_w\left(w\right)\ln\left(\frac{f_w\left(w\right)}{f_w^{\text{sim}}\left(w\right)}\right)$$

As Figure 7 shows, this process is effective.

For each wage, I can then calculate utility $v\left(w\left(\theta|\pi\right)|T_0\right)$, given an initial tax system T_0, by solving workers' labor supply problems for each value of θ. Expected utility for each level of productivity is then given by:

$$E_\theta\left(v\left(\theta|\pi_0,T_0\right)|q\right)=\underbrace{\boldsymbol{B}_0}_{n_q\times n_\theta}\times\underbrace{\boldsymbol{v}_0}_{n_\theta\times 1}$$

where $\boldsymbol{v}_0$ is a vector that stacks the utility realized at each value of θ and π_0 denotes employers' current and correct beliefs about the distribution of productivity. Combined with individuals' productivity choices and a value of β, this vector of expected utilities can then be used to back out an implied cost distribution.

E. Additional Figures

Table 3 provides summary statistics for data used to test for employer learning in Section 5. Figures 12 to 14 compare the mechanical effect, fiscal externality and belief externality between the naïve and optimal tax schedules, in each tax bracket for the simulation in Section 5. Additionally, Figure 15 plots the change in marginal social welfare weights starting from naïve taxation and transitioning to optimal taxation.

TABLE 3: SUMMARY STATISTICS FOR HIGH AND LOW AFQT WORKERS

	Low AFQT		**High AFQT**	
	Mean	Standard Deviation	Mean	Standard Deviation
AFQT	−0.67	(0.67)	1.00	(0.40)
Log(wage)	6.68	(0.46)	7.03	(0.54)
Experience	7.21	(5.99)	8.13	(6.05)
Years since left school	10.55	(6.49)	9.74	(6.29)
Urban (%)	74.4		78.6	
Education (%)				
– 12 years	59.7		35.7	
– 16 years	3.8		25.4	
– Other	36.5		38.9	
Observations	18921		18903	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table notes. Data are from the National Longitudinal Survey of Youth (NLSY79). The sample is restricted to working black and white men who have wages between one and one hundred dollars and at least eight years of schooling. AFQT is a worker's score on the armed forces qualification test, standardized by age to have zero mean and unit standard deviation. Experience is measured in years.

F. ALTERNATIVE PARAMETER VALUES

This subsection discusses simulations with alternative sets of parameters. As predicted by the theoretical results in Sections 3 to 4 and Appendix E, the most important factors in determining the impact of taking into account the belief externality are the responsiveness of productivity relative to income, and the accuracy of employer information. Other factors, such as the level of the elasticity of taxable income, are less important.

Figure 16 shows the results of these alternative calibrations. First, panel (a) shows a simulation identical to the baseline exercise in Section 5 but with changes to ε_l and β to ensure that the elasticity of taxable income is lower, at $\varepsilon_z^{LR} = 0.8$, but that the ratio $\varepsilon_w^{LR}/\varepsilon_z^{LR}$ is approximately unchanged at 0.6. The level of the optimal tax schedule is higher, but the shift between the naïve and optimal schedules is qualitatively unchanged. This is despite there being a very large difference in ε_z^{LR} between the two scenarios.

Panel (b) of figure 16 shows the adjustment to marginal tax rates between the naïve and optimal tax schedules in four scenarios. The gray dotted line shows the baseline results from Section 5. Next, the black line shows the adjustment in the scenario from panel (a), with a lower income elasticity. The red line shows a third case with $\varepsilon_w^{LR}/\varepsilon_z^{LR} = 0.5$, but ε_z^{LR} unchanged. Finally, the blue line shows a fourth scenario with $dE(w)/dq = 0.9$, which implies a smaller belief externality. As expected, these latter two scenarios lead to a smaller adjustment to the tax schedule.

