

Online Appendix

Optimal Income Taxation with Spillovers from Employer Learning

Ashley C. Craig

Contents

A	Generalizations of the Model	2
B	Continuity and Stability	11
C	Beyond the First Order Approach	16
D	Multiple Groups and Self-fulfilling Disparities	21
E	Approximately Optimal Taxation	24
F	Unproductive Signaling	26
G	Proofs and Derivations	32
H	Empirical Evidence	51
I	Simulation of the Model	57

A Generalizations of the Model

(For Online Publication)

A. GENERALIZING CONTRACTS

The model in Section 1 assumes that employers offer a wage to workers, as opposed to offering a general contract that specifies both a wage and labor supply. Here, I show that this is not restrictive. The reason is that workers with different types or productivity levels do not differ in their disutility of labor supply, fixing their wage.

To demonstrate this formally, I adopt all the assumptions of the baseline model except that I allow each employer to offer a contract $C_j = \{z_j, l_j\} \in \mathbb{R}_+ \times \mathbb{R}_+ = \mathbb{C}$ to the worker.¹ 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: $x : \mathcal{K} \times \mathcal{T} \rightarrow \mathbb{R}_+$; and $A : \mathcal{K} \times \mathcal{T} \times \Theta \times \mathbb{C}^{|\mathcal{J}|} \rightarrow \mathcal{J}$. Each employer's strategy is a function that maps signals and tax systems to *contract* offers $O_j : \Theta \times \mathcal{T} \rightarrow \mathbb{C}$. Despite the increased complexity, it remains true that every firm earns zero expected profit. Moreover, contracts can always be equivalently characterized as an offer of a wage $w_j = w(\theta|\pi)$ equal to the worker's expected marginal product given the signal θ , with the worker freely choosing how much labor to supply. In this sense, nothing substantive is changed from the baseline model.

Lemma 3. Fix a realized value of θ and assume that $E[q|\theta, \pi]$ is strictly positive and finite given equilibrium beliefs $\pi(q)$. In any pure-strategy equilibrium: all firms $j \in \mathcal{J}$ 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 $l_j \in \mathcal{L}_j^* = \operatorname{argmax}_{\tilde{l}_j \in \mathbb{R}_+} u(w_j \tilde{l}_j - T(w_j \tilde{l}_j), \tilde{l}_j)$.

B. WORKER SCREENING

If workers of different types do differ in their disutility of labor supply conditional on their hourly wage, screening by employers using menus of contracts may be possible. To see why, suppose workers' utility functions take the following quasilinear form:

$$U_k(c, l, x) = c - h_q(l) - kx \tag{28}$$

where $q = Q(x)$.

¹Employers could also offer menus of contracts, but this has no benefit because workers of different productivity levels have no reason to select different contracts.

The assumptions required for screening are documented by Spence (1978), and studied in the context of taxation by Stantcheva (2014). Here, it suffices to study a special case in which taxation is linear and workers have two different equilibrium productivity levels, $q_i = \{q_1, q_2\}$ with $q_1 < q_2$. However, the analysis can be generalized to non-linear taxation and many productivity levels (see Stantcheva 2014). Given a signal, θ , that the employer has observed, let the likelihood that an individual has productivity q_1 be λ_1 , and the likelihood that they have productivity q_2 be $\lambda_2 = 1 - \lambda_1$. Given these productivity levels, screening requires the following assumptions.

Assumption 4. (i) *Labor supply costs are increasing and convex: $h'_{q_i}(l) > 0, h''_{q_i}(l) > 0 \forall i$.*

(ii) *Total and marginal disutility of effort are zero with zero labor: $h_{q_i}(l) = h'_{q_i}(l) = 0 \forall i$.*

(iii) *The higher type experiences lower total disutility: $h_{q_2}(l) < h_{q_1}(l) \forall l > 0$.*

(iv) *The higher type experiences lower marginal disutility: $h'_{q_2}(l) < h'_{q_1}(l) \forall l > 0$.*

The reason screening is possible with these assumptions is that higher-productivity workers are willing to work longer hours given the *same wage*. It is plausible that these assumptions hold in some contexts, although it is unclear whether they hold in general.²

There are many models of screening. I focus on the Miyazaki-Wilson-Spence (MWS) equilibrium concept (Miyazaki 1977, Wilson 1977, Spence 1978). However, there are other screening models that would be reasonable, and the choice of which one is appropriate is sensitive to assumptions about the timing and nature of contract choices. Specifically, one of the types of MWS equilibria below involves cross-subsidization, and firms would always want to withdraw the loss-making contract if there were a second stage. Nonetheless, MWS is commonly used and has been justified in several different ways (see Fernandez and Rasmusen 1993, Netzer and Scheuer 2014, Mimra and Wambach 2019).

In the MWS framework, firms can offer an arbitrary menu of contracts, each specifying a salary $z_j \in \mathbb{R}_+$ and a quantity of labor $l_j \in \mathbb{R}_+$. Firms break even on their overall menu of contracts, and choose their menus recognizing that other firms may withdraw any contracts that are unprofitable.

Definition 2. *An equilibrium is a set of contracts such that firms break even across their entire menu of contracts, and there is no omitted contract that would be profitable after all contracts made unprofitable by its introduction have been withdrawn.*

²Indirect support for these assumptions comes from instances in which firms use this type of screening. For example, Landers, Rebitzer and Taylor (1996) study law firms that seem to screen associates by requiring them to work long hours before being promoted.

Formally, the firm solves the following problem.

$$\max_{z_1, z_2, l_1, l_2} (1 - \tau) z_2 - h_{q_2}(l_2) \quad (29)$$

subject to:

$$(1 - \tau) z_1 - h_{q_1}(l_1) \geq (1 - \tau) z_2 - h_{q_1}(l_2) \quad (\text{IC}_{12})$$

$$(1 - \tau) z_2 - h_{q_2}(l_2) \geq (1 - \tau) z_1 - h_{q_2}(l_1) \quad (\text{IC}_{21})$$

$$\lambda_1 z_1 + \lambda_2 z_2 = \lambda_1 q_1 l_1 + \lambda_2 q_2 l_2 \quad (\text{ZP})$$

$$(1 - \tau) z_1 - h_{q_1}(l_1) \geq (1 - \tau) z_1^{RS} - h_{q_1}(l_1^{RS}) \quad (\text{RS})$$

The first two constraints guarantee incentive compatibility: they require that individuals of each productivity level prefer the contract designed for them. The third constraint is a zero profit condition, pooled across both types. The fourth requires that the lower-productivity worker gets at least as much utility as in the ‘‘Rothschild-Stiglitz’’ separating allocation (Rothschild and Stiglitz 1976); this precludes possible profitable deviations by firms that would otherwise arise (see Miyazaki 1977).

For simplicity, I assume labor supply is isoelastic, which ensures that there is an adverse selection problem for all τ if there is an adverse selection problem for $\tau = 0$. The solution is provided by Stantcheva (2014), which is adapted and restated here.

Proposition 5. *For any tax rate τ , the profit constraint (ZP) is binding and the second IC constraint (IC_{21}) is slack. With isoelastic labor supply, the first IC constraint (IC_{12}) binds. The low type works an efficient amount of hours, $h_{q_1}^*(\tau)$. There are two possible configurations.*

- (i) *If the share of low-productivity types is high, $\lambda_1 > \tilde{\lambda}_1(t)$: the RS constraint binds, each worker earns her marginal product, and there is full separation. The higher-productivity type works more than the efficient level, with her labor supply characterized by:*

$$q_1 l_{q_1}^*(\tau) (1 - \tau) = q_2 l_{q_2}(\tau) (1 - \tau) - (h_{q_1}(l_{q_2}(\tau)) - h_{q_1}(l_{q_1}^*(\tau))).$$

- (ii) *If the share of low-productivity types is low, $\lambda_1 \leq \tilde{\lambda}_1(t)$: the RS constraint does not bind, and there is cross-subsidization from high to low productivity workers. The high-productivity type works more than is efficient, with her labor supply characterized by:*

$$h'_{q_2}(l_{q_2}(\tau)) = (1 - \tau) \lambda_2 q_2 + \lambda_1 h'_{q_1}(l_{q_2}(\tau)).$$

In both of these cases, the utilities of workers with different productivity levels are

compressed relative to an economy with symmetric information. In case (i), the hourly compensation of each worker is equal to her marginal product. However, the higher type's utility is lowered because her labor supply is distorted away from her optimum to ensure that low productivity workers do not want to pretend to have high productivity. In case (ii), this remains the case, but high-productivity workers cross-subsidize the wage of low-productivity workers as well.

