Optimal Income Taxation with Spillovers from Employer Learning

Ashley C. Craig*

December 2019

Abstract

I study optimal income taxation when human capital investment is imperfectly observable by employers. In my model, Bayesian employer inference about worker productivity compresses the wage distribution. Workers invest too little in human capital due to a wedge between the private and social returns to investment. The implied positive externality lowers optimal marginal tax rates. Its quantitative significance hinges on the accuracy of employers’ beliefs and the responsiveness of human capital investments to taxation. I calibrate the model to match empirical moments from the United States, including evidence on the speed of employer learning about labor market entrants. Taking into account the spillover from human capital investment introduced by employer inference reduces optimal marginal tax rates by 13 percentage points at 100,000 dollars of income, with little change in the tails of the income distribution. The welfare gain 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, belief formation, externalities, statistical discrimination, signaling

*Affiliation: University of Michigan, Ann Arbor (Department of Economics)
Email: ashcraig@umich.edu. Phone: +1 (734) 763-9561. Mail: 611 Tappan Avenue, 238 Lorch Hall, Ann Arbor MI 48109, United States.

Thanks: I am indebted to my advisors, Roland Fryer, Lawrence Katz, David Laibson and Stefanie Stantcheva. I am also grateful for valuable comments and suggestions from mentors, colleagues and interviewers, including: Lisa Abraham, Ole Agersnap, Joseph Altonji, Alex Bell, Raj Chetty, Oren Danieli, Adriano Fernandes, Edward Glaeser, Benjamin Golub, Nathaniel Hendren, James Hines, 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, Joel Slemrod, Aleh Tsyvinski and Matthew Weinzierl. Thanks also to participants at Harvard seminars, my job talks, the Young Economists Symposium, NBER Summer Institute 2019, and the National Tax Association Annual Conference.

Funding and Conflict of Interest: Financial support from the University of Michigan, Harvard University, 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

¹This 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.
²Viewed 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 optimal 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 (e.g., 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 private 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 small 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. 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 levels of income.

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 conditional on that signal, and less likely to send low signals where she had 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 to investment that may be positive or negative. In the extreme case in which productivity is entirely innate, the net externality is negative. More generally, however, less accurate employer information reinforces the positive component of the externality that arises from wage compression, and 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).7 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 decisions.

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 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 on-
going 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. Stantcheva (2014) and Bastani, Blumkin and Micheletto (2015) analyze the two-level screening problem that arises when labor disutility is related to a worker’s productivity, so that willingness to work long hours signals high ability. However, productivity is immutable in these models. Most similar in spirit, Hedlund (2018) studies bequest taxation in a model that features a similar belief externality, but with binary investment and 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 2019, 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

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 $\lim_{x \to 0} Q'(x) = \infty$.\(^8\)

Nature then distributes a signal of productivity to the worker and all firms. Specifi-

\(^8\)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.
Figure 1: Timeline of the Game

<table>
<thead>
<tr>
<th>Nature distributes investment cost</th>
<th>Firms see signal</th>
<th>Worker accepts highest wage</th>
<th>Payoffs realized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worker invests</td>
<td>Firms offer wages</td>
<td>Worker supplies labor</td>
<td></td>
</tr>
</tbody>
</table>

cally, let \( \theta \in \Theta \subseteq \mathbb{R}_+ \) be a signal with conditional density \( f(\theta|q) \), which is twice continuously differentiable in \( q \), and has full support for all \( q \). Let \( \theta = \sup(\Theta) \) and let \( f(\theta) \) be the marginal distribution of \( \theta \). I assume that \( f(\theta|q) \) is differentiable with respect to \( \theta \), and that the monotone likelihood ratio property holds: 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 \( \theta \) 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. 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 \( T \in T \subseteq C(\mathbb{R}_+, \mathbb{R}) \) is the exogenous tax system set by the social planner. In some parts of the paper, I will restrict \( T(z) \) to be twice continuously differentiable.

A. Worker and Firm Payoffs

The worker receives utility \( u(z - T(z), l) - klx \), 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.

B. 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: \( x : K \times T \to \mathbb{R}_+ \); \( A : K \times T \times \Theta \times \mathbb{R}_+^{|J|} \to J \); and \( L : K \times T \times \Theta \times \mathbb{R}_+^{|J|} \to \mathbb{R}_+ \). Each employer’s strategy maps signals and tax systems to wage offers \( O_j : \Theta \times T \to \mathbb{R}_+ \).

C. 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.

---

9Appendix A shows that this is isomorphic to a model with contracts specifying both labor supply and income, because the marginal rate of substitution between consumption and the quantity of labor supplied is independent of labor quality. If this is relaxed, employers may be able to use contracts to screen workers (see Stantcheva 2014), but workers still underinvest. A more important assumption is that employers and workers cannot write binding contracts on future outcomes at the time that workers invest.

10I use \( C(A, B) \) to denote the space of continuous functions mapping from \( A \) to \( B \).
D. Optimal Firm and Worker Behavior

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

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

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 \( \theta \) and the equilibrium distribution of productivity.

**Lemma 1.** Fix a value of \( \theta \) 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.\(^{11}\)

\[
l(\theta|\pi, T) \in L^* = \arg\max_{l_j \in \mathbb{R}^+} u\left(w(\theta|\pi) \bar{l} - T\left(w(\theta|\pi) \bar{l}\right), \bar{l}\right)
\]

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.

\[
v(\theta|\pi, T) = u\left(z(\theta|\pi, T) - T \left(z(\theta|\pi, T), \frac{z(\theta|\pi, T)}{w(\theta|\pi)}\right)\right)
\]

Evaluating the expectation of \( v(\theta|\pi, T) \) by integrating over \( \theta \), 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 \( q(k|\pi, T) \).

\[
x(k|\pi, T) \in X^* = \arg\max_{\bar{x} \in \mathbb{R}^+} \int_\Theta v(\theta|\pi, T) f(Q(\bar{x})) d\theta - k\bar{x}
\]

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

\[
w(\theta|\pi) = \frac{\int_K q(k|\pi, T) f(\theta|q(k|\pi, T)) dG(k)}{\int_K f(\theta|q(k|\pi, T)) dG(k)}
\]

\(^{11}\)Throughout the paper, optimal choices of labor supply and investment will be unique.
The monotone likelihood ratio property ensures that the equilibrium wage is strictly increasing in $\theta$, and that $V(q|\pi, T)$ increases with $q$.

E. Characterizing and Selecting Equilibria

Equations 1, 3 and 4 describe 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 $E(T)$. I consider a selection of these equilibria, defined by choosing one equilibrium $E^\dagger(T) \in E(T)$ for each $T$. The expected utility of a worker with investment cost $k$ is then defined as her expected utility given the tax schedule and this selection: $\bar{V}(k, T) = V(k, E^\dagger(T), T)$. 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 in my exposition of 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 more intuitive terminology throughout much of the paper.

F. 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 $W(T)$, which is defined as the average across types of the worker’s expected utility levels, $\bar{V}(k, T)$, after they have been transformed by a social welfare function $W$.

$$\max_{T \in T} W(T) = \int_K W(\bar{V}(k, T)) \, dG(k)$$

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

---

12 An alternative is to assume that the initial equilibrium is stable under a simple dynamic adjustment process. I define stability in Appendix B, and show how it ensures that investment responses to changes in $T$ can be written as the sum of an infinite sequence in much the same way as in Sachs et al. (2019).

13 I omit profits from welfare because they are zero in expectation. Allowing welfare to depend on realized utilities makes no qualitative difference, although the planner may then disrespect individual preferences.
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 $T(z)$ 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.

$$\max_{T \in T} W(T) = W(\overline{V}(k, T)) \, dG(k)$$

where:

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

$$x(k|\pi, T) \in \arg\max_{\bar{x} \in \mathbb{R}^+} \int_{\Theta} v(\theta|\pi, T) f(\theta|Q(\bar{x})) \, d\theta - k \bar{x}$$

$$l(\theta|\pi, T) \in \arg\max_{\bar{l} \in \mathbb{R}^+} u(w(\theta|\pi) \bar{l} - T(w(\theta|\pi) \bar{l}, \bar{l}))$$

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

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

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.\(^{14}\)

### 3 A Simple Example with Linear Taxation

I begin with an example in which the planner is restricted to choosing a linear tax, $\tau$.\(^{15}\)

Each worker’s consumption is then an average of her own income, $z$, and the mean income, $\bar{z}$: $c = (1 - \tau) z + \tau \bar{z}$. The example highlights an effect that is not present in classic models of income taxation: when productivity falls in response to higher taxes, employers

\(^{14}\)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.

\(^{15}\)For simplicity only, I also assume that the government’s revenue requirement, $R$, is zero.
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.\(^{16}\)

\[
\begin{align*}
  u &= c - l^{1 + \frac{1}{\epsilon_l}} \left/ (1 + 1/\epsilon_l) \right. \\
  q &= x^\beta
\end{align*}
\]

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 \(\theta\) and productivity \(q\) is: \(\ln \theta = \ln q + \ln \xi\), where \(\ln \xi \sim \mathcal{N}(0, \sigma_\xi^2)\). Secondly, investment costs \(k\) are distributed log-normally: \(k \sim \mathcal{LN}(\ln \mu_k - \sigma_k^2/2, \sigma_k^2)\).

B. EQUILIBRIUM

Given any tax rate \(\tau\), there is an equilibrium in which productivity and income are both log-normally distributed. In this equilibrium, a worker’s wage is a weighted geometric average of her own productivity \(q\), average productivity \(\mu_q\), 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 the extent that variation in it reflects differences in productivity.

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

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

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

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 \(\theta\) comes from noise (\(\sigma_\xi^2\) 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 \(\sigma_\xi^2\) is small, then \(s\) is close to one, and the private return to investment is close to the social return.

\(^{16}\)I assume \(\beta (1 + \epsilon_l) < 1\) so that investment returns are concave and choices finite (see Appendix G).
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 $\tau$, workers invest more, and $\mu_q$ rises; this further increases investment returns, and amplifies the response of productivity to taxation. The two effects of more or less 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 ($\varepsilon_q$) and income ($\varepsilon_z$) with respect to the retention rate $1 - \tau$ are:

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

**C. Optimal Taxation**

Building on Lemma 2, Proposition 2 provides a formula for the optimal linear tax, $\tau^*$. In equation 11, $\psi_k = W' (\bar{V}(k, \tau))$ is the marginal social welfare weight placed on workers with cost $k$, and $\bar{\psi}$ is the average welfare weight. The average income for individuals with cost $k$ is $z_k$, and $\bar{z}$ is the average income across workers of all types.

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

$$
\frac{\tau^*}{1 - \tau^*} = \frac{1 - \gamma - \gamma (1 - s) \varepsilon_q}{\varepsilon_z} \geq 0
$$

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

Equation 11 is similar to the optimal tax formula in the standard setting 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.