FIGURE 12: COMPARISON OF FISCAL EXTERNALITY

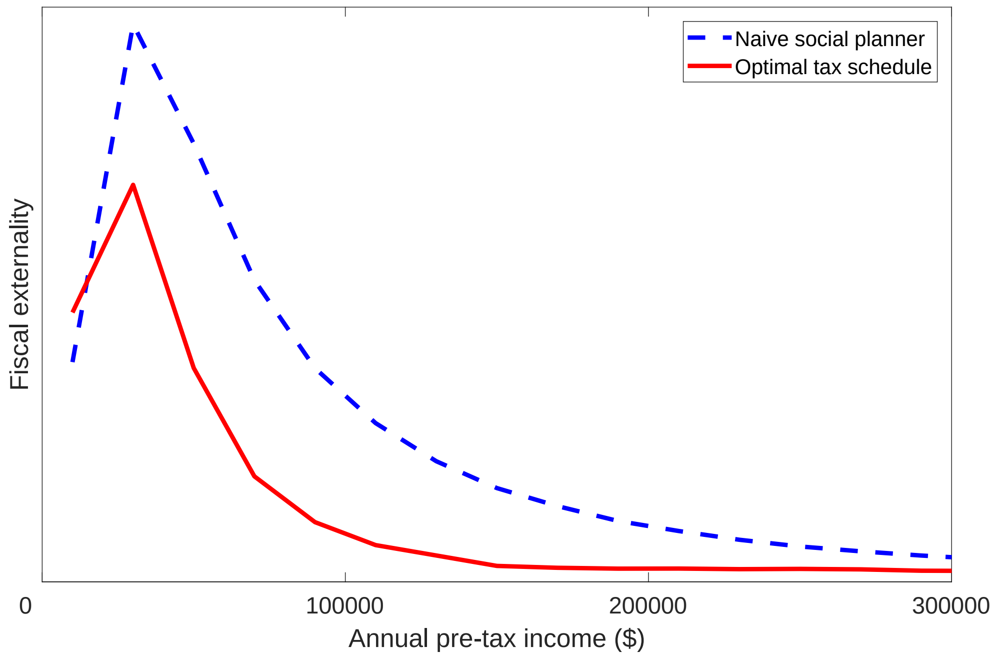


Figure notes. This figure compares the fiscal externality in each tax bracket under naïve and optimal taxation, in the simulation described in Section 5.

FIGURE 13: COMPARISON OF BELIEF EXTERNALITY

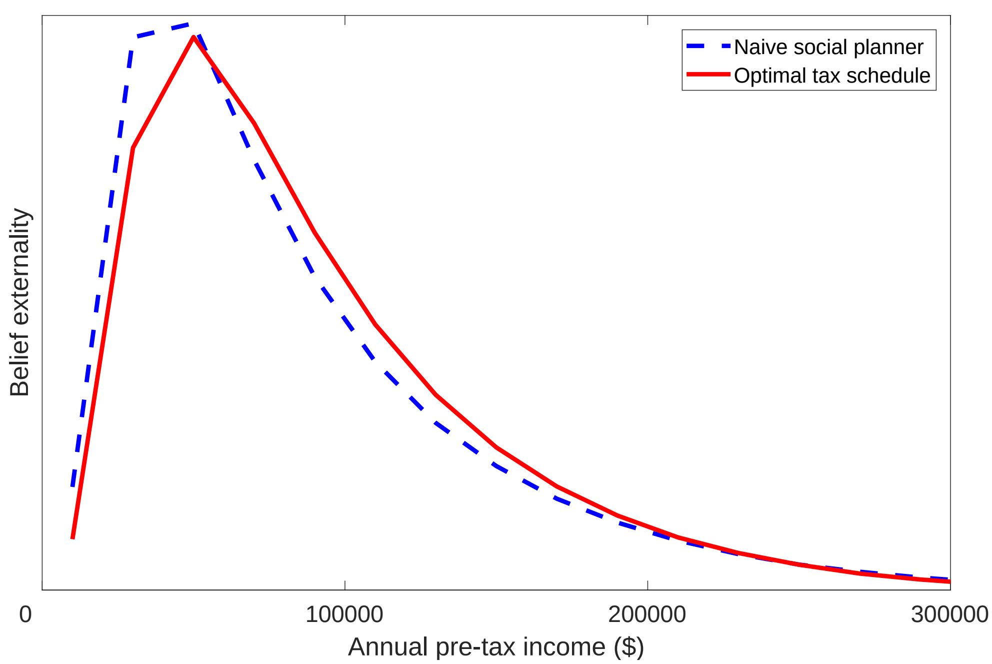


Figure notes. This figure compares the belief externality in each tax bracket under naïve and optimal taxation, in the simulation described in Section 5.