From the point of view of the human capital investments that are the focus of this paper, this compression of workers' utility levels acts in the same way as the wage compression that occurs due to Bayesian belief formation when screening by firms is not possible. The utility gain from increasing one's productivity is lower than it would be if employers could directly observe productivity, which undermines a worker's incentive to acquire human capital, and implies a spillover when workers invest.

C. REPEATED INTERACTIONS IN THE LABOR MARKET

An important simplification embedded in the model in Section 1 is that workers invest in human capital once, and then interact with employers once in the labor market. In reality, human capital investment and labor supply decisions are made repeatedly. As discussed in Section 4, evidence suggests that employers gradually learn a worker's productivity, and that the return to a worker's initial level of skill therefore increases over time. In this appendix, I briefly discuss the implications of this.

Starting with the thought experiment in Section 4, let q be the worker's human capital and assume human capital investment occurs before entry into the labor market. (Generalizations follow below.) However, her marginal product, $MP_t(q)$ may depend on experience, t . In each period, let employers see a different signal $\theta_t \in \Theta_t$, which may become arbitrarily precise over time. The worker is still paid her expected marginal product, $E(MP_t(q)|\theta_t, \pi)$, and her discount factor is δ .

If utility is quasilinear, taxation is linear and labor supply is perfectly inelastic, the fraction of the social return to increasing q that the worker captures is:

$$s = \frac{\sum_{t=0}^T \delta^t \int_{\Theta_t} E(MP_t(q)|\theta_t, \pi) \frac{\partial f_t(\theta_t|q)}{\partial q} d\theta_t}{\sum_{t=0}^T \delta^t \frac{\partial MP_t(q)}{\partial q}} \quad (30)$$

where $f(\theta_t|q)$ is the conditional distribution of the period t signal.

The expression on the right of equation 30 is what I use in the calibration in Section 4. It is the present-discounted private return to higher productivity divided by the present-discounted social return. The expression takes into account the fact that employers gradually learn a worker's productivity, so that $E(MP_t(q)|\theta_t, \pi)$ approaches $MP_t(q)$ and the

worker is fully remunerated for her productivity later in life. Nonetheless, at earlier stages in her career – which are more important due to discounting – the marginal private return is lower than the social return. In addition, all my results take into account the fact that workers’ tax rates, labor supplies and welfare weights change with their wage.

If human capital investments were spread out over the lifecycle, the model would become more complex. The learning externality would then be present in every period for the marginal investment, raising the question of how taxes should vary over the lifecycle. On one hand, the fact that investment is still likely be concentrated at earlier ages suggests that the externality would be more important – and taxes lower – at younger ages. But on the other hand, the externality is exacerbated for later investments, all else equal. The reason for this is that a worker’s marginal investment decision can be predicted based on behavior in *previous* periods, which reduces the weight placed by employers on any signals of contemporaneous investment. This mechanism is likely to push toward lower taxes at older ages, to the extent that human capital investment remains important.

The model could be extended further still, to accommodate possibilities that arise in a dynamic context. For example, with repeated interaction, employers might find a way to provide incentives with dynamic contracts. There are many versions of this, which generally involve workers effectively being remunerated in later periods for having had higher productivity in earlier periods than was expected at the time (and vice versa). Such dynamic incentives would need to be supported by commitment power or reputation-building, and many such models involve workers agreeing to work for less than their marginal product in earlier periods (possibly implying negative initial wages). This is plausible in some settings, which makes this a rich but complex direction for future work.

D. UNEQUAL OR ASYMMETRIC LEARNING

The model discussed throughout this paper features symmetric information across employers. An alternative assumption would be that incumbent employers have more information about a worker’s productivity than do other firms. Acemoglu and Pischke (1998) is one example. They model learning more more simply than in this paper, focusing on the asymmetry. Specifically, a worker’s first-period employer knows her productivity in the second period, but outside employers do not.

The key result of Acemoglu and Pischke (1998) is that the incumbent employer has ex-post monopsony power because of their informational advantage, and earns a profit. Importantly, that profit is increasing in a worker’s skill. In turn, this produces the motivation for a firm to train its workers that is the focus of their paper. However, the flip-side is that a worker’s wage increases by less than her skill despite the fact that her productivity

is known with certainty by the incumbent firm in the second period. This means, in turn, that there would be less incentive for the worker to invest in her own skill.

If the model here incorporated asymmetric learning, it would feature a similar effect. Learning itself would mean that the relationship between expected productivity and skill would be flatter than the relationship between true productivity and skill. That is why there is under-investment in the present paper. Asymmetric learning would lead wages to be further compressed even relative to expectations, causing a further distortion.

E. STRUCTURAL CHANGES IN THE LABOR MARKET

The framework in this paper has implications for empirical tests designed to detect discrimination and its effects. Here, I consider the results of Hsieh, Hurst, Jones and Klenow (2019) who show that women and black men have chosen increasingly high-skilled occupations over the past 60 years, converging to a set of choices that is more similar to white men. Building on this observation, they ask how much of GDP growth over this period can be explained by falling barriers such as discrimination which had previously prevented women and black men from choosing occupations that reflected their comparative advantage. They draw a distinction between two types of barriers: those in labor markets and those in education markets.

Hsieh et al. (2019) link these two types of barriers directly to a “taste-based” discrimination framework in the spirit of Becker (1957). That framework provides a micro-foundation for two types of implicit “taxes” on workers from each disadvantaged group, relative to white men. First, due to prejudice against hiring them, members of the disadvantaged workers are paid less per efficiency unit than white men: Specifically, the ratio of the minority to white male wage is $1 - \tau_{ig}^w$. Second, there is a prejudice against providing educational services to disadvantaged workers, so that their cost of attaining human capital is $1 + \tau_{ig}^h$ higher than white men. These wedges take much the same form as a tax (τ_{ig}^w) on minority income and tax on their educational inputs (τ_{ig}^h). The “taxes” distort educational investments, and reduce utility levels. In turn, this deters members of the disadvantaged groups from entering the labor force or choosing high-skill occupations.

My model also features distortions to human capital investments due to implicit “taxes”. When I introduce exogenous worker characteristics in Section 5 Part B, those distortions may well vary between different groups. As I show, the model then produces statistical discrimination against disadvantaged groups. Specifically, there are two groups: advantaged workers (A) and disadvantaged workers (D). The only difference between the groups is that disadvantaged workers have proportionally higher investment costs. To put this in the framework of Hsieh et al. (2019), this cost disadvantage is indistinguish-

able from saying that disadvantaged workers have the same distribution of investment costs as majority workers but face a “tax” of $1 + \tau_{ig}^h$ on those investments.

There is also a “tax” in the labor market. The expected wage of a disadvantaged worker is lower than that of an advantaged worker with the same productivity level. The fraction of the overall wage gap between the two groups explained by that wedge is $1 - s$.

$$\ln \left(\frac{E(w | q, A)}{E(w | q, D)} \right) = (1 - s) \left(\frac{E(w | A)}{E(w | D)} \right) \quad (31)$$

Following Hsieh et al. (2019), we can let $\tau_{ig}^w = 1 - s$ be the tax wedge from labor market discrimination. In many respects, this wedge is again very similar to the wedge in the taste-based discrimination framework above.

At this point, we are in a position to examine the distortion to human capital investments from the two “taxes”. And it is here that we find a stark deviation from Hsieh et al. (2019). Unlike their taste-based discrimination framework, the entire gap in productivity levels is explained by the human capital rather than labor market discrimination.

$$\frac{q_B}{q_{AA}} = \left(\frac{1}{1 + \tau_B^h} \right)^{\frac{\beta}{1-\beta}} \quad (32)$$

The reason for this is that statistical discrimination here is “rational” in the sense that it reflects a true gap in human capital. In the labor market, one would observe that a fraction s of the overall log wage gap between the groups is explained by an individual’s own gap in the signal they send to employers (which reflects their own true productivity). A fraction $1 - s$ is explained by labor market discrimination, but that discrimination also reflects a gap in productivity between the two groups.