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

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.

\[
FE = -\tau \overline{\psi} \varepsilon z \frac{z}{1 - \tau}
\]

When workers re-optimize in response to a change in \( \tau \), 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.

\[
BE = -E_k (\psi_k z_k) (1 - s) \varepsilon_q
\]

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, \( 1 - s \). This wedge is large when employers’ information about their workers is imprecise. Second, it scales with \( \varepsilon_q \), because the externality arises from workers becoming more productive. Third, the welfare impact scales with \( E_k (\psi_k z_k) \), because higher income individuals – who supply more labor – are most-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- 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 11 collapses to the standard case.

\[
\frac{\tau^*}{1 - \tau^*} \bigg|_{s=1} = \frac{1 - \gamma}{\varepsilon_z}
\]

While it is critical that \(\varepsilon_z\) 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., \(\psi_k = 1\forall k\)).

\[
\frac{\tau^*}{1 - \tau^*} \bigg|_{\psi_k=1\forall k} = -\frac{(1 - s) \varepsilon_q}{\varepsilon_z}
\]

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, \(\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon_z}\), maximizes government revenue, and is unchanged from the standard case because the highest-cost worker is unaffected by the belief externality.

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 productivity distributions differ by group, 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 \((\mu^D_k > \mu^A_k)\). 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

\(^{17}\)In this case, both the condition for optimal taxes and the elasticity \(\varepsilon_z\) are identical to the results of Bovenberg and Jacobs (2005) under the assumption that the planner cannot directly subsidize investment.
Figure notes. These figures show the effects of a reduction in $\tau$ 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 wage impacts in panel (a) measure how much a worker’s expected wage changes per percentage point of the tax rate reduction. The utility impacts in panel (b) are in dollars of after-tax income, but are scaled by the productivity density so that the area under each curve is proportional to the average impact.
reveal a positive wage gap between $A$ and $D$ workers with identical signals.

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

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, $\varepsilon_z$, may differ between groups; and second, the covariance between incomes and welfare weights, $\gamma$, may differ. 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 need not imply that a group-specific subsidy is optimal.

**Corollary 1.** The standard deviation of log wages ($\sigma_q^2$) and signal-to-noise ratio $s = \frac{\sigma_q^2}{\sigma_q^2 + \sigma_\xi^2}$ are pinned down by the following condition.

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

The optimal tax rate is thus independent of the level of costs $\mu_k$, and does not directly depend on the level of log productivity $\mu_q$.

This conclusion contrasts with subsidy programs suggested based on classic models of “self-fulfilling” statistical discrimination (Coate and Loury 1993). The strong prescription based on those models stems from three assumptions. First, the investment decision is binary. Second, the social benefit to investment is constant. Third, the two groups are assumed to be identical in their fundamentals. Any equilibrium with a lower rate of investment must then feature a larger externality. This externality underlies the potential for multiple equilibria, and would warrant a subsidy for the disadvantaged group.

Another classic example is to suppose that the cost distributions of groups $A$ and $D$ are identical but that the employer’s signal is noisier for group $D$. This also yields statistical

---

18For 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.

19See also Arrow (1973), Moro and Norman (2004), and Craig and Fryer (2017). These papers generally focus on equilibrium selection. I consider self-fulfilling disparities and multiple equilibria in Appendix D.
discrimination (see Phelps 1972, Aigner and Cain 1977, Lundberg and Startz 1983). In this case, the private return to investment is again lower for the disadvantaged group, and the belief externality is larger. The social planner therefore has a motive to set $\tau_D < \tau_A$.

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 private 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 $\varepsilon_z$, $\varepsilon_q$ and $\gamma$, such evidence provides guidance as to whether group-specific tax rates are theoretically warranted.

4 Non-linear Taxation

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 $dz$.\(^{20}\) 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.

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 ($\text{ME}$), a fiscal externality ($\text{FE}$) and – new to this model – a belief externality ($\text{BE}$).

A. 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, $E^*(T)$.\(^{21}\) 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.\(^{22}\)

\(^{20}\)My exposition parallels Saez (2001) but little changes with general perturbations (see Gerritsen 2016). Requiring $d\tau$ to be small abstracts from bunching and gaps from introducing a kink in the tax schedule.

\(^{21}\)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.

\(^{22}\)I discuss the existence of a locally unique selection of equilibria in Appendix B. Appendix C discusses why the planner may in some cases choose to locate at a singularity where these conditions break down.
**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.

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). Second, I assume that individuals’ second-order conditions for labor supply hold strictly (Assumption 2). As discussed by Saez (2001), this requires that $1 - T'(z) + \varepsilon c z T''(z) > 0$, where $\varepsilon c z$ 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 $T''(z) \geq 0$.\(^{23}\)

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

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

Third, I assume that investment returns are strictly concave, so 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 cost distributions and investment technologies such that condition 14 holds. It can also be relaxed,\(^{23}\)

\(^{23}\)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.
with the key requirement being that workers are not indifferent between two decisions. With finitely many cost types, this is satisfied generically; and with a continuum of cost types, the analysis is unchanged if it is violated for countably many cost types.

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

\[
- \frac{Q''(x)}{Q'(x)^2} > \frac{\int_\Theta v(\theta|\pi, T) \frac{\partial^2 f(\theta|q)}{\partial q^2} |_{q=Q(x)} d\theta}{\int_\Theta v(\theta|\pi, T) \frac{\partial f(\theta|q)}{\partial q} |_{q=Q(x)} d\theta}
\]  

(14)

**B. Mechanical Effect**

Subject to these regularity assumptions, there are three effects of the proposed perturbation. I begin with the mechanical effect (ME) of taxation. 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.

Assumptions 1 and 2 jointly ensure that income is strictly increasing in \( \theta \). 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 \( \theta \), and letting \( \psi(k) = W'(\bar{V}(k, T)) \), the mechanical gain in welfare is:

\[
\int_\Theta u_c(\theta) \int_K \psi(k) dG(k|\theta) d\theta \times \int_{\theta(z|\pi, T)}^{\bar{V}(k, T)} f(\theta) d\theta - \int_{\theta(z|\pi, T)}^{\bar{V}(k, T)} u_c(\theta) \int_K \psi(k) dG(k|\theta) f(\theta) d\theta
\]

Value of transfer to average worker

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:

\[
\text{ME}(z) = d\tau dz \times \{ \Psi(z) - H(z) \}.
\]  

(15)

Since \( \psi_z(z) \) is decreasing in \( z \), the welfare weight below any finite level of income is higher than the population weight.\(^{24}\) This in turn implies that \( \text{ME}(z) > 0 \). 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.

\(^{24}\)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 \( \theta \). Finally, \( u_c(\theta) \) is weakly decreasing and \( z(\theta|\pi, T) \) strictly increasing, so \( \psi_z(z) \) is strictly decreasing.
C. 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 16, 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. In turn, the income response for a given signal realization 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.

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

The derivatives in equation 16 are causal responses to this perturbation to the tax schedule. They are thus not directly related to the properties of objects such as the utility function. However, Appendix B shows how these equilibrium responses can be rewritten as infinite series in terms of fundamentals, in an analogous way to Sachs et al. (2019).

D. Belief Externality

The final effect of the perturbation is new to this model. When individuals re-optimize their investment decisions, they disregard the effect of this 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 a worker’s investment: an increase in productivity, and a shift in her signal distribution.

\[
\frac{dw(\tilde{\theta}|\pi)}{d[1-T'(z)]} f(\tilde{\theta}) = \int_{K} \left( \frac{dq(k|\pi, T)}{d[1-T'(z)]} \left( f(\tilde{\theta}|q(k|\pi, T)) \right) \right) dG(k)
\]

\[
+ \left[ q(k|\pi, T) - E(q|\tilde{\theta}, \pi) \right] \left( \frac{\partial f(\tilde{\theta}|q)}{\partial q} \bigg|_{q=q(k|\pi, T)} \right) dG(k)
\]

(17)
The first component of equation 17 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 own investment and productivity being unchanged.

The second component is the rent transfer effect, which is generally negative. The intuition behind this effect is related to job market signaling à la Spence (1973). As shown in Figure 4, a worker $i$ who invests more becomes more likely to send high signals, and thus to be pooled with workers who have higher productivity than herself. By sending more of these signals, since $q_i < E(q|\tilde{\theta}, \pi)$, she lowers the wages of other workers who send $\tilde{\theta}$. At the same time, investing more makes this worker less likely to send low signals where $q_i > E(q|\tilde{\theta}, \pi)$. This reduces the wages of other workers who send such low signals.

The productivity and rent transfer effects differ in both sign and incidence. As workers re-optimize, the productivity effect raises the equilibrium wages of workers whose signal distributions overlap most with those who increased their productivity. This gain is proportional to the level of the conditional signal distribution, $f(\tilde{\theta}|q)$. In contrast, the rent transfer effect reduces the wages of workers who send signals in regions where $f(\tilde{\theta}|q)$ changes the most. As Figure 5 suggests, this is a different set of workers: in terms of their productivity levels, they are likely to be less similar to those who changed their investment decisions than are the beneficiaries of the productivity effect.

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
**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 of a reduction in the marginal tax rate on income between $60,000 and $61,000. The baseline economy is the linear taxation example, calibrated to the United States. The wage and income distributions are the same in both panels. The effect at each productivity level is scaled by the density so that the area under each curve is proportional to the average wage change due to that component. The dotted line shows the direct impact of due to the change in each agent’s investment decision, holding beliefs constant. The solid line shows the total wage impact. Finally, the shaded area is the externality.
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 rent transfer 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, $\psi_z$. Third, the total impact can be calculated by integrating over the signal distribution.

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

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, the incidence of a wage change is irrelevant, and the belief externality is proportional to the difference between the average productivity increase and the average private gain from investment.

\[ -\text{BE}(z) \propto \frac{d\tilde{q}}{d(1 - T'(z))} - \int_K \left\{ \int_{\Theta} \left( \frac{df(\theta|q)}{d(1 - T'(z))} \right) \bigg|_{q=q(k|\pi, T)} w(\theta|\pi) d\theta \right\} dG(k) \]

The private gain to investment on the right-hand side depends on the information structure of the economy, while the social benefit on the left is independent of it.

More generally, the impacts of wage changes on social welfare are re-weighted in a way that is important in driving the shape of the optimal tax schedule. The weights are given by: $\Omega(\theta) = \psi_z(z(\theta|\pi, T)) \left[ 1 - T'(z(\theta|\pi, T)) \right] l(\theta|\pi, T)$. A given wage change is thus more important if it affects a worker who supplies a large amount of labor, and who receives significant social welfare weight. This implies larger weights at intermediate incomes, contributing to a “U”-shaped optimal tax schedule. Additionally, the weights are proportional to the retention rate, $1 - T'$. This amplifies the effects of other forces in the model, compounding the “U” shape that already arises from the usual trade-off between the mechanical effect of taxation and the fiscal externality (see Diamond 1998).

F. A Necessary Condition for Optimality

Bringing everything together, this perturbation leads to three effects: $\text{ME}(z)$, $\text{FE}(z)$ and $\text{BE}(z)$. Figure 6 shows these effects graphically. Just as in the example in Section 3, the
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.