FIGURE 14: COMPARISON OF MECHANICAL EFFECT

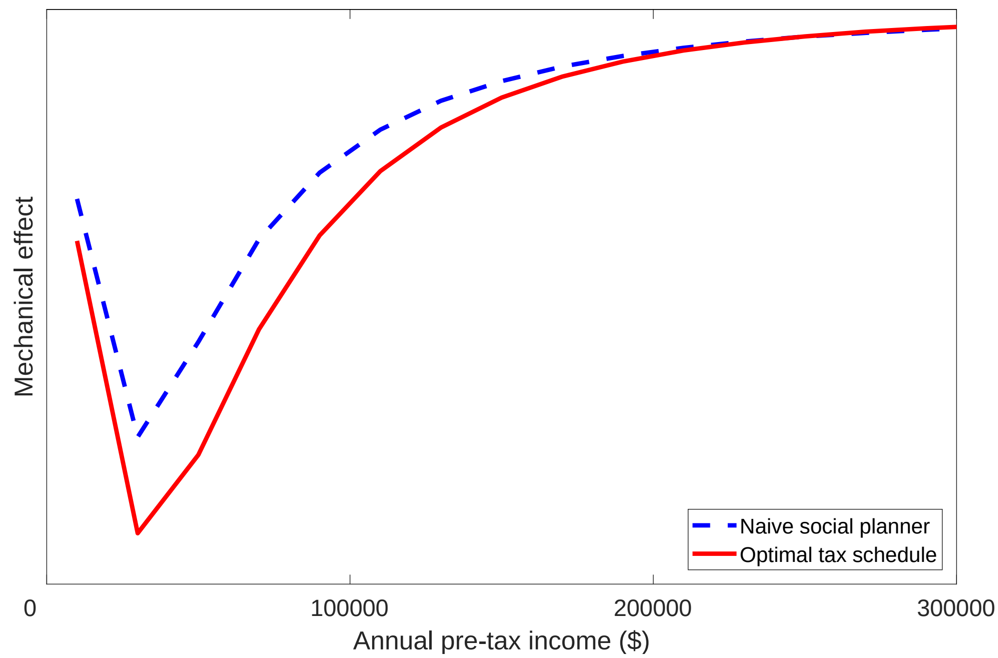


Figure notes. This figure compares the mechanical effect in each tax bracket under naïve and optimal taxation, in the simulation described in Section 5.

FIGURE 15: MARGINAL SOCIAL WELFARE WEIGHTS

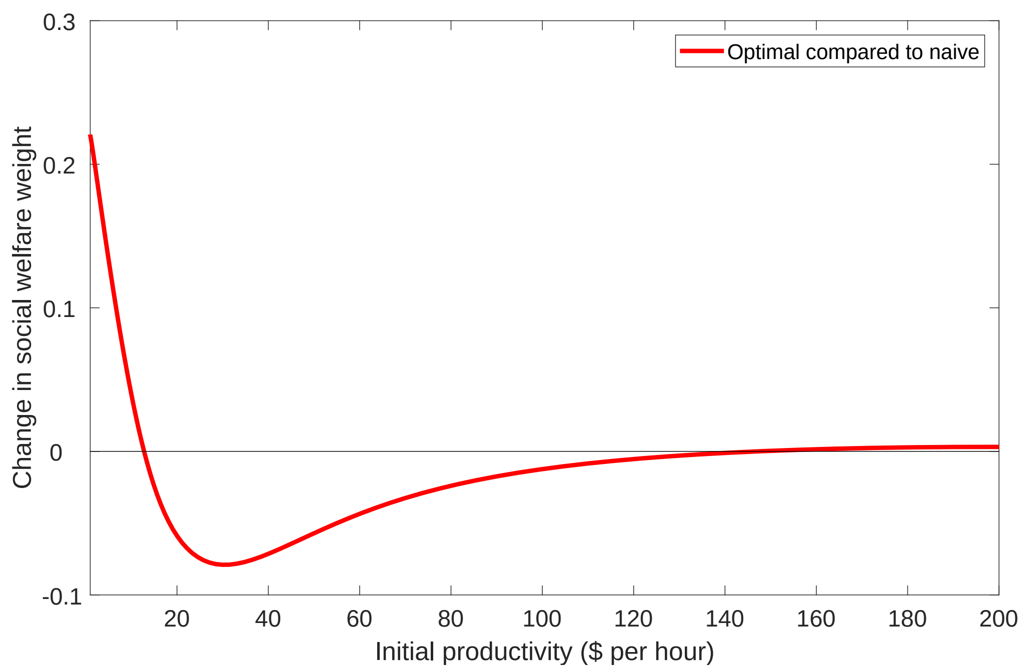
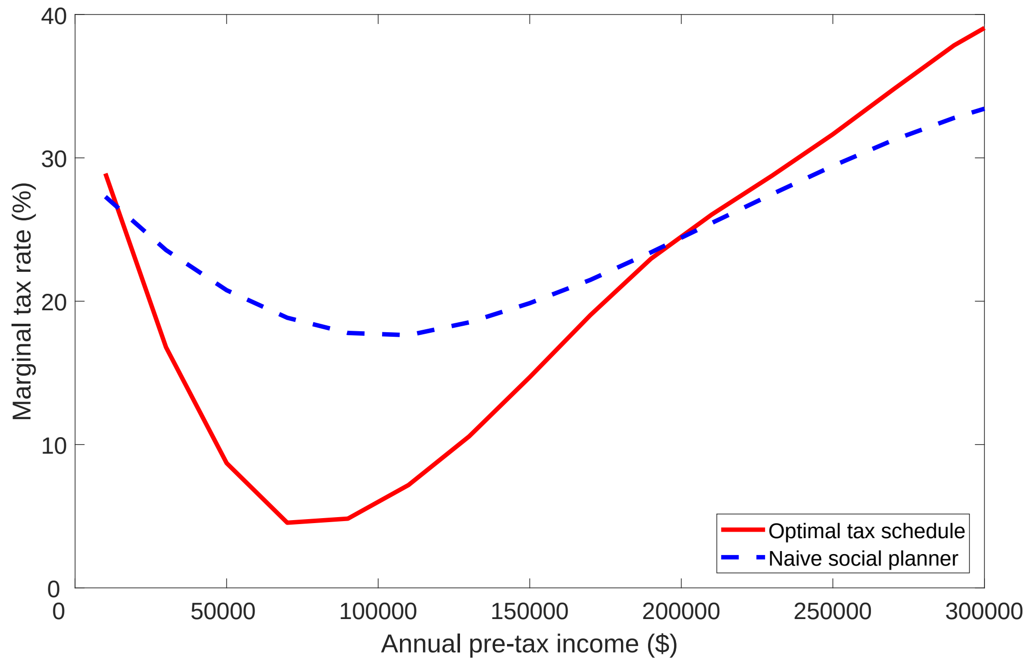