At this point, we can turn our attention toward the empirical identification strategy of Hsieh et al. (2019). The fundamental logic of their analysis is that: (i) human capital disadvantages are reflected in a cohort’s outcomes in *all time periods*, because investments are locked in at an early stage; and (ii) labor market barriers are reflected in outcomes in a given time period for *all cohorts* at a point in time because (for example) employers’ prejudice-based discrimination does not distinguish between cohorts.

Here, there is another sharp divergence between the taste-based discrimination model Hsieh et al. (2019) use to interpret their results, and the statistical discrimination model. Rational statistical discrimination implies that belief formation should occur within a demographic cell, so that discrimination is proportional to the gap in human capital between white male and disadvantaged workers in that cohort. In other words, labor market ac-

tors who statistically discriminate rationally distinguish between cohorts, and in a way that is inseparable from direct human capital barriers. Regardless of the ultimate source of the gap in outcomes between advantaged and disadvantaged workers in the statistical discrimination model (e.g., cost disadvantages, differences in screening technologies, or self-fulfilling disparities), the conclusions from Hsieh et al.'s (2019) analysis would therefore always be that distortions arise due to a barrier to human capital formation.

The final point to note is that there is a distortion to human capital formation in my model that affects both the advantaged and disadvantaged group: both groups underinvest in human capital because of the belief externality. However, Hsieh et al. (2019) normalize the distortion to zero for white men. This observation is important because it implies that there may be misallocation for all groups even if there is no *differential* misallocation across groups.

F. LINEAR TAXATION WITHOUT PARAMETRIC ASSUMPTIONS

Without the parametric assumptions in Section 2, the analysis of linear taxation closely follows the analysis of non-linear taxation in Section 3. Moreover, the same logic applies at an equilibrium other than the log-normal one considered in Propositions 1 and 2. In fact, the analysis is identical except that the tax system is restricted to be linear.

As in Section 3 then, let $w(\theta | \pi)$ be the wage of a worker with signal realization θ . Similarly, those workers have average welfare weight $\psi(\theta)$, labor supply $l(\theta | \pi, T)$ and income $z(\theta | \pi, T)$. The density of workers with type θ is $f(\theta)$. Finally, we need to define $\varepsilon_{\bar{q}}(\theta)$ as the elasticity of the productivity of workers with signal θ with respect to $1 - \tau$.

Putting all of that together, we obtain the same optimal tax formula except that the belief externality is generalized as follows:

$$\frac{\tau^*}{1 - \tau^*} = \frac{1 - \alpha}{\varepsilon_z} - \underbrace{\int_{\Theta} \psi(\theta) \left[\frac{z(\theta | \pi, T)}{\bar{z}} \right] \left[\frac{\varepsilon_w^{\text{BE}}(\theta) \varepsilon_{\bar{q}}(\theta)}{\varepsilon_{\bar{q}}(\theta) \varepsilon_z} \right] f(\theta) d\theta}_{\text{Generalized belief externality}} \quad (33)$$

where:

$$\varepsilon_w^{\text{BE}}(\theta) = \frac{dw(\theta | \pi)}{d(1 - \tau)} \frac{1 - \tau}{w(\theta | \pi)} \quad (34)$$

is the change in the the wage paid given signal realization θ .

The log-normal equilibrium in Section 2 is a special case. First, the response of productivity is uniform, so that $\varepsilon_{\bar{q}}(\theta) = \varepsilon_q$. Second, the fraction of the return to investment that workers capture is a constant: specifically, $\varepsilon_w^{\text{EXT}}(\theta) / \varepsilon_{\bar{q}}(\theta) = 1 - s$ for all θ . If we impose both of these restrictions here, we are left with the same formula as in Proposition 2.

This highlights the key difference in general. Compared to the log-normal equilibrium,

the belief externality will be larger (and the optimal tax lower all else equal) if there is a greater response among workers for whom the externality is larger, or if the externality of those who respond disproportionately affects workers with higher incomes and higher welfare weights. This closely mirrors the discussion of incidence in Sections 3 and 4 of the paper with non-linear taxation, with the main difference being that non-linear taxation is not available to respond to such differences across the distribution.

B Continuity and Stability

(For Online Publication)

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. Assume that there is a finite number of cost types, indexed by $i = 1, \dots, |\mathcal{K}|$, let \mathbf{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.

$$\lambda_i(\mathbf{x}, T) = Q'(x_i) \int_{\Theta} v(\theta|\pi, T) \frac{\partial f(\theta|q)}{\partial q} \Big|_{q=q_i} d\theta - k_i = 0 \quad (35)$$

Differentiating $\lambda_i(\mathbf{x}, T)$ 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:

$$\frac{\partial \lambda_i}{\partial x_j}(\mathbf{x}) = \begin{cases} \lambda_{ii}^q + \lambda_{ij}^w & \text{if } i = j \\ \lambda_{ij}^w & \text{if } i \neq j \end{cases} \quad (36)$$

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

$$\lambda_{ii}^q = Q''(x_i) \int_{\Theta} v(\theta|\pi, T) \frac{\partial f(\theta|q)}{\partial q} \Big|_{q=q_i} d\theta + Q'(x_i)^2 \int_{\Theta} v(\theta|\pi, T) \frac{\partial^2 f(\theta|q)}{\partial q^2} \Big|_{q=q_i} d\theta \quad (37)$$

$$\lambda_{ij}^w = Q'(x_j) \int_{\Theta} u_c(\theta) [1 - T'(z(\theta|\pi, T))] l(\theta|\pi, T) \frac{\partial w(\theta|\pi)}{\partial q_j} f(\theta|q_i) d\theta \quad (38)$$

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

$$\frac{\partial w(\theta|\pi)}{\partial q_j} = \left(f(\theta|q_j) + [q_j - w(\theta|\pi)] \frac{\partial f(\theta|q)}{\partial q} \Big|_{q=q_j} \right) p(k_j) \quad (39)$$

The partial derivatives (equation 36) can be arranged to form the Jacobian $\mathbf{J}_{f,\mathbf{x}}$.

$$\mathbf{J}_{f,\mathbf{x}} = \begin{bmatrix} \frac{\partial \lambda_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial \lambda_1}{\partial x_{|\mathcal{K}|}}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial \lambda_{|\mathcal{K}|}}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial \lambda_{|\mathcal{K}|}}{\partial x_{|\mathcal{K}|}}(\mathbf{x}) \end{bmatrix} \quad (40)$$

Next, let $dc(\theta|\pi, T) = -dT(z(\theta|\pi, T))$ be the Fréchet derivative with respect to T of

consumption by a worker with signal θ . The Fréchet derivative of $v(\theta|\pi, T)$ is then:

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

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

$$d\lambda_i(\mathbf{x}, T) = Q'(x_i) \int_{\Theta} dv(\theta|\pi, T) \left. \frac{\partial f(\theta|q)}{\partial q} \right|_{q=q_i} d\theta \quad (41)$$

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

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

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

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

First, I construct a parameterized family of functions, $\tilde{Q}(x|\mathbf{c})$, where \mathbf{c} is a vector of strictly negative real numbers $c_1, \dots, c_{|\mathcal{K}|}$. Each function in this family retains the key properties of $Q(x)$, but the second derivative of $\tilde{Q}(x|\mathbf{c})$ 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 $\tilde{Q}(x|\mathbf{c})$ with $\tilde{Q}(0|\mathbf{c}) = 0$, $\tilde{Q}'(x|\mathbf{c}) > 0$, $\tilde{Q}''(x|\mathbf{c}) > 0$ and $\lim_{x \rightarrow 0} \tilde{Q}'(x|\mathbf{c}) = \infty$. This is always possible, since r is small and Q strictly concave.
3. Let $\hat{Q}(x|\mathbf{c}, \alpha) = \alpha\tilde{Q}(x|\mathbf{c}) + (1 - \alpha)Q(x)$ with $\alpha \in (0, 1)$.