If $T$ is optimal, the three effects must sum to zero for all $z$. Otherwise, there exists a change to the tax schedule that raises welfare. This is summarized in Proposition 3.

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:

\[
\text{ME}(z) + \text{FE}(z) - d\tau dz \int_Z \bar{z} \psi_z(\bar{z}) \left( \frac{1 - T' (\bar{z})}{1 - T'(z)} \right) \varepsilon_{w(\bar{z}),1-T'(z)} dH(\bar{z})
\]

where:

\[
\varepsilon_{w(\bar{z}),1-T'(z)} = \frac{dw(\theta (\bar{z}) | \pi)}{d[1 - T'(z)]} \times \frac{1 - T'(z)}{w(\theta (\bar{z}) | \pi)}.
\]

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

If condition 18 is zero, there is no first-order gain from perturbing the tax schedule and moving to an equilibrium near the status quo. Stated differently, following Hen-
dren (2016), this is a requirement that the cost and benefit of marginal policy changes be equated.

\[
\frac{1 - \Psi(z)}{1 - H(z)} = 1 + \frac{BE(z) + FE(z)}{1 - H(z)}
\]  

(19)

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 \(1 + \Psi(z) > 1 - H(z)\) per dollar of mechanical expenditure, while the cost in addition to the mechanical expenditure is the sum of the two externalities.

A limitation to Proposition 3 is that equation 18 is necessary for optimality at an equilibrium around which there exists a continuous selection from the equilibrium correspondence, but not at a point of discontinuity if one exists. Assumptions 1 to 3 ensure continuity for a generic tax schedule, but I show in Appendix C how one can construct examples in which the planner chooses to locate at a point at which equilibrium decisions and social welfare jump discontinuously in response to a perturbation in a particular direction. At any such point, the first-order approach breaks down, and the social planner simply locates the economy on the side of the discontinuity with higher welfare.

5 Quantitative Analysis

A. Evidence on Employer Learning

My next step is to assess the size of the belief externality using new and existing empirical evidence. First, I look to the employer learning literature for evidence on the speed at which employers learn about worker productivity. The dominant approach, pioneered by Farber and Gibbons (1996) and Altonji and Pierret (2001), posits that the econometrician observes a productivity correlate that employers do not. This is usually a score on the Armed Forces Qualification Test (AFQT) from before the worker entered the labor market.

Using workers’ AFQT scores and wages, these studies involve estimation of a version of equation 20, often using data from the National Longitudinal Survey of Youth (NLSY).

\[
\ln w = \alpha_0 + \rho_0 \text{AFQT} + \rho_1 \text{AFQT} \times \text{Experience} + \gamma_0 \text{Education} + \gamma_1 \text{Education} \times \text{Experience} + \lambda_0 \text{Experience} + \lambda_1 \text{Experience}^2 + \lambda_2 \text{Experience}^3 + X'\beta + \varepsilon
\]  

(20)

The typical finding is that \(\rho_1\) is strictly positive. This is interpreted as evidence that employers do not fully reward workers for their productivity at first, but that the reward
increases over time. A simultaneous finding that $\gamma_1 < 0$ further supports the hypothesis that learning is occurring: employers initially use education to gauge productivity, but they obtain more direct information as time progresses. As they observe workers and learn more, they rely less on pre-existing productivity 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 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 present discounted private and social returns to improving it. In other words, these results imply that the belief externality is large on average.

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 wages.

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. This implies that 10 to 40 percent of the social return accrues to others, which is exactly the statistic required to calibrate the size of the externality in 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. Consistent with learning, 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).

Finally, numerous studies uncover evidence of statistical discrimination, which is itself evidence of 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

$^{25}$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.

$^{26}$These data come from a firm in the United States, first analyzed by Baker, Gibbs and Holmstrom (1994).
(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 existing evidence on how employer learning varies with 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 20 by interacting the variables of interest with indicators $I_A = 1 \ (\text{AFQT} > m)$ and $I_B = 1 \ (\text{AFQT} \leq m)$ for whether a worker’s AFQT score is above or below the median, $m$.

$$
\ln w = \sum_{j \in \{A, B\}} \left\{ \rho_{0,j} \text{AFQT} + \rho_{1,j} \text{AFQT} \times \text{Experience} \\
+ \gamma_{0,j} \text{Education} + \gamma_{1,j} \text{Education} \times \text{Experience} \\
+ \lambda_{0,j} + \lambda_{1,j} \text{Exper.} + \lambda_2 \text{Exper.}^2 + \lambda_3 \text{Exper.}^3 \right\} \times I_j + X' \beta + \varepsilon
$$

I then estimate equations 20 and 21 using data from the NLSY79 survey.

The sample follows Arcidiacono et al. (2010). 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. 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 gain increases by with each year of experience.

---

27 Appendix Table H1 provides summary statistics for workers with high and low AFQT scores.
28 The relationship between log wages, AFQT and experience is approximately linear in this region.
Table 1: Heterogeneity in Employer Learning

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

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.
Below the median, there is strong evidence of learning. The weight on AFQT rises steeply with experience, and the weight on education falls. There is less evidence of learning above the median, where the coefficient on the interaction between AFQT and experience is close to zero. The large direct effect of AFQT in the upper half of the distribution suggests that the results are not driven by AFQT scores being unimportant at the high end. The less negative interaction between education and experience above the median suggests that learning is driving the difference in results. These conclusions are all robust to restricting the sample to workers with twelve or sixteen years of education.

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. Blomquist and Selin (2010) provide a short-run estimate using a difference-in-differences approach and a tax reform in Sweden. Their results suggest that around three quarters of the response of taxable income comes through wages.\(^{29}\) This is consistent with Trostel (1993), whose calibration suggests that 60 to 80 percent of the long run response of income to taxation comes from 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.\(^{30}\) 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 employers receive more information and wages begin to more closely track ability, *average* wages rise by seven percent given the same formal education. This rise in wages is consistent with a response of human capital investment to the higher return to 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.

---

\(^{29}\) This may be conservative since Blomquist and Selin cannot capture long-run human capital responses.

\(^{30}\) *Kibbutzim* are small collective communities in Israel.
Table 2: Calibrated and Implied Objects

<table>
<thead>
<tr>
<th>Assumed object</th>
<th>Assumption</th>
<th>Implied object</th>
<th>Implied value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social welfare function</td>
<td>$\log(E(U))$</td>
<td>Income elasticity</td>
<td>$\varepsilon_{LR}^I = 1.1$</td>
</tr>
<tr>
<td>Noise distribution</td>
<td>LN, $\text{var}(\theta</td>
<td>q) = 7q$</td>
<td>Wage elasticity</td>
</tr>
<tr>
<td>Labor supply elasticity</td>
<td>$\varepsilon_l = 0.25$</td>
<td>Productivity</td>
<td>Champernowne: $\lambda = 1.3,$</td>
</tr>
<tr>
<td>Production concavity</td>
<td>$\beta = 0.3$</td>
<td>distribution</td>
<td>$\alpha = 2.46, y_0 = 2.7$</td>
</tr>
<tr>
<td>Wage distribution</td>
<td>Pareto LN: $a = 2,$ $\sigma_q^2 = 0.56, \mu_q = 2.76$</td>
<td>External fraction</td>
<td>0.15 (average)</td>
</tr>
</tbody>
</table>

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.

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, $f(\theta|q)$, and then find a distribution for productivity that yields a wage distribution as close as possible to the target.\footnote{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.} 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 conservative end of the estimates provided by Kahn and Lange (2014). In line with my empirical results on heterogeneity in learning, I also ensure that there is a flatter relationship between productivity and wages 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 with $\text{var}(\theta|q)$ 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 $\varepsilon_l = 0.25$, in line with estimates of the intensive-margin labor supply elasticity (e.g., Chetty 2012). Second, I calibrate the equilibrium (“long run”) elasticity of each variable with respect to the retention rate, $1 - T'$. Although these elasticities cannot be directly assumed, they are closely connected to the concavity of the production function, $\beta$.\footnote{For instance, the wage elasticity in Section 3’s example is $\beta(1 + \varepsilon_l) / ([\varepsilon_l + \beta(1 + \varepsilon_l)]$. Given a fixed value of $\varepsilon_l$, the elasticity is therefore fully determined by $\beta$.} I choose a value for $\beta$ that produces long-run elasticities of wages and labor supply of 0.7 and 0.4 respectively. This implies that 60 percent of the long run response of taxable in-

\[31\]
\[32\]
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. Further details of the calibration procedure are available in Appendix H. 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. On average, a worker who increases her productivity by one dollar per hour receives an 85 cent increase in her expected wage.
come comes from labor productivity, which is at the low end of the estimates above.\textsuperscript{33} The same estimates imply an overall elasticity of taxable income of 1.1, which is comparatively large because it incorporates the long-run response of human capital investment.\textsuperscript{34}

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 $\bar{z}$ to $\overline{z}$.

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

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 each marginal perturbation is approximately zero. At this point, condition 18 of Proposition 3 approximately holds. I refer to this final tax schedule as optimal.\textsuperscript{35} Further details of this process and the specific simulation below 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 social planner who neglects to take into account the fact that part of the response of wages to a change in the tax system arises due to an externality. 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 the equivalent model with perfect employer information.

\textsuperscript{33}Long-run equilibrium responses vary with the tax system and over the income distribution. The statistics here are based on responses of aggregate wages and labor supply to a small change to a flat tax.

\textsuperscript{34}There is also a multiplier effect: more productive workers work more, and working more raises the return to investment. An elasticity of 1.1 is in line with Mertens and Montiel-Olea (2018), whose estimates are conservative since they cannot incorporate long-run human capital responses. I provide simulations with other parameter values in Appendix H. Holding $\varepsilon^Lq / \varepsilon^Lz$ constant, the level of $\varepsilon^Lz$ is not important to my conclusions about the importance of the belief externality.

\textsuperscript{35}Global optimality cannot be guaranteed, but following this process leads to substantial welfare gains.
Comparison of the naïve and optimal tax schedules 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), which involves solving for an allocation that satisfies a social planner who ignores the endogeneity of wages. 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. The only thing that the planner is unaware of is that part of the change in equilibrium wages arises due to a spillover that workers ignore when re-optimizing.

G. SIMULATION RESULTS

Figure 8 shows the resulting optimal and naïve tax schedules. The red line shows a tax schedule that satisfies equation 18, so that there is no 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. Marginal tax rates are generally much lower under the optimal than the naïve schedule, reflecting the fact that the belief externality provides the planner with an 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 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 itself.