Figure notes. This figure plots the change in marginal social welfare weights starting from naïve taxation and transitioning to optimal taxation, for the simulation described in Section 5.

FIGURE 16: NON-LINEAR TAXATION WITH ALTERNATIVE PARAMETER VALUES

(a) Less Elastic Income ($\varepsilon_z = 0.8$)



(b) Tax Rate Adjustment in Alternative Scenarios

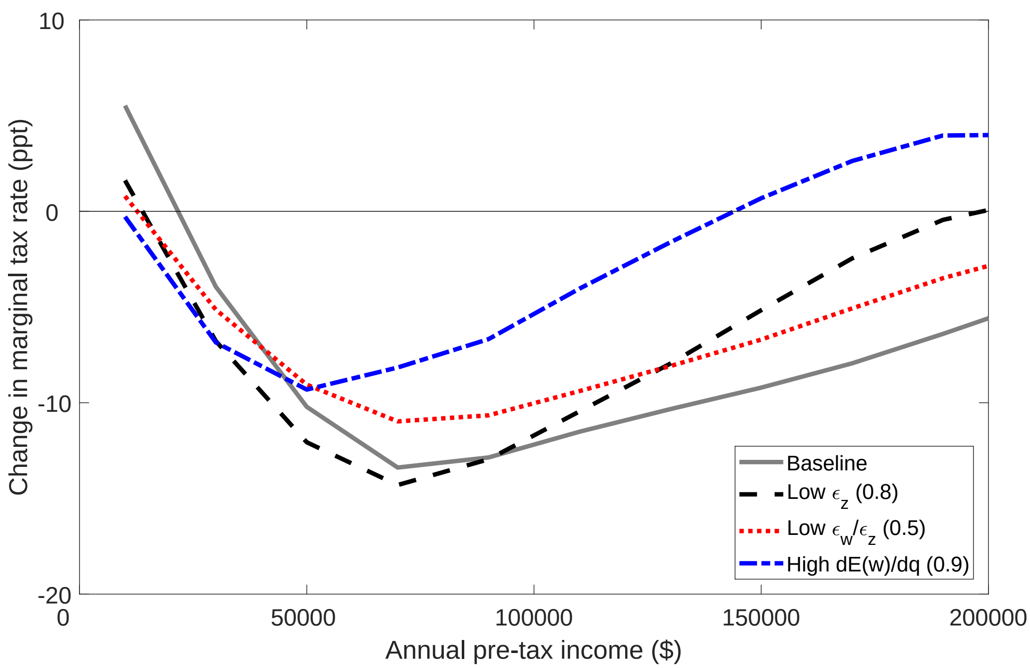


Figure notes. This figure shows the results of exercises similar to that described in Section 5 but with alternative parameter values. Panel (a) shows the optimal and naïve tax schedules when $\varepsilon_l = 0.2$ and $\beta = 0.25$, which still yields $\varepsilon_w^{LR}/\varepsilon_z^{LR} = 0.6$ but with $\varepsilon_z^{LR} = 0.8$. Panel (b) shows the adjustment to marginal tax rates in: (i) the baseline exercise; (ii) the alternative case with $\varepsilon_z^{LR} = 0.8$; (iii) a third case with $\varepsilon_w^{LR}/\varepsilon_z^{LR} = 0.5$ but ε_z^{LR} unchanged; and (iv) a fourth with $dE(w)/dq = 0.9$, which implies a smaller belief externality.