Next, I replace $Q(x)$ with $\hat{Q}(x|\mathbf{c}, \alpha)$ in the economy described in Section 1. For any \mathbf{c} , there remains an equilibrium with the same investment decisions. However, the diagonal elements of the Jacobian $\mathbf{J}_{f,\mathbf{x}}$ are replaced by:

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

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

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

C. STABILITY OF EQUILIBRIA

Restricting the set of equilibria to those that are stable is one way to ensure that the economy does not switch equilibria in response to a perturbation such as that described in Section 4. To define such a notion of stability, suppose that the economy evolves according to the following backward-looking dynamic adjustment process:

$$x_{k,t+1} \in X_{k,t+1}^* = \operatorname{argmax}_{\tilde{x} \in \mathbb{R}_+} \int_{\Theta} v(\theta|\pi_t, T) f(\theta|Q(\tilde{x})) d\theta - k\tilde{x} \quad (42)$$

where:

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

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

Thus, letting $\bar{x}(T)$ be a set of equilibrium investment decisions, the dynamic adjustment process above can be approximated locally around $\bar{x}(T)$ by a first-order linear system $\mathbf{x}_{t+1}(T) - \bar{x}(T) = \mathcal{B}[\mathbf{x}_t(T) - \bar{x}(T)]$. If all the eigenvalues of the matrix \mathcal{B} have moduli strictly less than one, then the equilibrium is locally asymptotically stable. Providing that $\mathbf{J}_{f,x}$ is invertible (see Part A above) so that there is a locally unique Fréchet differentiable function mapping T to \mathbf{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. EXISTENCE OF PURE STRATEGY EQUILIBRIA

In this appendix, I discuss the existence of a set of pure strategy equilibria. I begin with a practical observation: fixing any signal technology $f(\theta|q)$, there is always a production

function $Q(x)$ that satisfies the assumptions in Section 1 while also guaranteeing the existence of an equilibrium with the empirically observed wage distribution, labor supply elasticity and wage elasticity. It can also be ensured that this equilibrium is stable. Indeed, I use these facts in my simulations in Section 4.

Despite the fact that there is always an equilibrium that matches any observed wage distribution, it is important to note that uniqueness is not guaranteed. In fact, there is always a “bad” equilibrium in which no worker invests at all and wages are zero for all signal realizations. This equilibrium is always unstable given the assumptions in Section 1. However, it is not generally possible to rule out the potential for multiple internal stable equilibria. I discuss this issue further in Appendix D.

It is harder to provide general weak conditions for the existence of a stable equilibrium with positive investment in the general model *without* backing out a set of fundamentals from one that is observed. To see why, let $q_i^*(\mathbf{q})$ be an expected utility maximizing productivity level of type i , given that employers believe that the vector of productivity levels is \mathbf{q} . The assumption that $\lim_{x \rightarrow 0} Q'(x) = \infty$ combines with the regularity assumptions on the signal technology to ensure that all individuals’ optimal productivity levels are greater than employers believe them to be: i.e., $q_i^*(\mathbf{q}) > q_i$ for all $i = 1, \dots, |\mathcal{K}|$ if \mathbf{q} is close enough to zero. If we were also to assume that the marginal return to investment is zero above some \bar{q} (effectively truncating the relevant strategy space), then $q_i^*(\mathbf{q}) < \bar{q}$ if employers believe that all workers have productivity level \bar{q} . Thus, if $q_i^*(\mathbf{q})$ were globally continuous, we would be guaranteed another equilibrium. The problem is that this need not be the case without restrictions that guarantee concavity of investment returns.

Despite this difficulty in providing general conditions, it is also clear that this is an issue that need not be of great concern. Not only does the calibration in Section 4 back out a well-defined stable equilibrium from the data, but the equilibrium remains well-defined as the optimization algorithm explores many different parts of a multi-dimensional space. In addition, the examples in Section 2 and Appendix D have closed-form solutions. There are many other such examples.

E. 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 \mathcal{S} be the diagonal matrix of second-order conditions:

$$\mathcal{S} = \begin{bmatrix} \mathcal{S}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{S}_{|\mathcal{K}|} \end{bmatrix}$$

where:

$$\mathcal{S}_i = Q''(x_i) \int_{\Theta} v(\theta|\pi, T) \frac{\partial f(\theta|q)}{\partial q} \Big|_{q=q_i} d\theta + Q'(x_i)^2 \int_{\Theta} v(\theta|\pi, T) \frac{\partial^2 f(\theta|q)}{\partial q^2} \Big|_{q=q_i} d\theta$$

The response of investment can then be written as $d\mathbf{x} = -\mathbf{J}_{f,\mathbf{x}}^{-1} \mathcal{S} \mathcal{S}^{-1} d\lambda(\mathbf{x}, T)$. Letting \mathcal{I} be the identity matrix, the matrix $\mathbf{J}_{f,\mathbf{x}}^{-1} \mathcal{S}$ can be rewritten as the following Neumann series, providing that series is convergent.

$$\mathbf{J}_{f,\mathbf{x}}^{-1} \mathcal{S} = \sum_{k=0}^{\infty} (\mathcal{I} - \mathcal{S}^{-1} \mathcal{J})^k$$

The matrix $\mathcal{B} = \mathcal{I} - \mathcal{S}^{-1} \mathbf{J}_{f,\mathbf{x}}$ 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 Neumann series above corresponds to the case of stability discussed in Part D. 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.

$$d\mathbf{x} = \underbrace{-\mathcal{S}^{-1} d\lambda(\mathbf{x}, T)}_{\text{Direct impact}} - \underbrace{\sum_{k=1}^{\infty} \mathcal{B}^k \mathcal{S}^{-1} d\lambda(\mathbf{x}, T)}_{\text{Indirect wage response}} \quad (43)$$

The intuition here is similar to Proposition 1 of Sachs, Tsyvinski and Werquin (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 43 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

(For Online Publication)

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.³ This greatly simplifies the analysis of this subset of issues, while providing insights that are conceptually general.

A. SPECIAL CASE OF THE MODEL WITH BINARY INVESTMENT

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

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

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

The worker accepts her best offer, supplies a unit of labor and receives that wage. If she invested, she obtains utility $v(\theta|\pi, \tau) - k = u((1 - \tau)w(\theta|\pi) + \tau\bar{w}) - k$, where τ 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: $\lim_{c \rightarrow 0} u'(c) = \infty$ and $\lim_{c \rightarrow \infty} u'(c) = 0$.

³The model with binary investment is similar to Moro and Norman (2004).

Integrating over θ , the expected utilities (gross of investment costs) for an investor (\bar{v}_q) and non-investor (\bar{v}_u) are given by equations 44 and 45.

$$\bar{v}_q(\pi|\tau) = \int_0^1 v(\theta|\pi, \tau) dF_q(\theta) - k \quad (44)$$

$$\bar{v}_u(\pi|\tau) = \int_0^1 v(\theta|\pi, \tau) dF_u(\theta) \quad (45)$$

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

The final requirement of equilibrium is that workers invest at a rate that coincides with employers' beliefs. This is embodied in equation 46, which states that the fraction of investors must be equal to the fraction of workers that employers believe are qualified.

$$\pi = G(\beta(\pi|\tau)) \quad (46)$$

For a given tax rate τ , equation 46 defines a fixed point as shown in Figure C1. An employer belief π , combined with the tax τ , pins down the investment return and an investment rate, $G(\beta(\pi|\tau))$.

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.⁴ 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 π such that $G(\beta(\pi|\tau)) > \pi$ then there are multiple solutions to condition 46.*

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

B. OPTIMAL TAXATION WITH BINARY INVESTMENT

Tax policy can be analyzed in the same way as in the general model. Raising the linear tax τ causes $G(\beta(\pi|\tau))$ to shift down for every employer belief π . As a result, the location

⁴Tweaking the assumptions so that $G(0) > 0$ eliminates the equilibrium with zero investment.