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

---

36 The SCPE could be used as an alternative benchmark, but ignoring the endogeneity of wages is unappealing in a model with human capital, and the exercise would exaggerate welfare gains.
37 Technical details are provided in Appendix H and in the accompanying replication code.
38 The algorithm to find the naïve schedule is conceptually identical to that used for the optimal schedule.
39 Relative to some other simulations in the literature (e.g. Saez 2001), even the naïve tax schedule is lower. This is partly due to the larger income elasticity, which incorporates the response of human capital. A second important contribution comes from my assumption that agents are risk-neutral. This avoids income effects, and greatly simplifies the exposition, but allowing for risk aversion would raise the levels of both schedules.
40 The shape is further amplified because the belief externality scales with with the retention rate, $1 - T'(z)$. 33
**Figure 8: Optimal Non-linear Taxation**

(a) Non-linear Taxation

![Graph showing the optimal tax schedule and a naive social planner's tax schedule.](image)

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

![Graph showing the welfare gain from each effect.](image)

*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 naive 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.
of the time – implying a smaller fiscal externality from their re-optimization.\footnote{Comparisons of the mechanical effect, fiscal externality and belief externality for marginal perturbations starting from the naïve and optimal tax schedules are available in Appendix H.}

At very low incomes, optimal marginal tax rates are also higher. The main reason for this is subtle. As taxes are adjusted downward throughout most of the income distribution, most workers experience an increase in their expected utility levels, but those at the bottom do not (see Figure 9). As a result, relative welfare weights rise at the lowest incomes, which increases the mechanical welfare gain from raising marginal taxes at low incomes and collecting infra-marginal income from all but the lowest-income workers.

**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, along with a plot of the net transfer to workers at each initial productivity level.

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 with moderate levels of productivity experience large gains in utility. But workers with very low productivity are worse off because marginal tax rates are much lower on average. This means that the transfer to the lowest-income worker is five percent smaller. Workers with very high productivity are also hurt, due to higher tax rates at top incomes.

**H. Alternative Approaches to Calibration**

Although the first-order condition for the optimal tax schedule cannot be written directly in terms of simple sufficient statistics, I show in Appendix E that 60 percent of the gain
from optimal taxation can be obtained using a simple approximation, which is derived assuming that the incidence of the belief externality is local. In addition, I show in Appendix B how equilibrium responses can be expanded as infinite series in terms of fundamentals such as structural elasticities; this approach is similar to Sachs et al. (2019).

6 Unproductive Signaling

I have so far assumed that investment is productive in the sense that its sole effect is to increase productivity. In this Section, I introduce the possibility that investment plays an unproductive ‘pure’ signaling role as in Spence (1973). Investment then reflects both human capital accumulation and innate differences in productivity due to variation in the immutable ability levels of workers. In this case, the externality from investment may, in general, be more positive or more negative than in the model without innate ability.\footnote{It is hard to assess the importance of unproductive signaling empirically. Evidence from large-scale school reforms demonstrate substantial productive effects of formal education (Meghir and Palme 2005, Aakvik, Salvanes and Vaage 2010, Oreopoulos 2006), but there is also some evidence to suggest a role for pure signaling (Lang and Kropp 1986, Bedard 2001). See Lange and Topel (2006) for a discussion.}

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 cost 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 Cobb-Douglas combination of human capital \( h \) and innate ability \( n \). In turn, assume that the worker invests \( x \) to attain human capital \( h = x^{\beta} \). Finally, let inherent ability be negatively related to a worker’s investment cost: \( n = 1/k \). As before, assume that the distribution of ability and the conditional signal distribution are log-normal.

\[
\begin{align*}
    n &\sim \mathcal{LN}(\ln \mu_n - \frac{\sigma_n^2}{2}, \sigma_n^2) \\
    \ln \theta &\sim \ln x + \ln \xi \\
    \ln \xi &\sim \mathcal{N}(0, \sigma_\xi^2)
\end{align*}
\]

With these assumptions, there again exists an equilibrium in which income and productivity are log-normally distributed. Similar to the original example in Section 3, the elasticities of productivity and income depend only on the labor supply elasticity \( \varepsilon_l \), the
concavity of the production function $\beta$, and the relative importance of innate ability $\alpha$.

Most importantly, these elasticities do not vary with the tax rate.

**Proposition 4.** For any tax rate $\tau$, 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.

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

This nests the original example in Section 3 in which investment is purely productive: when $\alpha = 0$ so that $q = h$, the two elasticities $\varepsilon_q$ and $\varepsilon_z$ collapse to that case, and equation 22 collapses to equation 11. By contrast, if $\alpha = 1$ so that $q = n$, 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 other workers. The logic here is similar to the rent transfer effect in Section 4: holding fixed the decisions of others, a worker who invests more hurts other workers, because she is now pooled by employers with workers who have higher productivity than herself, thereby lowering the wages of those other workers.

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

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

(22)

where $s = \frac{\sigma^2_x}{\sigma^2_x+\sigma^2_z}$ and $\gamma = E_n \left( \frac{\psi_n \pi_n}{\psi^2} \right)$.

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

$$
\beta (1-\alpha) = s \left[ \frac{\alpha + \beta (1-\alpha)}{1+\alpha(1+\varepsilon_l)} \right] \iff so (1+\varepsilon_l) = (1-s) \varepsilon_q
$$

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 mag-
ntitude 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 motive for intervention.

An implication of the equations above is that noisier 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, holding fixed the degree to which productivity reflects human capital as opposed to pure signaling. In this sense, evidence of residual employer uncertainty (Lange 2007, Kahn and Lange 2014) suggests a more positive externality, and implies lower optimal tax rates than if employers had more accurate beliefs.

**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. This results in a deterministic equilibrium mapping from investment to wages, $w(x)$. Taking this as given, the worker’s investment problem is:

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

(23)

where:

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

(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:

$$
uc(z(k) - T(z(k)), l(k))[1 - T'(z(k))] l(k) w'(x(k)) = k
$$

(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 $k$.

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.

$$
\frac{Q_k(x(k), k)}{x'(k)} = \frac{w'(x(k)) - Q_x(x(k), k)}{x'(k)}
$$

(26)
If \( Q_k (x, k) < 0 \) 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 \( Q_k (x, k) > 0 \), 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.

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

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:

\[
FE (z) + ME (z) + \int \tilde{z} \psi (\tilde{z}) \left( \frac{1 - T'(\tilde{z})}{1 - T'(z)} \right) \varepsilon_{\tilde{x}(\tilde{z}), 1 - T'(z)} \left[ \varepsilon_{\text{Private} \tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})} - \varepsilon_{\text{Social} \tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})} \right] dH (\tilde{z}) = 0
\]

(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.

\[
\varepsilon_{\text{Private} \tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})} = w' (x(k)) \frac{x(k)}{w(k)} \quad \varepsilon_{\tilde{x}(\tilde{z}), 1 - T'(z)} = \frac{dx(\tilde{z})}{d\left[ 1 - T'(z) \right]} \frac{1 - T'(z)}{x(\tilde{z})}
\]

Note the similarity between conditions 18 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. The equation for the belief externality, \( BE (z) \), remains very similar to Section 4. There remain distinct productivity and rent transfer effects, with the change in the equilibrium wage given signal realization \( \tilde{\theta} \) given by equation 28.

\[
\frac{dw(\tilde{\theta})}{d\left[ 1 - T'(z) \right]} f(\tilde{\theta}) = \int_K \left( \frac{dx(k|\pi, T)}{d\left[ 1 - T'(z) \right]} \right) \left[ Q_x (x(k|\pi, T), k) f(\tilde{\theta}|x(k|\pi, T)) \right]
\]

(28)
\[ + [Q(x(k, π, T), k) - E(q|\tilde{θ}, π)](\frac{\partial f(\tilde{θ}|x)}{\partial x}\bigg|_{x=x(k|π, T)})]dG(k) \]

Rent transfer effect

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 innate 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 rent transfer effect then becomes less negative, and may even be positive, since a worker who considers increasing her investment has higher innate ability 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 could be expanded to provide new insights into these contexts and many others.

References


See Wagner, Banerjee, Mohanan and Sood (2019) for an example of how imperfect information distorts the return to quality in a healthcare setting.


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 \( C_j = \{ z_j, l_j \} \in \mathbb{R}_+ \times \mathbb{R}_+ = C \) 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: \( x: K \times T \rightarrow \mathbb{R}_+ \); and \( A: K \times T \times \Theta \times C \rightarrow J \). Each employer’s strategy is a a function that maps signals and tax systems to contract offers \( O_j: \Theta \times T \rightarrow C \). 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 \( \theta \), 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 \( \theta \) 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 \( l_j \in L_j^* = \arg\max_{l_j \in \mathbb{R}_+} u(w_j\tilde{l}_j - T(w_j\tilde{l}_j), \tilde{l}_j) \).

B Continuity and Stability

A. Continuity of Investment Responses

In this appendix, I discuss the conditions required for equilibrium indeterminacy to be avoided, and for a given equilibrium to shift continuously in response to the perturbations that I consider. I assume that there is a finite number of cost types, indexed by \( i = 1, \ldots, |K| \), let \( x \) be the vector of investment decisions \( x_i \), 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.

\[
\lambda_i(x, T) = Q'(x_i) \int_\Theta v(\theta|\pi, T) \frac{\partial f(\theta|q)}{\partial q} \bigg|_{q=q_i} d\theta - k_i = 0
\]
Differentiating $\lambda_i (x, T)$ 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:

$$\frac{\partial \lambda_i}{\partial x_j} (x) = \begin{cases} \lambda_{ii}^q + \lambda_{ij}^w & \text{if } i = j \\ \lambda_{ij}^w & \text{if } i \neq j \end{cases}$$  \hspace{1cm} (30)

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

$$\lambda_{ii}^q = Q''(x_i) \int_\Theta v (\theta | \pi, T) \frac{\partial f (\theta | q)}{\partial q} \bigg|_{q=q_i} d\theta + Q'(x_i)^2 \int_\Theta v (\theta | \pi, T) \frac{\partial^2 f (\theta | q)}{\partial q^2} \bigg|_{q=q_i} d\theta$$ \hspace{1cm} (31)

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

$$\frac{\partial w (\theta | \pi)}{\partial q_j} = \left( f (\theta | q_j) + \left[ q_j - w (\theta | \pi) \right] \frac{\partial f (\theta | q)}{\partial q} \bigg|_{q=q_j} \right) p (k_j)$$ \hspace{1cm} (33)

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

$$\mathcal{J}_{f,x} = \begin{bmatrix} \frac{\partial \lambda_1}{\partial x_1} (x) & \cdots & \frac{\partial \lambda_{|K|}}{\partial x_1} (x) \\ \vdots & \ddots & \vdots \\ \frac{\partial \lambda_1}{\partial x_{|K|}} (x) & \cdots & \frac{\partial \lambda_{|K|}}{\partial x_{|K|}} (x) \end{bmatrix}$$ \hspace{1cm} (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 $\theta$. 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 $f_i (x, T)$ is given by $df_i (x, T)$.

$$d\lambda_i (x, T) = Q' (x_i) \int_\Theta dv (\theta | \pi, T) \frac{\partial f (\theta | q)}{\partial q} \bigg|_{q=q_i} d\theta$$ \hspace{1cm} (35)

These derivatives can be stacked into a $|K| \times 1$ vector $d\lambda (x, T)$.