FIGURE C1: EQUILIBRIA AND TAXATION

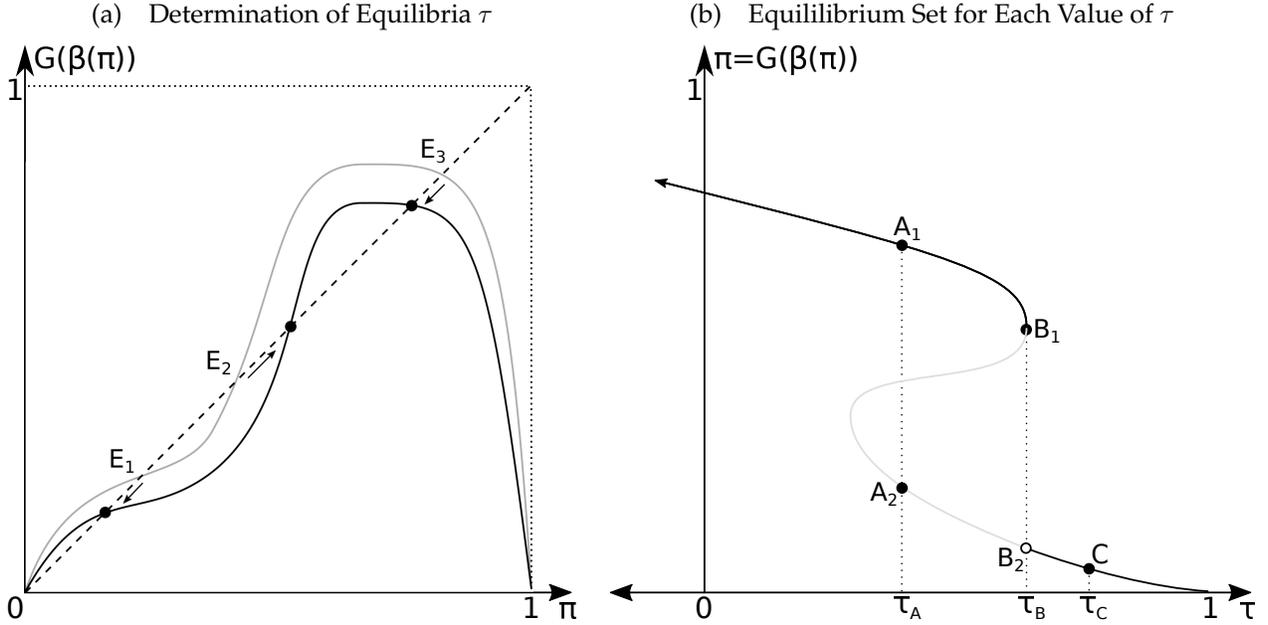


Figure notes. This figure shows an example economy with binary investment. In panel (a), the aggregate rate of investment implied by worker and firm optimization, $G(\beta(\pi))$, is plotted against the employer prior, π . Any intersection between this line and the 45 degree line is an equilibrium. The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

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 τ is lowered.⁵ 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 π satisfy equation 46 for a given tax rate τ . Let π_i and π_j be two solutions. Welfare is higher for every worker under π_i than π_j if and only if $\pi_i > \pi_j$. Moreover, investment in the Pareto-dominant equilibrium increases as τ is lowered.*

Next, to characterize optimal taxation, define ε_z as the elasticity of average income with respect to the retention rate. Second, let u'_θ be the marginal utility of consumption of

⁵The 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 τ rises. In Figure C1, both the zero investment equilibrium and E_2 are unstable.

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

Proposition 8 provides a necessary condition for the optimality of τ , in the same form as Propositions 2 and 3. As before, there is a trade-off between redistribution from high-wage to low-wage workers, a fiscal externality and a belief externality. Ignoring the belief externality, an optimal τ at which this condition holds would always be strictly positive. The belief externality \bar{w}_z provides an efficiency motive for intervention and pushes toward lower tax rates.

Proposition 8. *Fix a value of τ and an investment rate $\pi^*(\tau) > 0$, which satisfies equation 46. If $g(\beta(\pi^*(\tau)|\tau))\beta'(\pi^*(\tau)|\tau) \neq 1$ and τ is optimal, then the following condition holds:*

$$\frac{\tau}{1-\tau} = \frac{\bar{v}_\tau - \varepsilon_z \bar{w}_z}{\varepsilon_z} \quad (47)$$

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

Proposition 8 parallels the results from the linear tax example (Proposition 2) and non-linear taxation (Proposition 3). The requirement that $g(\beta(\pi^*(\tau)|\tau))\beta'(\pi^*(\tau)|\tau) \neq 1$ simply suffices to ensure the investment rate varies continuously with τ at the optimum, which is equivalent to invertibility of the Jacobian, $\mathbf{J}_{f,x}$, 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 τ_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 is 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 τ as we approach τ_B from below and *also* as we approach τ_B from above, so that τ_B is the optimal tax rate. However, equation 47 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

(For Online Publication)

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

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

Definition 3. A tax system T is a triple $(\tau_A, \tau_D, T_{A \rightarrow D})$, comprised of a marginal tax rate τ_j for each group combined with an intergroup transfer $T_{A \rightarrow D}$.

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

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

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

Gross of investment costs, a worker's expected utility is thus $\bar{v}_q^j(\pi_j|T)$ if she invests, and $\bar{v}_u^j(\pi_j|T)$ if she does not.

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

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 3, which applies given any continuous selection of equilibria. Specifically, for any given tax schedule T , let $\pi(T)$ be the set of pairs (π_A, π_D) such that $\pi_j(T) = G(\beta_j(\pi(T)|T))$ for $j \in \{A, D\}$. The correspondence

$\pi(T)$ suffices to characterize the set of equilibria for each tax schedule. I define a selection by choosing one equilibrium pair $\pi^\dagger(T)$ for each tax schedule from this set.

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

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

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

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

$$\frac{\tau_j}{1 - \tau_j} = \frac{\bar{v}_{j,\tau} - \varepsilon_z^j \bar{w}_z^j}{\varepsilon_z^j} \quad (48)$$

$$\int_\theta u'_{A,\theta} dF(\theta) = \int_\theta u'_{B,\theta} dF(\theta) \quad (49)$$

where $\bar{v}_{j,\tau} = (1 - \pi_j) \int_0^1 \tilde{u}'_{j,\theta} [f_u(\theta) - f_q(\theta)] d\theta$, ε_z^j is the income elasticity of group j , and $\bar{w}_z^j = \frac{1}{\omega} \int_0^1 \tilde{u}'_{j,\theta} \frac{\partial w(\theta|\pi_j)}{\partial \pi_j} [\pi_j f_q(\theta) + (1 - \pi_j) f_u(\theta)] d(\theta)$ 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 τ_j is then isomorphic to the model with a single group, so $\tau_A = \tau_B$ and $\pi_A = \pi_B$. Moreover, if condition 48 holds, equation 49 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 *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

(For Online Publication)

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 belief externality falls on workers with similar welfare weight, labor supply and tax rate to those with income z .

Part E of Section 3 shows that the welfare impact of a wage change due to the belief externality is weighted by $\Omega(z, \tilde{\theta}) = \psi_z(z(\tilde{\theta}|\pi, T)) [1 - T'(z(\tilde{\theta}|\pi, T))] l(\tilde{\theta}|\pi)$. Using this, and letting $l(z)$ be the labor supply at income z , I define $\tilde{\Omega}(z, \tilde{\theta})$ as the difference between the weight on externalities at income $z(\tilde{\theta}|\pi, T)$ and the weight at income z .

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

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

$$\begin{aligned} \mathbf{BE}(z) = & -d\tau dz \left\{ \psi_z(z) [1 - T'(z)] l(z) \left[\int_{\Theta} \left(\frac{dw(\tilde{\theta}|\pi)}{d[1 - T'(z)]} \right) f(\tilde{\theta}) d\theta \right] \right. \\ & \left. + \underbrace{\int_{\Theta} \tilde{\Omega}(z, \tilde{\theta}) \left(\frac{dw(\tilde{\theta}|\pi)}{d[1 - T'(z)]} \right) f(\tilde{\theta}) d\theta}_{\text{Covariance bias}} \right\} \end{aligned} \quad (50)$$

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

$$\int_{\Theta} \frac{dw(\tilde{\theta}|\pi)}{d[1 - T'(z)]} f(\tilde{\theta}) d\tilde{\theta} = (1 - s(z)) \frac{d\bar{w}}{d[1 - T'(z)]} \quad (51)$$

Bringing everything together, expression 18 is approximately zero if:

$$\mathbf{FE}(z) + \mathbf{ME}(z) - (1 - s(z)) \psi_z(z) l(z) [1 - T'(z)] \frac{d\bar{w}}{d[1 - T'(z)]} = 0. \quad (52)$$

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 3, the correction term in equation 52 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 52 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

FIGURE E1: APPROXIMATELY OPTIMAL TAXATION

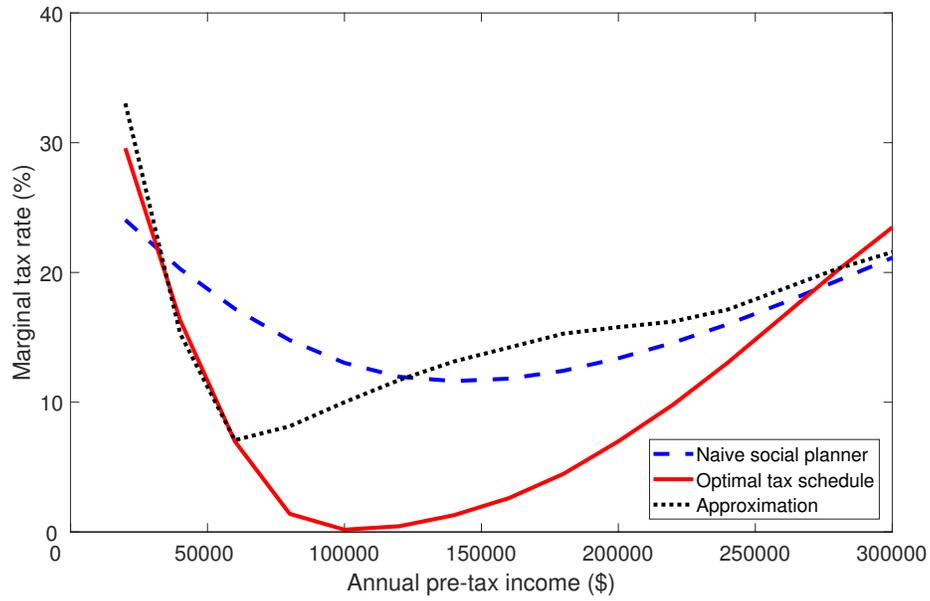


Figure notes. This figure shows the results of the simulation described in Appendix I. The solid red line shows the optimal tax schedule, the dashed blue line shows the naïve schedule, and the dotted black line shows a schedule what would be accepted by a planner who implemented equation 52.

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.

F Unproductive Signaling

(For Online Publication)

If the productivity of a worker depends directly on her type as well as her human capital investment, it is possible for investment to play an unproductive or ‘pure’ signaling role as in Spence (1973). Investment returns then reflect both a genuine increase in skill, and partial revelation of innate ability. In general, the overall externality from investment may then be more positive or more negative than in the model without innate ability.⁶

A. UNPRODUCTIVE SIGNALING: EXAMPLE WITH LINEAR TAXATION

I start with an extension of the example in Section 2, and then study the general case. Productivity, $q = n^\alpha h^{1-\alpha}$, is a Cobb-Douglas combination of human capital h and innate ability n . Human capital, $h = x^\beta$, is attained via investment, x . Inherent ability is negatively related to a worker’s investment cost: $n = 1/k$. Finally, the ability distribution and the conditional signal distribution are log-normal.

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

There is again an equilibrium in which income and productivity are log-normally distributed. The elasticities of productivity and income are functions of the labor supply elasticity, ε_l , the production function elasticity, β , and the importance of innate ability, α .

Proposition 10. *For any tax rate τ , there is an equilibrium in which productivity and income are log-normally distributed. Assuming this equilibrium is played, the elasticities of productivity and investment with respect $1 - \tau$ are as follows.*

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

This example nests the version in Section 2 in which investment is purely productive. When $\alpha = 0$ so that $q = h$, the two elasticities ε_q and ε_z collapse to that case, and equation 53 collapses to equation 11. When $\alpha = 1$ 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 11. It features a second externality correction, $1 + s\alpha(1 + \varepsilon_l)$, which pushes toward *higher* rather than

⁶The empirical importance of unproductive signaling is hard to assess. For formal education, evidence from school reforms demonstrate substantial productive effects (Meghir and Palme 2005, Aakvik, Salvanes and Vaage 2010, Oreopoulos 2006), but there is also evidence to suggest a role for pure signaling (Lang and Kropp 1986, Bedard 2001, Aryal, Bhuller and Lange 2020). See Lange and Topel (2006) for a discussion.

lower taxes. Intuitively, there is no social benefit from the part of the private return to investment that comes from signaling innate ability, which implies that this return comes at the expense of other workers. The logic here is similar to the rent transfer effect in Section 3. Holding fixed the decisions of others, a worker who invests more hurts other workers, because she becomes more likely to be pooled by employers with workers who have higher productivity than herself, thereby lowering the wages of those other workers.

Proposition 11. *Assume that the log-normal equilibrium described in Proposition 10 is played. Then the first-order condition for the optimal linear tax τ^* is:*

$$\frac{\tau^*}{1 - \tau^*} = \frac{1 - \gamma \left[\frac{1 + (1-s)\varepsilon_q}{1 + s\alpha(1 + \varepsilon_l)} \right]}{\varepsilon_z} \quad (53)$$

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

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

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

The condition on the left states that the social and private benefits of investment are aligned. The one on the right states that the unproductive component of the private return is equal in magnitude to the part of the productive component that is not captured by the individual. If these conditions hold, condition 53 collapses to the standard optimal tax formula. Any other parameter values imply a correction on efficiency grounds.

As these equations show, noisier employer information implies a smaller private benefit of investment for a given social benefit. Specifically, lower s dampens the signaling externality but strengthens the learning externality. 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 better information.

B. UNPRODUCTIVE SIGNALING: GENERAL CASE WITH OBSERVABLE INVESTMENT

I next move beyond the simple example, and extend the general model outlined in Section 1 to allow for unproductive signaling. However, I start with the simpler case in which investment is perfectly observable. This entails replacing the production function with $q = Q(x, k)$, so that productivity is a direct function of the worker's type. Employers observe productivity, x , but do not observe productivity, q .

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

$$\max_{x \in \mathbb{R}_+} v(w(x)|T) - kx \quad (54)$$

where:

$$v(w(x)|T) = \max_{l \in \mathbb{R}_+} u(w(x)l - T(w(x)l), l). \quad (55)$$

The solutions to problem 54 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. Given this, I provide conditions in Part C that guarantee $x(k)$ and $w(x)$ are differentiable, which ensures that the investment choice for a worker with cost k is characterized by a first-order condition:

$$u_c(z(k) - T(z(k)), l(k)) [1 - T'(z(k))] l(k) w'(x(k)) = k \quad (56)$$

where $l(k)$ is the level of labor supply that solves problem 55, and $z(k) = w(x(k))l(k)$ is the equilibrium income of a worker with cost k .

This relationship between innate ability and investment drives a wedge between the private and social returns, which I refer to as the unproductive component.

$$\underbrace{\frac{Q_k(x(k), k)}{x'(k)}}_{\text{Unproductive}} = \underbrace{w'(x(k))}_{\text{Private}} - \underbrace{Q_x(x(k), k)}_{\text{Productive (social)}} \quad (57)$$

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

These results provide a foundation for policy analysis that mirrors Section 3. 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.

$$\mathbf{BE}(z) = -d\tau dz \int_{\mathcal{K}} \psi(k) [1 - T'(z(k))] l(k) \frac{dx(k)}{d[1 - T'(z)]} [w'(x(k)) - Q_x(x(k), k)] dG(k)$$

This equation for $\mathbf{BE}(z)$ can again be written in terms of the observable income distribution, and combined with the fiscal externality and mechanical effect to obtain a necessary

condition for optimality of the tax system:

$$\mathbf{FE}(z) + \mathbf{ME}(z) + \int_{\mathcal{Z}} \tilde{z} \psi(\tilde{z}) \left(\frac{1 - T'(\tilde{z})}{1 - T'(z)} \right) \varepsilon_{\tilde{x}(\tilde{z}), 1 - T'(z)} \left[\varepsilon_{\tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})}^{\text{Private}} - \varepsilon_{\tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})}^{\text{Social}} \right] dH(\tilde{z}) = 0 \quad (58)$$

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.