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

I next show that, if $\mathcal{J}_{f,x}$ 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 $\mathcal{J}_{f,x}$ is invertible, this clearly remains the case after a similarly small perturbation. In these two senses, invertibility of $\mathcal{J}_{f,x}$ is generic.

First, I construct a parameterized family of functions, $\tilde{Q}(x|c)$, where $c$ 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 $\tilde{Q}(x|c)$ 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 $\tilde{Q}(x|c)$ with $\tilde{Q}(0|c) = 0$, $\tilde{Q}'(x|c) > 0$, $\tilde{Q}''(x|c) > 0$ and $\lim_{x \to 0} \tilde{Q}'(x|c) = \infty$. This is always possible, since $r$ is small and $Q$ strictly concave.

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

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

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

Moreover, $\lambda^q_{ii}$ 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 $\mathcal{J}_{f,x}$ are unchanged.

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

\[
x_{k,t+1} \in X_{k,t+1}^* = \arg\max_{\bar{x} \in \mathbb{R}_+} \int_{\Theta} v (\theta | \pi_t, T) f (\theta | Q (\bar{x})) d\theta - k\bar{x}
\]  

(36)

where:

\[
v (\theta | \pi_t, T) = w (\theta | \pi_t, T) l (\theta | \pi_t, T)
\]

\[
l (\theta | \pi_t, T) \in L^* (\theta | \pi_t, T) = \arg\max_{\bar{l} \in \mathbb{R}_+} u (w (\theta | \pi_t) \bar{l} - T (w (\theta | \pi_t) \bar{l}), \bar{l})
\]

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 \(l (\theta | \pi_t, T)\) are both uniquely pinned down and vary continuously with other agents’ investment decisions.

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

D. Series Expansion, Direct and Indirect Effects

At a stable equilibrium, the investment response can be decomposed into direct and indirect effects of a tax change. First, let \(S\) be the diagonal matrix of second-order conditions:

\[
S = \begin{bmatrix}
S_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & S_{|K|}
\end{bmatrix}
\]

where:

\[
S_i = Q'' (x_i) \int_{\Theta} v (\theta | \pi, T) \frac{\partial f (\theta | q)}{\partial q} \bigg|_{q=q_i} d\theta + Q' (x_i)^2 \int_{\Theta} v (\theta | \pi, T) \frac{\partial^2 f (\theta | q)}{\partial q^2} \bigg|_{q=q_i} d\theta
\]

The response of investment can then be written as

\[
dx = -J_{f,x}^{-1} SS^{-1} d\lambda (x, T).
\]

Letting \(I\) be the identity matrix, the matrix \(J_{f,x}^{-1} S\) can be rewritten as the following Neumann series,
providing that series is convergent.

\[ J_{f,x}^{-1}S = \sum_{k=0}^{\infty} (I - S^{-1}J)^k \]

The matrix \( B = I - S^{-1}J_{f,x} \) captures the effect of a change in each worker \( i \)'s investment decision on the investment decision of each other worker \( j \).

Convergence of the Neuman series above corresponds to the case of stability discussed in part C. At any stable equilibrium, we can thus write the response of the vector of investment choices to a change in the tax schedule as the following infinite series.

\[
dx = -S^{-1}d\lambda(x, T) - \sum_{k=1}^{\infty} B^k S^{-1}d\lambda(x, T) \tag{37}
\]

The intuition here is similar to Proposition 1 of Sachs et al. (2019). The first term captures the partial equilibrium response of investment to a change in the tax schedule. The second term accounts for general equilibrium cross-wage effects.

Each term in the infinite series on the right-hand side of equation 37 captures a “round” of cross-wage effects. The first term measures the indirect effect of partial equilibrium investment responses on investment choices. The \( n \)th term then captures the successive impact of changes induced by round \( n - 1 \). At a stable equilibrium, each round is smaller than the last, and the series converges. The sum of all of these rounds of adjustment measures the total shift in equilibrium investments.

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.\textsuperscript{44} This greatly simplifies the analysis of this subset of issues, while providing insights that are conceptually general.

\section*{A. \textbf{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 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 \(\theta\). In equilibrium, firms’ prior beliefs coincide with the true equilibrium probability \(\pi\) 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 \(v(\theta|\pi, \tau) - k = u((1 - \tau)w(\theta|\pi) + \tau\bar{w}) - k\), where \(\tau\) is a linear income tax, and \(\bar{w} = \pi \omega\) is the average wage. If she did not invest, she receives \(v(\theta|\pi, \tau) = u((1 - \tau)w(\theta|\pi) + \tau\bar{w})\). 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 \(\theta\), the expected utilities for an investor \((\nu_q)\) and non-investor \((\nu_u)\) are given by equations 38 and 39.

\[
\nu_q(\pi|\tau) = \int_0^1 v(\theta|\pi, \tau) dF_q(\theta) - k 
\]

\[
\nu_u(\pi|\tau) = \int_0^1 v(\theta|\pi, \tau) dF_u(\theta)
\]

Since workers invest if their expected return is greater than their cost, this implies an investment rate of \(G(\beta(\pi|\tau))\) where \(\beta(\pi|\tau) = \nu_q(\pi|\tau) - \nu_u(\pi|\tau)\).

The final requirement of equilibrium is that workers invest at a rate that coincides with employers’ beliefs. This is embodied in equation 40, which states that the fraction of

\textsuperscript{44}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.

\[ \pi = G(\beta(\pi|\tau)) \]  

(40)

For a given tax rate \(\tau\), equation 40 defines a fixed point as shown in Figure C1. An employer belief \(\pi\), combined with the tax \(\tau\), pins down the investment return and an investment rate, \(G(\beta(\pi|\tau))\).

**Figure C1: Equilibria and Taxation**

(a) Determination of Equilibria \(\tau\)  
(b) Equilibrium Set for Each Value of \(\tau\)

*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, \(G(\beta(\pi))\), is plotted against the employer prior, \(\pi\). Any intersection between this line and the 45 degree line is an equilibrium. The arrows show the direction in which each equilibrium moves as \(\tau\) rises. Panel (b) shows the set of equilibria over a range of values of \(\tau\). 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.\(^{45}\) Proposition 6 provides sufficient conditions for there to be others. For example, the economy in Figure C1 has four equilibria: 0, \(E_1\), \(E_2\) and \(E_3\).

**Proposition 6.** Assume that \(\phi(\theta) = f_u(\theta) / f_q(\theta)\) is continuous and strictly positive on \([0, 1]\). If there exists \(\pi\) such that \(G(\beta(\pi|\tau)) > \pi\) then there are multiple solutions to condition 40.

\(^{45}\)Tweaking the assumptions so that \(G(0) > 0\) eliminates the equilibrium with zero investment.
Intuitively, these conditions are satisfied if the returns to investment are high enough, as ensured by a large value of $\omega$ or a low enough tax rate. In turn, this means there is some employer belief $\pi$ such that the fraction of investors given that belief, $G(\beta(\pi|\tau))$, is higher than $\pi$. Since $G(\beta(1|\tau)) = 0$, and the regularity assumptions ensure that $G(\beta(\pi|\tau))$ is continuous $\pi$, 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 $\tau$ causes $G(\beta(\pi|\tau))$ to shift down for every employer belief $\pi$. 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 C1.

For simplicity, I assume that agents play the planner’s preferred equilibrium, which ensures that investment and welfare always increase as $\tau$ is lowered.\footnote{The set of equilibria can alternatively be refined by requiring stability in the sense introduced in Appendix B. In this case, the dynamic adjustment process is: $\pi_{t+1} = G(\beta(\pi_t|\tau))$. Stability amounts to a requirement that the absolute value of the slope of $G(\beta(\pi|\tau))$ is less than one, which implies that investment falls when $\tau$ rises. In Figure C1, both the zero investment equilibrium and $E_2$ are unstable.} 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 C1, the black line traces out the Pareto-dominant equilibria.

Proposition 7. Assume that multiple values of $\pi$ satisfy equation 40 for a given tax rate $\tau$. Let $\pi_i$ and $\pi_j$ be two solutions. Welfare is higher for every worker under $\pi_i$ than $\pi_j$ if and only if $\pi_i > \pi_j$. Moreover, investment in the Pareto-dominant equilibrium increases as $\tau$ is lowered.

Next, to characterize optimal taxation, define $\varepsilon_z$ as the elasticity of average income with respect to the retention rate. Second, let $u_\theta'$ be the marginal utility of consumption of an individual who sends signal $\theta$ 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 $\tau$, 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 $\tau$ at which this condition holds would always be strictly positive. The belief externality $\bar{w}_z$ provides an efficiency motive for intervention and pushes toward lower tax rates.
Proposition 8. Fix a value of \( \tau \) and an investment rate \( \pi^* (\tau) > 0 \), which satisfies equation 40. If \( g (\beta (\pi^* (\tau) | \tau)) \beta' (\pi^* (\tau) | \tau) \neq 1 \) and \( \tau \) is optimal, then the following condition holds:

\[
\frac{\tau}{1 - \tau} = \frac{\nu_\tau - \varepsilon_z w_z}{\varepsilon_z}
\]

where \( \nu_\tau = (1 - \pi) \int_0^1 \tilde{u}_\theta [f_u (\theta) - f_q (\theta)] d\theta, \varepsilon_z \) is the elasticity of income to the retention rate \( 1 - \tau \), and \( w_z = \frac{1}{\omega} \int_0^1 \tilde{u}_\theta \left[ \frac{\partial w (\theta | \pi)}{\partial \pi} \right] [\pi f_q (\theta) + (1 - \pi) f_u (\theta)] d\theta \) is the belief externality.

Proposition 8 parallels the results from the linear tax example (Proposition 2) and nonlinear 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 \( \tau \) at the optimum, which is equivalent to invertibility of the Jacobian, \( J_{f,x} \), discussed in Appendix B. Graphically, it amounts to a requirement that \( G (\beta (\pi | \tau)) \) is not tangent to the 45 degree line in Figure C1. If it were tangent, then it would imply an upward discontinuity in the equilibrium correspondence as at \( \tau_B \) in Panel (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 C1, but the first order condition may also hold at \( C \). This a natural limitation of the first-order approach, which is not specific to this model.

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 C1. 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 \( \tau \) as we approach \( \tau_B \) from below and also as we approach \( \tau_B \) from above, so that \( \tau_B \) is the optimal tax rate. However, equation 41 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 $\lambda_A$ 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 $\tau_j$ 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 $\tau_j$ 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 $\theta$ 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 $\bar{v}_q^j (\pi_j | T)$ if she invests, and $\bar{v}_u^j (\pi_j | T)$ if she does not.

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

The model remains otherwise unchanged from Appendix C. Workers invest if the return, $\beta_j (\pi_j | T) = \bar{v}_q^j (\pi_j | T) - \bar{v}_u^j (\pi_j | T)$, 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 $\pi (T)$ be the set of pairs $(\pi_A, \pi_D)$ such that $\pi_j (T) = G (\beta_j (\pi (T) | T))$ for $j \in \{A, D\}$. The correspondence $\pi (T)$ suffices to characterize the set of equilibria for each tax schedule. I define a selection by choosing one equilibrium pair $\bar{\pi} (T)$ 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:
\[ W_j = \pi_j \bar{v}_j^q (\pi_j | T) + (1 - \pi_j) \bar{v}_j^q (\pi_j | T) - \int_0^1 \bar{v}_j^q (\pi_j | T) - \bar{v}_j^d (\pi_j | T) \, k \, dG_j (k). \]

Within each group, the same perturbation arguments apply and the condition required for \( \tau_j \) 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.

\[
\frac{\tau_j}{1 - \tau_j} = \frac{\bar{v}_{j, \tau} - \varepsilon_z \bar{w}_z^j}{\varepsilon_z}
\]
\[
\int_0^1 u'_{A, \theta} dF (\theta) = \int_0^1 u'_{B, \theta} dF (\theta)
\]

where \( \bar{v}_{j, \tau} = (1 - \pi_j) \int_0^1 \bar{u}_{j, \theta} [f_u (\theta) - f_q (\theta)] d\theta, \varepsilon_z \) is the income elasticity of group \( j \), and \( \bar{w}_z^j = \frac{1}{\omega} \int_0^1 \bar{u}_{j, \theta} \frac{\partial w_{\theta, \pi_j}}{\partial \pi_j} [\pi_j f_q (\theta) + (1 - \pi_j) f_u (\theta)] d\theta \) is the belief externality.

To build intuition, consider the case in which \( T_{A \rightarrow D} \) 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 \( \tau_j \) is then isomorphic to the model with a single group, so \( \tau_A = \tau_B \) and \( \pi_A = \pi_B \). Moreover, if condition 42 holds, equation 43 must as well. Starting from equal treatment (\( \tau_A = \tau_B \) and \( T_{A \rightarrow D} = 0 \)), there is therefore no first order gain from slightly changing the tax system. This implies that the planner would not want to set \( T_{A \rightarrow D} \neq 0 \), 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 \textit{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

This appendix provides a way of calculating an approximately optimal tax schedule given only a few measurable statistics. Two general principles underlie the approach. 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 $	ilde{\Omega}(z, \tilde{\theta})$ as the difference between the weight on externalities at income $z(\tilde{\theta}|\pi, T)$ and the weight at income $z$.

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

The belief externality can then be rewritten as an approximation, plus a covariance bias.

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

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

$$
\int_\Theta \frac{dw(\tilde{\theta}|\pi)}{d[1 - T'(z)]} f(\tilde{\theta}) d\tilde{\theta} = (1 - s(z)) \frac{d\bar{w}}{d[1 - T'(z)]} \tag{45}
$$

Bringing everything together, condition 18 can then be approximated by:

$$
FE(z) + ME(z) - (1 - s(z)) \psi_z(z) l(z) \left[ 1 - T'(z) \right] \frac{d\bar{w}}{d[1 - T'(z)]} = 0. \tag{46}
$$

An advantage of this equation is that it facilitates assumptions about how the belief externality varies with income without finding corresponding distributional assumptions. As in Section 4, the correction term in equation 46 is larger if: (i) investment is more responsive; (ii) workers capture little of their return to investment; or (iii) a worker supplies a large amount of labor, faces a low tax rate, and receives substantial welfare weight.

Figure E1 shows the results when equation 46 is implemented in my simulated economy. The optimal and approximately optimal tax schedules are similar at lower levels of income, but the approximation deteriorates at higher levels of income where the true impact of the externality is more disperse. Starting from the naïve benchmark, 60 percent of the gains from optimal taxation are achieved via the approximation.
Figure E1: 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 46.

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 47):

\[
w u_c(w l^*(w) - T(w l^*(w))) + u_l(w l^*(w) - T(w l^*(w))) = 0
\]

where \( l^*(w) = \arg\max_{l \in \mathbb{R}_+} u(wl - T(wl), l) \).

Next, I define \( \hat{\vartheta}(x) = v(w(x)|T) \), and let \( x_{FB}(k) = \arg\max_x v(Q(x,k)|T) - kx \) 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 \( k \in K \), \( \hat{\vartheta}(x) \) is strictly concave around \( x_{FB}(k) \).

Assumption 6. \( \exists \kappa > 0 \) such that \( \hat{\vartheta}''(x) \geq 0 \Rightarrow \hat{\vartheta}'(x) > \kappa \) 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 \( k \):

\[
uc(z(k) - T(z(k)), l(k)) [1 - T'(z(k))] l(k) w'(x(k)) = k
\]  

(48)

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 \( \pi \) 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 \( \tilde{u}(w) = u(w l^*(w) - T(w l^*(w)), l^*(w)) \) represent the utility that the worker receives from accepting wage \( w \) 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 \( P_{k, \theta} \leq 0 \) 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 \( P_{k, \theta} > 0 \) 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 \( P_{\text{MAX}} \).

\[
P_{\text{MAX}} = \max_w [E[q|\theta, \pi] - w] l^*(w) \text{ s.t. } u(w l^*(w) - T(w l^*(w)), l^*(w)) \geq u(T(0), 0)
\]

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.

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

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

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

$$= \left(\frac{\sigma_q^2}{\sigma_q^2 + \sigma_\xi^2}\right) \ln q + \left(\frac{\sigma_\xi^2}{\sigma_q^2 + \sigma_\xi^2}\right) \left(\ln \mu_q - \frac{\sigma_q^2}{2}\right) + \left(\frac{\sigma_q^2}{\sigma_q^2 + \sigma_\xi^2}\right) \ln \xi$$

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:

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

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

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

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

$$q = \left[\frac{\beta}{k} \left((1 - \tau)^{1+\varepsilon_l} \mu_q^{(1-s)(1+\varepsilon_l)}\right) E\left[\xi^{s(1+\varepsilon_l)}\right]\right]^\frac{1}{1-\beta s(1+\varepsilon_l)}$$
Similarly, we can derive the elasticity of income effect, and an effect via average productivity. Combining these yields the total elasticity.

\[ \ln q \sim \mathcal{N} \left( \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \ln \beta + \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \ln s + \frac{\beta (1 + \varepsilon_l)}{1 - \beta s (1 + \varepsilon_l)} \ln (1 - \tau) \right. \\
+ (1 - s) \frac{\beta (1 + \varepsilon_l)}{1 - \beta s (1 + \varepsilon_l)} \ln \mu_q + \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \ln \mathbb{E} \left[ \xi^s (1 + \varepsilon_l) \right] \\
- \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \left( \ln \mu_k - \frac{\sigma^2_k}{2} \right), \left( \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \right)^2 \sigma^2_k \right) \]

Finally, we can obtain expressions for \( \mu_q \) and \( \sigma^2_q \) by matching coefficients.

\[ \sigma^2_q = \left( \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \right)^2 \sigma^2_k \tag{49} \]

\[ \mu_q = \left\{ \frac{\beta s (1 - \tau)^{1 + \varepsilon_l} \mathbb{E} \left[ \xi^s (1 + \varepsilon_l) \right]}{\mu_k} \right\} \exp \left[ \left( 1 + \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \right) \frac{\sigma^2_k}{2} \right] \frac{\beta}{1 - \beta (1 + \varepsilon_l)} \tag{50} \]

Equation 49 implicitly pins down \( \sigma^2_q \) in terms of \( \sigma^2_k, \beta, \varepsilon_l \) and \( \xi^2 \). It is independent of \( \mu_k \). In turn, equation 50 characterizes \( \mu_q \) as a function of the same set of parameters plus \( \mu_k \). The elasticity of \( \mu_q \) with respect to \( \mu_k \) is \( -\beta / [1 - \beta (1 + \varepsilon_l)] \). \( \square \)

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.

\[ \sigma_q = \frac{dq}{d(1 - \tau)} \times \frac{1 - \tau}{q} = \left[ \frac{\partial q}{\partial (1 - \tau)} + \frac{\partial q}{\partial \mu_q} \frac{d\mu_q}{d(1 - \tau)} \right] \frac{1 - \tau}{q} \]

\[ = \left[ \frac{\beta (1 + \varepsilon_l)}{1 - \beta s (1 + \varepsilon_l)} + (1 - s) \frac{\beta (1 + \varepsilon_l)}{1 - \beta s (1 + \varepsilon_l)} \frac{\beta (1 + \varepsilon_l)}{1 - \beta (1 + \varepsilon_l)} \right] \]

\[ = \frac{\beta (1 + \varepsilon_l)}{1 - \beta (1 + \varepsilon_l)} \]

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

\[ \sigma_z = \frac{dz}{d(1 - \tau)} \times \frac{1 - \tau}{z} = \left[ \frac{\partial z}{\partial (1 - \tau)} + \frac{\partial z}{\partial \mu_q} \frac{d\mu_q}{d(1 - \tau)} + \frac{\partial z_i}{\partial \mu_q} \frac{d\mu_q}{d(1 - \tau)} \right] \frac{1 - \tau}{z} \]

\[ = \varepsilon_l + (1 + \varepsilon_l) \frac{\beta (1 + \varepsilon_l)}{1 - \beta (1 + \varepsilon_l)} \]

\[ = \varepsilon_l + \beta (1 + \varepsilon_l) \frac{1}{1 - \beta (1 + \varepsilon_l)} \]
Proof of Proposition 2. The utility of a worker with noise realization $\xi$ and cost $k$ is:

$$v = \frac{\left[(1-\tau)q\mu_{q}^{1-s}\xi^{1+\varepsilon_{l}}\right]^{1+\varepsilon_{l}}}{1+\varepsilon_{l}} - kx + \tau \bar{z}$$

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 $\xi$, the expected utility for an individual with cost $k$ is:

$$\left[\frac{1 - \beta s (1 + \varepsilon_{l})}{1 + \varepsilon_{l}}\right] \left((1-\tau)^{1+\varepsilon_{l}} \mu_{q}^{(1-s)(1+\varepsilon_{l})} \right) E \left[\xi^{s(1+\varepsilon_{l})}\right] q^{s(1+\varepsilon_{l})} + \tau \bar{z}$$

Then, substituting in the optimal choice of $q$, and weighting by the worker’s welfare weight $\psi_{k}$, we get expected welfare in terms of $\mu_{q}$ and $\xi$.

$$E_{\xi} [\psi_{k} v_{k,\xi}|k] = \psi_{k} \left[\frac{1 - \beta s (1 + \varepsilon_{l})}{1 + \varepsilon_{l}}\right] \left((1-\tau)^{1+\varepsilon_{l}} \mu_{q}^{(1-s)(1+\varepsilon_{l})} \right) E \left[\xi^{s(1+\varepsilon_{l})}\right]$$

$$\times \left(\frac{\beta s}{k}\right)^{\frac{1 - \beta s (1 + \varepsilon_{l})}{1 - \beta s (1+\varepsilon_{l})}} \left\{ E \left[\xi^{s(1+\varepsilon_{l})}\right] \right\}^{\frac{1}{1 - \beta s (1+\varepsilon_{l})}} + \psi_{k} \tau \bar{z}$$

$$= (1-\tau) \psi_{k} \bar{z} \left[\frac{1 - \beta s (1 + \varepsilon_{l})}{1 + \varepsilon_{l}}\right] + \psi_{k} \tau \bar{z}$$