$$\begin{aligned} \varepsilon_{\tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})}^{\text{Private}} &= w'(x(k)) \frac{x(k)}{w(k)} & \varepsilon_{\tilde{x}(\tilde{z}), 1 - T'(z)} &= \frac{dx(\tilde{z})}{d[1 - T'(z)]} \frac{1 - T'(z)}{x(\tilde{z})} \\ \varepsilon_{\tilde{w}(\tilde{z}), \tilde{x}(\tilde{z})}^{\text{Social}} &= Q_x(x(k), k) \frac{x(k)}{w(k)} \end{aligned}$$

Note the similarity between expression 18 and equation 58. This is not coincidental: 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. DIFFERENTIABILITY OF $w(x)$ AND $x(k)$

I next provide conditions under which $w(x)$ and $x(k)$ are differentiable in Part B above. As in Section 3, I assume that problem 55 is strictly concave given a wage $w = w(x)$ so that the labor supply choice can be characterized by a first-order condition (equation 59):

$$w u_c(w l^*(w) - T(w l^*(w)), l^*(w)) [1 - T'(w l^*(w))] + u_l(w l^*(w) - T(w l^*(w)), l^*(w)) = 0 \quad (59)$$

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

Next, I define $\hat{v}(x) = v(w(x) | T)$, and let $x_{FB}(k) = \operatorname{argmax}_x v(Q(x, k) | T) - kx$ be the investment level chosen by an agent with cost k in the equivalent problem with perfect employer information. Using these definitions, I adopt three assumptions regarding problem 54, which can be viewed as restrictions on the investment technology, $Q(x, k)$.

Assumption 5. *The solution to the first best contracting problem, $x_{FB}(k)$, is unique for all k .*

Assumption 6. *For all $k \in K$, $\hat{v}(x)$ is strictly concave around $x_{FB}(k)$.*

Assumption 7. *$\exists \kappa > 0$ such that $\hat{v}''(x) \geq 0 \Rightarrow \hat{v}'(x) > \kappa$ for all $(k, x) \in \mathcal{K} \times \mathbb{R}_+$.*

A sufficient condition for Assumption 5 to hold is that the first best contracting problem is strictly concave, which is always true given sufficient concavity of the investment technology. Assumption 6 simply states that problem 54 is *locally* strictly concave around the first-best investment choice, while assumption 7 is a *global* equivalent that is weaker than strict concavity but stronger than strict quasi-concavity.

Assumptions 5, 6 and 7 jointly ensure that $x(k)$ is differentiable for all $k \in \mathcal{K}$ (see Mailath and von Thadden 2013), which in turn implies that $w(x)$ is differentiable and that the following condition holds for all k :

$$u_c(z(k) - T(z(k)), l(k)) [1 - T'(z(k))] l(k) w'(x(k)) = k \quad (60)$$

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

D. UNPRODUCTIVE SIGNALING: IMPERFECTLY OBSERVABLE INVESTMENT

My final step is to consider the general case in which investment is unobservable. The remaining difference from Section 1 is that employers now observe a noisy 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 1.

The equation for the belief externality, $\mathbf{BE}(z)$, remains very similar to Section 3. There remain distinct *productivity* and *rent transfer* effects, with the change in the equilibrium wage given signal realization $\tilde{\theta}$ given by equation 61.

$$\begin{aligned} \frac{dw(\tilde{\theta}|\pi)}{d[1 - T'(z)]} f(\tilde{\theta}) = \int_{\mathcal{K}} \left(\frac{dx(k|\pi, T)}{d[1 - T'(z)]} \right) & \left[\overbrace{Q_x(x(k|\pi, T), k) f(\tilde{\theta}|x(k|\pi, T))}^{\text{Productivity effect}} \right. \\ & \left. + \underbrace{[Q(x(k|\pi, T), k) - E(q|\tilde{\theta}, \pi)] \left(\frac{\partial f(\tilde{\theta}|x)}{\partial x} \Big|_{x=x(k|\pi, T)} \right)}_{\text{Rent transfer effect}} \right] dG(k) \end{aligned} \quad (61)$$

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.

G Proofs and Derivations

(For Online Publication)

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

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

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

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

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

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

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, \bar{P}_{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 $w = E[q|\theta, \pi]$. \square

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} \ln q \sim \mathcal{N}\left(\frac{\beta}{1 - \beta s(1 + \varepsilon_l)} \ln \beta + \frac{\beta}{1 - \beta s(1 + \varepsilon_l)} \ln s + \frac{\beta(1 + \varepsilon_l)}{1 - \beta s(1 + \varepsilon_l)} \ln (1 - \tau) \right. \\ \left. + (1 - s) \frac{\beta(1 + \varepsilon_l)}{1 - \beta s(1 + \varepsilon_l)} \ln \mu_q + \frac{\beta}{1 - \beta s(1 + \varepsilon_l)} \ln E\left[\xi^{s(1+\varepsilon_l)}\right] \right) \end{aligned}$$

$$-\frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \left(\ln \mu_k - \frac{\sigma_k^2}{2} \right), \left(\frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \right)^2 \sigma_k^2$$

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

$$\sigma_q^2 = \left(\frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \right)^2 \sigma_k^2 \quad (62)$$

$$\mu_q = \left\{ \frac{\beta s (1 - \tau)^{1 + \varepsilon_l} E \left[\xi^{s(1 + \varepsilon_l)} \right]}{\mu_k} \exp \left[\left(1 + \frac{\beta}{1 - \beta s (1 + \varepsilon_l)} \right) \frac{\sigma_k^2}{2} \right] \right\}^{\frac{\beta}{1 - \beta (1 + \varepsilon_l)}} \quad (63)$$

Equation 62 implicitly pins down σ_q^2 in terms of σ_k^2 , β , ε_l and σ_ξ^2 . It is independent of μ_k . In turn, equation 63 characterizes μ_q as a function of the same set of parameters plus μ_k . The elasticity of μ_q with respect to μ_k is $-\beta / [1 - \beta (1 + \varepsilon_l)]$. \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.

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

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

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

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

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

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

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

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

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

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

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

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} .

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

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

$$\mathbf{BE} = (1 - s) E_k (\psi_k \bar{z}_k) \varepsilon_q$$

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

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

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

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

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 $\mathcal{W}(T)$ subject to the four constraints of Problem 5. This problem is restated here for convenience.

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

where:

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

subject to:

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

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

$$v(\theta|\pi, T) = u(w(\theta|\pi) l(\theta|\pi, T) - T(w(\theta|\pi) l(\theta|\pi, T)), l(\theta|\pi, T)) \quad (64)$$

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

$$\mathbf{WL} = -d\tau dz \left\{ \int_{\theta(z|\pi, T)}^{\bar{\theta}} u_c(\theta) \int_{\mathcal{K}} \psi(k) dG(k|\theta) f(\theta) d\theta \right\} \quad (65)$$

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

$$\lambda = \int_{\Theta} u_c(\theta) \int_{\mathcal{K}} \psi(k) dG(k|\theta) f(\theta) d\theta \quad (66)$$

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

$$\mathbf{WG} = d\tau dz \left\{ \int_{\theta(z|\pi, T)}^{\bar{\theta}} f(\theta) d\theta \right\} \lambda. \quad (67)$$

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

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

Since the focal selection $(E(T), T)$ 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 θ 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 λ , the total fiscal externality is as follows.

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

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

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

Applying the envelope theorem and again dividing by λ , the effect of this wage change on social welfare is simply scaled by the affected worker's labor supply, retention rate and

the average welfare weight of an individual with signal realization θ .

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

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

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

Proof of Lemma 3. Firm beliefs about the distribution of productivity in the population must be confirmed in equilibrium and identical across firms. Let π denote the equilibrium set of beliefs. Firm j 's expectation of the worker's productivity is $E[q|\theta, \pi, A_j = 1] \geq 0$. 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 $\bar{P}_{k,\theta} \leq 0$ 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 $\bar{P}_{k,\theta} > 0$ 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 \bar{P}_{MAX} .

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

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, \bar{P}_{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 $l_j \notin \mathcal{L}^*$ where:

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

Some other firm k could offer a contract with the same implied wage as C_j but with $l_k \in \mathcal{L}^*$. 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 $l_j \in \mathcal{L}^*$.