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

$$E [\psi_{k} v_{k,\xi}] = (1-\tau) E [\psi_{k} \bar{z} \left[\frac{1 - \beta s (1 + \varepsilon_{l})}{1 + \varepsilon_{l}}\right] + \tau \bar{z}]$$

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, $\bar{z}$.

$$\text{FE} = \tau \psi \bar{z} \frac{\bar{z}}{1 - \tau}$$

Second, welfare rises due to the externality via employer beliefs. Specifically, differentiating with respect to $\mu_{q}$ and aggregating over $k$, the gain in social welfare is as follows.

$$\text{BE} = (1-s) E_{k} (\psi_{k} \bar{z} \left[1 - \beta s (1 + \varepsilon_{l})\right] \varepsilon_{q}$$
Finally, there is a mechanical welfare loss due to the transfer from the average worker to high-income workers:

$$ME = E_k (\psi_k \zeta_k) - \bar{\psi} \bar{z}$$

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

$$FE + ME + BE = \frac{\tau}{1 - \tau} \zeta_q \bar{z} + E_k (\psi_k \zeta_k) [1 + (1 - s) \zeta_q] \bar{z} - \bar{\psi} \bar{z}$$

Then setting this to zero yields the first order condition shown in the proposition. \(\square\)

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

$$\max_T W (T) = W (\bar{V} (k, T)) \, dG (k)$$

where:

$$\bar{V} (k, T) = \int_{\Theta} (v (\theta | \pi, T) - k x (k, \pi, T)) f (\theta, q (k | \pi, T)) \, d\theta$$

subject to:

$$x (k | \pi, T) \in \arg\max_{\tilde{x} \in \mathbb{R}^+} \int_{\Theta} v (\theta | \pi, T) f (\theta | Q (\tilde{x})) \, d\theta - k \tilde{x}$$

$$l (\theta | \pi, T) \in \arg\max_{\tilde{l} \in \mathbb{R}^+} u (w (\theta | \pi) \tilde{l} - T (w (\theta | \pi) \tilde{l}), \tilde{l})$$

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

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

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.

$$v (\theta | \pi, T) = u (w (\theta | \pi) l (\theta | \pi, T) - T (w (\theta | \pi) l (\theta | \pi, T)), l (\theta | \pi, T))$$  \hspace{1cm} (51)

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 \(z\).

$$WL = -dT dz \left\{ \int_{\theta (z | \pi, T)} \bar{\theta} u_c (\theta) \int_K \psi (k) dG (k | \theta) f (\theta) \, d\theta \right\}$$  \hspace{1cm} (52)
Since the revenue raised is returned to all individuals equally via an increase in the intercept of the tax schedule, it is worth $\lambda$ per dollar in terms of social welfare, where:

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

(53)

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

$$WG = d\tau dz \left\{ \int_{\theta(z|\pi,T)} f(\theta) d\theta \right\} \lambda. \quad (54)$$

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

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

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 $\lambda$, the total fiscal externality is as follows.

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

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 $\theta$ is as follows.

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

Applying the envelope theorem and again dividing by $\lambda$, 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 $\theta$.

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

To obtain the total belief externality shown in the main text, we then integrate over the distributions of $\theta$ 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 $FE + BE + ME \neq 0$, 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.

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

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 $\xi$.

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

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

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

where \( s = \sigma_q^2 / (\sigma_q^2 + \sigma_x^2) = \sigma_x^2 / (\sigma_x^2 + \sigma_\xi^2) \). A worker’s expected level of productivity is therefore a geometric weighted average of \( A, x, \xi \) and \( \mu_q \).

\[ w = \tilde{\theta}^s \mu_q^{1-s} = A^s x^s B^s \xi^s \mu_q^{1-s} \]

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

\[ (1 - \tau) z = (1 - \tau)^{1+\varepsilon_l} w^{1+\varepsilon_l} = (1 - \tau)^{1+\varepsilon_l} \left[ A^s x^s B^s \xi^s \mu_q^{1-s} \right]^{1+\varepsilon_l}. \]

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

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

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.

\[ \ln x = \frac{1}{1 - sB(1 + \varepsilon_l)} \left[ \ln n + \ln (sB) + (1 + \varepsilon_l) \ln (1 - \tau) + (1 - s) (1 + \varepsilon_l) \ln \mu_q \right. \\
\left. + \ln E \left[ \xi^{s(1+\varepsilon_l)} \right] + s (1 + \varepsilon_l) \ln A \right] \]

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 $\ln x$.

\[
\ln x = \frac{1 + s \alpha (1 + \xi_l)}{1 - s \beta (1 - \alpha) (1 + \xi_l)} \ln (n) + \frac{1}{1 - s \beta (1 - \alpha) (1 + \xi_l)} \left[ \ln s + \ln \left( \frac{\alpha + \beta (1 - \alpha)}{1 + s \alpha (1 + \xi_l)} \right) \right] + (1 + \xi_l) \ln (1 - \tau) + (1 - s) (1 + \xi_l) \ln \mu_q + \ln E \left[ \xi^n(1 + \xi_l) \right]
\]

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 $\mu_q$ and $\sigma_q^2$ can be found by matching coefficients.

\[
\sigma_q^2 = \left[ \frac{\alpha + \beta (1 - \alpha)}{1 - \beta s (1 - \alpha) (1 + \xi_l)} \right]^2 \sigma_n^2
\]

\[
\ln \mu_q = \left[ \frac{\alpha + \beta (1 - \alpha)}{1 - (1 - \alpha) \beta (1 + \xi_l)} \right] \ln n + \left[ \frac{\beta (1 - \alpha)}{1 - (1 - \alpha) \beta (1 + \xi_l)} \right] \ln s + \ln \left[ \frac{\alpha + \beta (1 - \alpha)}{1 + s \alpha (1 + \xi_l)} \right] + (1 + \xi_l) \ln (1 - \tau) + \ln E \left[ \xi^n(1 + \xi_l) \right]
\]

\[
\ln \mu_q = \frac{\sigma_q^2}{2} \left[ \frac{\alpha + \beta (1 - \alpha)}{1 - \beta s (1 - \alpha) (1 + \xi_l)} \right] \left[ \frac{\alpha + \beta (1 - \alpha)}{1 - (1 - \alpha) \beta (1 + \xi_l)} \right]
\]

The elasticity of productivity follows directly.

\[
\frac{d \ln \mu_q}{d \ln (1 - \tau)} = \left( \frac{\beta (1 - \alpha) (1 + \xi_l)}{1 - \beta (1 - \alpha) (1 + \xi_l)} \right)
\]

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

\[
\frac{d \ln z}{d \ln (1 - \tau)} = \frac{\partial z}{\partial (1 - \tau)} + \frac{\partial \ln z}{\partial \ln q} \frac{\partial \ln q}{\partial \ln (1 - \tau)} + \frac{\partial \ln z}{\partial \ln \mu_q} \frac{d \ln \mu_q}{d \ln (1 - \tau)}
\]

\[
= \xi_l + (1 + \xi_l) \left[ \frac{\beta (1 - \alpha) (1 + \xi_l)}{1 - \beta (1 - \alpha) (1 + \xi_l)} \right] s + \frac{\beta (1 - \alpha) (1 + \xi_l)}{1 - \beta (1 - \alpha) (1 + \xi_l)} (1 - s)
\]

\[
= \xi_l + (1 + \xi_l) \frac{\beta (1 - \alpha)}{1 - \beta (1 - \alpha) (1 + \xi_l)}
\]

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

\[
\bar{v}_n = n \frac{s(1 + \xi_l)(\alpha + \beta(1 - \alpha))}{1 - s \beta (1 - \alpha) (1 + \xi_l)} (1 - \tau)^{(1 + \xi_l)} \mu_q^{(1 - s)(1 + \xi_l)} E \left[ \xi^n(1 + \xi_l) \right] \left[ \frac{1}{1 - s \beta (1 - \alpha) (1 + \xi_l)} \right] \left[ \frac{\beta s (1 + \xi_l)(1 - \alpha)}{1 - \beta s (1 - \alpha) (1 + \xi_l)} \right] \left[ \frac{1}{1 + \xi_l} \right] + \tau^\pi
\]

71
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.

\[
(1 - \tau) z_n = n^{s(1 + \varepsilon_l)/[\alpha + \beta (1 - \alpha)]} \left[ (1 - \tau) \mu_{q(1 - s)} \right]^{1 + \varepsilon_l} \frac{1}{1 - \beta s(1 - \alpha)(1 + \varepsilon_l)} E \left[ \varepsilon^{s(1 + \varepsilon_l)} \right]^{1 - \beta s(1 - \alpha)(1 + \varepsilon_l)} \left[ sB \right]^{\beta s(1 + \varepsilon_l)/(1 - \alpha)} \frac{1}{1 - \beta s(1 - \alpha)(1 + \varepsilon_l)}
\]

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

\[
\psi_n \tau_n = (1 - \tau) \psi_n z_n^{1 - (1 + \varepsilon_l) sB} + \tau \psi_n \tau
\]

Differentiating \( \psi_n \tau_n \) 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 \( \tau_n \). Then taking the expectation over ability types, \( n \), we obtain:

\[
\text{MEU} = E [\tau_n \psi_n] \left[ \frac{1}{1 + s \alpha (1 + \varepsilon_l)} \right] - \overline{\psi} \tau
\]

Next, we can calculate the belief externality. This is again captured by the effect via \( \mu_{q} \).

Using the elasticities from Proposition 4 and the expression for \( \tau_n \), the effect on the welfare of a worker with ability \( n \) is:

\[
\frac{(1 + \varepsilon_l) (1 - s)}{1 - \beta s (1 - \alpha) (1 + \varepsilon_l)} \frac{\tau_n - \tau \overline{\psi}}{1 - \beta (1 - \alpha) (1 + \varepsilon_l)} \frac{\beta (1 - \alpha) (1 + \varepsilon_l) \mu_{q}}{1 - \beta (1 - \alpha) (1 + \varepsilon_l) (1 - \tau)}
\]

Weighting by \( \psi_n \), using the expression for \( \tau_n \) and taking the expectation over ability types, this gives us the total belief externality.

\[
\text{BE} = (1 - s) E [\tau_n \psi_n] \left[ \frac{1}{1 + s \alpha (1 + \varepsilon_l)} \right] - \frac{\beta (1 - \alpha) (1 + \varepsilon_l)}{1 - \beta (1 - \alpha) (1 + \varepsilon_l)}
\]

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

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

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

\[
\frac{\tau}{1 - \tau} = \frac{1 - E_n \left( \frac{\bar{p}_n \bar{w}_n}{q} \right)}{1 + s \alpha (1 + \epsilon_l) \frac{(1+\epsilon_l)\beta(1-\alpha)}{1-\beta(1-\alpha)(1+\epsilon_l) - \epsilon_l + (1+\epsilon_l)\beta(1-\alpha)}} \left[ 1 + (1 - \epsilon_l) \right]
\]

Proof of Lemma 3. Firm beliefs about the distribution of productivity in the population must be confirmed in equilibrium and identical across firms. Let \( \pi \) denote the equilibrium set of beliefs. Firm \( j \)'s expectation of the worker's productivity is

\[
E_n (z_n | \psi_n) \left[ 1 + (1 - s) \alpha \right] \left[ 1 - \beta (1-\alpha)(1+\epsilon_l) \right]
\]

Finally, let \( \tilde{u} (C_j) = u (z_j - T (z_j), l_j) \) 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 \( \tilde{u} (C_j) \geq \tilde{u} (C_k) \) 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: \( \tilde{u} (C_j) > \tilde{u} (C_k) \) 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: \( \tilde{u} (C_j) = \tilde{u} (C_k) \) for all \( C_k \), and \( P_{k, \theta} \leq 0 \) for some \( k \).

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

Case 3: \( \tilde{u} (C_j) = \tilde{u} (C_k) \) and \( P_{k, \theta} > 0 \) for all \( k \).

Since the worker always accepts an offer, \( E [q|\theta, \pi, C_j] \) is bounded weakly below \( E [q|\theta, \pi] \) for at least one firm. This firm’s expected profit is bounded below \( P_{\text{MAX}} \).

\[
P_{\text{MAX}} = \max_{l, z} E [q|\theta, \pi] l - z \quad \text{s.t.} \quad u (z - T (z), l) \geq u (T (0), 0)
\]

The assumptions on the worker’s utility function ensure that this yields finite labor supply for any finite \( E [q|\theta, \pi] \). Since \( z_j \) is restricted to be greater than zero and \( E [q|\theta, \pi] \) is finite, \( P_{\text{MAX}} \) 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:

\[
 L^* = \arg\max_{l_j} \left( E[q|\theta, \pi] l_j - T \left( E[q|\theta, \pi] l_j \right) , l_j \right). 
\]

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:

\[
 L^*_j = \arg\max_{l_j} \left( E[q|\theta, \pi, A_j = 1] l_j - T \left( E[q|\theta, \pi, A_j = 1] l_j \right) , l_j \right). 
\]

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:

\[
 u \left( w_j l^*_j - T \left( w_j l^*_j \right) , l^*_j \right) \geq u \left( w_k l^*_k - T \left( w_k l^*_k \right) , l^*_k \right) .
\]

Suppose now that \( u \left( w_j l^*_j - T \left( w_j l^*_j \right) , l^*_j \right) \leq u \left( w_k l^*_k - T \left( w_k l^*_k \right) , l^*_k \right) \). Then firm \( j \) can alter its offer to \( z_j = w_k l^*_k < w_j l^*_j \) and set \( l_j \) below but arbitrarily close to \( l_k \). Firm
with the fact that better off. Thus, welfare is higher for all workers. The accompanying change in individual investment decisions can only make those marginal individuals better off. Thus, welfare is higher for all workers.

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 \( \pi \), that \( w(\theta|0) = 0 \) for all \( \theta \) and that \( w(\theta|1) = \omega \) for all \( \theta \). 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 \( \pi^* \) then the continuity of \( \phi(\theta) \) and \( G \) combined with the fact that \( G(\beta(1|\tau)) = 0 \) implies that \( G(\beta(\tilde{\pi}|\tau)) = \tilde{\pi} \) for some \( \tilde{\pi} > \pi^* \). There are therefore at least two solutions to equilibrium condition 40.

Proof of Proposition 7. Social welfare is given by:

\[
\pi \overline{v}_q(\pi) + (1 - \pi) \overline{v}_u(\pi) - \int_0^1 \overline{v}_q(\pi) - \overline{v}_u(\pi) \, kdG(k). 
\]

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 \( \pi \).

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

In turn, this means that \( v(\theta|\pi, \tau) = u((1 - \tau) w(\theta|\pi) + \tau \pi w) \) is increasing in \( \pi \). Thus, holding investment decisions and \( \tau \) constant, welfare increases with \( \pi \).
Next, let $\pi^*(\tau)$ be the investment rate in the planner’s preferred equilibrium for each tax rate. The proof that $\pi^*(\tau)$ rises as $\tau$ falls is simple. First, if $\pi^*(\tau) = 0$, it cannot fall. Alternatively, suppose that $\pi^*(\tau_0) > 0$. Since lowering $\tau$ from $\tau_0$ to $\tau_1$ raises $G(\beta(\tau|\tau))$ for any $\pi$, it must be true that $G(\beta(\pi^*(\tau) | \tau_1)) > \pi^*(\tau_0)$. Since $G(\beta(\pi|\tau))$ is continuous and $G(\beta(1|\tau)) = 0$, there must be some higher investment rate $\hat{\tau}$ such that $G(\beta(\hat{\tau}|\tau_1)) = \hat{\tau} > \pi^*(\tau_0)$. 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 $\tau$. First, there is a mechanical effect. For a worker with signal $\theta$, this is as follows.

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

$$= u' \pi (1-\pi) \omega \left[ \frac{f_q(\theta) - f_u(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)} \right]$$

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

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

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

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

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 $\pi^*(\tau)$ again:

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

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

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

$$\frac{\tau}{1-\tau} = \frac{(1-\pi) \int_0^1 u' \left[ f_u(\theta) - f_q(\theta) \right] d\theta - \varepsilon \int_0^1 u' \left[ \frac{f_u(\theta) f_q(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)} \right] d\theta}{\varepsilon \int_0^1 u' \left[ \pi f_q(\theta) + (1-\pi) f_u(\theta) \right] d\theta}$$

$$= \frac{\overline{\tau} - \varepsilon \overline{w}}{\varepsilon \overline{z}}$$

76
Proof of Proposition 9. Fixing a value of $T_{A \rightarrow D}$, the proof that condition 42 must hold at the optimum is analogous to the proof of Proposition 8. A similar perturbation argument can be used to establish that condition 43 must hold. An increase in $T_{A \rightarrow D}$ leads to the following gain in welfare for type $A$ and $D$ individuals:

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

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 43.

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 $n_\theta$ 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 $B_0$, 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 $\bar{z}$ and $\bar{z}$ by $\Delta T'$. This yields a new tax schedule, $T_1$.

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

Take the existing wage given each $\theta$ but apply $T_1$ instead of $T_0$. Re-optimize labor supply decisions and calculate $v(w(\theta|\pi_0)|T_1)$ for each $\theta$, yielding a candidate vector of utilities $v_1^{(0)}$. Using $v_1^{(0)}$, calculate $E_\theta(v(\theta|\pi_0, T_1)|q)$ and adjust workers’ investment decisions toward their preferred choice. This yields a new distribution of productivity, $\delta_1^{(0)}(q|\pi_0, T_1)$.

In the discretized space, $\delta_1^{(0)}(q|\pi_0, T_1)$ implies a new candidate vector of productivity choices $q_1^{(1)}$. Use these choices to reconstruct a new candidate $B_1^{(1)}$ matrix. Then solve for employers’ rational productivity inferences at each value of $\theta$, yielding a candidate set of employer beliefs $\pi_1^{(1)}(q)$ and a hypothesized vector of wages $w_1^{(1)}$.

$$w_1^{(1)} = \left[ \text{diag} \left( B_1^{(1)} \times \delta_1^{(1)}(q|\pi_1^{(1)}, T_1) \right) \right]^{-1} \times \left[ B_1^{(1)} \times \text{diag} \left( q_1^{(1)} \right) \times \delta_1^{(1)}(q|\pi_1^{(1)}, T_1) \right]$$
Recalculate utilities to obtain \( v_1^{(1)} \) and adjust workers’ investment decisions again, yielding \( q^{(2)} \). Iterate this process until individuals do not want to adjust their investment decisions given the hypothesized employer beliefs: i.e., when \( \pi_1^{(k)}(q) \approx \delta_1^{(k)}(q|\pi_1^{(k)}(q), T_1) \). 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, hold all decisions (wages, labor supply and investment) constant and evaluate the mechanical change in welfare. 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 \( \theta \). 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, \( \Delta T \). 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:

\[
\frac{1}{\frac{1}{2} \exp(\alpha (z-z_0)) + \lambda + \frac{1}{2} \exp(-\alpha (z-z_0))}
\]

To choose the parameters, I use MATLAB’s \texttt{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 \( f_w^{\text{sim}} \).

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

As Figure 7 shows, this process is effective.

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

\[
E_\theta \left( u(w(\theta|\pi_0,T_0)|q) \right) = B \times v_0
\]

where \( v_0 \) is a vector that stacks the utility realized at each value of \( \theta \) and \( \pi_0 \) denotes employers’ current and correct beliefs about the distribution of productivity. Combined with individuals’ productivity choices and a value of \( \beta \), this vector of expected utilities can then be used to back out an implied cost distribution.

**E. Additional Figures**

Table H1 provides summary statistics for data used to test for employer learning in Section 5. Figures H1 to H3 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. Figure H4 shows the expected net transfer from the government for workers of each initial productivity level. Finally, Figure H5 plots the change in marginal social welfare weights starting from naïve taxation and transitioning to optimal taxation.

**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 H6 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 \( \varepsilon_l \) and \( \beta \) to ensure that the elasticity of taxable income is lower, at \( \varepsilon_{LR} = 0.8 \), but that the ratio \( \varepsilon_{LR}^{LZ}/\varepsilon_{LR}^{LR} \) 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 \( \varepsilon_{LR}^{LZ} \) between the two scenarios.
Table H1: Summary Statistics for High and Low AFQT Workers

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

* 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.

Second, panel (b) of Figure H6 shows the adjustment to marginal tax rates between the naïve and optimal tax schedules in four different 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 \( \varepsilon_{z}^{LR} \) 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 when the belief externality is taken into account by the planner.
**Figure H1: Comparison of Fiscal Externality**

![Fiscal externality graph]

*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 H2: Comparison of Belief Externality**

![Belief externality graph]

*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 H3: Comparison of Mechanical Effect**

![Figure H3: Comparison of Mechanical Effect](image)

*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 H4: Expected Net Transfer**

![Figure H4: Expected Net Transfer](image)

*Figure notes.* This figure plots the expected net transfer from the government for workers of each initial productivity level for the simulation described in Section 5. The solid red line shows the transfer under the optimal tax schedule, while the dashed blue line shows the transfer under the naïve tax schedule.
Figure H5: 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 H6: Non-linear Taxation with Alternative Parameter Values

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

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 $\varepsilon_L^{LR}/\varepsilon_z^{LR} = 0.6$ 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 $\varepsilon_L^{LR} = 0.8$; (iii) a third case with $\varepsilon_L^{LR}/\varepsilon_z^{LR} = 0.5$ but $\varepsilon_z^{LR}$ unchanged; and (iv) a fourth with $dE(w)/dq = 0.9$, which implies a smaller belief externality.