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:

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

Otherwise, firm j could offer a contract with the same implied wage but with $l_j \in \mathcal{L}_j^*$, 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 $l_j^* \in \mathcal{L}_j^*$ be the labor supply offered by firm j . By the definition of \mathcal{L}_j^* we know that:

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

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

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

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

Proof of Proposition 4. Assume – subject to verification – that formal education and unobservable investment are jointly log-normally distributed.

$$\begin{bmatrix} \ln x \\ \ln e \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \ln \mu_x - (1 - \rho_i)^2 \frac{\sigma_x^2}{2} \\ \ln \mu_e - (1 - \rho_i)^2 \frac{\sigma_e^2}{2} \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \rho_i \sigma_x \sigma_e \\ \rho_i \sigma_x \sigma_e & \sigma_e^2 \end{bmatrix} \right)$$

where: σ_x^2 and σ_e^2 are the equilibrium variances of x and e , and ρ_i is the correlation between the two investments.

This implies the following conditional distribution of x .

$$\ln x \sim \mathcal{N} \left(\ln \mu_x - (1 - \rho_i)^2 \frac{\sigma_x^2}{2} + \rho_i \frac{\sigma_x}{\sigma_e} \left(\ln e - \ln \mu_e + (1 - \rho_i)^2 \frac{\sigma_e^2}{2} \right), (1 - \rho_i)^2 \sigma_x^2 \right)$$

Given this, employers face a log-normal signal extraction problem. Conditional on observable investment level e and signal θ , the expectation of log-investment is as follows.

$$\begin{aligned} E(\ln x|\theta, e) &= \tilde{s} \ln x + (1 - \tilde{s}) \left[\ln \mu_x + \rho_i \frac{\sigma_x}{\sigma_e} \left(\ln e - \ln \mu_e + (1 - \rho_i)^2 \frac{\sigma_e^2}{2} \right) \right] \\ &\quad - \frac{1}{2} \frac{(1 - \rho_i)^2 \sigma_\xi^2 \sigma_x^2}{(1 - \rho_i)^2 \sigma_x^2 + \sigma_\xi^2} \right) + \tilde{s} \ln \xi \end{aligned}$$

where $\tilde{s} = \frac{(1 - \rho_i)^2 \sigma_x^2 + \sigma_\xi^2}{(1 - \rho_i)^2 \sigma_x^2 + \sigma_\xi^2}$ is the signal-to-noise ratio for θ conditional on e .

Next, we can calculate a worker's wage, which is equal to her marginal product. Noting that $E(\ln q) = \beta\alpha \ln e + \beta(1-\alpha) E(\ln x)$, we have:

$$\ln w = \kappa_e \ln e + \kappa_x \ln x + \ln \tilde{\mu}_x$$

where:

$$\begin{aligned}\kappa_e &= \beta\alpha + \beta(1-\alpha)(1-\tilde{s})\rho_i \frac{\sigma_x}{\sigma_e} \\ \kappa_x &= \beta(1-\alpha)\tilde{s} \\ \ln \tilde{\mu}_x &= \beta(1-\alpha)(1-\tilde{s}) \left[\ln \mu_x - \rho_i \frac{\sigma_x}{\sigma_e} \left(\ln \mu_e - (1-\rho_i)^2 \frac{\sigma_e^2}{2} \right) \right]\end{aligned}$$

Exponentiating, the level of after-tax wages is: $w = (1-\tau) e^{\kappa_e} x^{\kappa_x} \tilde{\mu}_x \xi^{\kappa_x}$ and labor supply is $l = (1-\tau)^{\varepsilon_l} w^{\varepsilon_l}$. After-tax income is therefore $(1-\tau)z = [(1-\tau) e^{\kappa_e} x^{\kappa_x} \tilde{\mu}_x \xi^{\kappa_x}]^{1+\varepsilon_l}$. Finally, expected utility is:

$$(1-\tau)^{1+\varepsilon_l} \tilde{\mu}_x^{(1+\varepsilon_l)} E \left[\xi^{\kappa_x(1+\varepsilon_l)} \right] \frac{e^{\kappa_e(1+\varepsilon_l)} x^{\kappa_x(1+\varepsilon_l)}}{1+\varepsilon_l} - k_x x - k_e (1-\tau_e) e + \tau \bar{z} - \tau_e \bar{k}_e e.$$

Assuming that $\kappa_x(1+\varepsilon_l) < 1$ and $\kappa_e(1+\varepsilon_l) < 1$ so that individual decisions are characterized by their first-order conditions, the optimal choices are as follows.

$$\begin{aligned}x &= \left(\frac{\kappa_x (1-\tau)^{1+\varepsilon_l} \tilde{\mu}_x^{(1+\varepsilon_l)} e^{\kappa_e(1+\varepsilon_l)} E \left[\xi^{\kappa_x(1+\varepsilon_l)} \right]}{k_x} \right)^{\frac{1}{1-\kappa_x(1+\varepsilon_l)}} \\ e &= \left(\frac{\kappa_e (1-\tau)^{1+\varepsilon_l} \tilde{\mu}_x^{(1+\varepsilon_l)} x^{\kappa_x(1+\varepsilon_l)} E \left[\xi^{\kappa_x(1+\varepsilon_l)} \right]}{k_e (1-\tau_e)} \right)^{\frac{1}{1-\kappa_e(1+\varepsilon_l)}}\end{aligned}$$

Solving this pair of simultaneous equations yields explicit solutions.

$$\begin{aligned}\ln x &= \frac{(1+\varepsilon_l) [\ln(1-\tau) + \ln \tilde{\mu}_x] + [1-\kappa_e(1+\varepsilon_l)] \ln \kappa_x + \kappa_e(1+\varepsilon_l) \ln \kappa_e + E \left[\xi^{\kappa_x(1+\varepsilon_l)} \right]}{1 - (\kappa_x + \kappa_e)(1+\varepsilon_l)} \\ &\quad - \frac{[1-\kappa_e(1+\varepsilon_l)] \ln k_x + \kappa_e(1+\varepsilon_l) \ln k_e}{1 - (\kappa_x + \kappa_e)(1+\varepsilon_l)} - \frac{\kappa_e(1+\varepsilon_l) \ln(1-\tau_e)}{1 - (\kappa_x + \kappa_e)(1+\varepsilon_l)}\end{aligned}\tag{68}$$

$$\begin{aligned}\ln e &= \frac{(1+\varepsilon_l) [\ln(1-\tau) + \ln \tilde{\mu}_x] + \kappa_x(1+\varepsilon_l) \ln \kappa_x + [1-\kappa_x(1+\varepsilon_l)] \ln \kappa_e + E \left[\xi^{\kappa_x(1+\varepsilon_l)} \right]}{1 - (\kappa_x + \kappa_e)(1+\varepsilon_l)} \\ &\quad - \frac{\kappa_x(1+\varepsilon_l) \ln k_x + [1-\kappa_x(1+\varepsilon_l)] \ln k_e}{1 - (\kappa_x + \kappa_e)(1+\varepsilon_l)} - \frac{[1-\kappa_x(1+\varepsilon_l)] \ln(1-\tau_e)}{1 - (\kappa_x + \kappa_e)(1+\varepsilon_l)}\end{aligned}\tag{69}$$

These two equations can be written in matrix form:

$$\begin{bmatrix} \ln x \\ \ln e \end{bmatrix} = c + B \begin{bmatrix} \ln k_x \\ \ln k_e \end{bmatrix}$$

where c is a 2×1 vector of constants, and B is a 2×2 matrix of constants. Since k_x and k_e are jointly log-normal, so are x and e . This proves the first part of the proposition.

Using the equations for $\ln x$ and $\ln e$ above, it is straightforward to derive the elasticities of x and e with respect to $1 - \tau$ and $1 - \tau_e$.

$$\begin{aligned} \varepsilon_{x\tau} &= \frac{1 + \varepsilon_l}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} & \varepsilon_{e\tau} &= \frac{1 + \varepsilon_l}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} \\ \varepsilon_{x\tau_e} &= -\frac{\kappa_e(1 + \varepsilon_l)}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} & \varepsilon_{e\tau_e} &= -\frac{1 - \kappa_x(1 + \varepsilon_l)}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} \end{aligned}$$

Using the fact that $\ln q = \beta\alpha \ln e + \beta(1 - \alpha) \ln x$, we can then derive the elasticities of overall productivity with respect to $1 - \tau$ and $1 - \tau_e$.

$$\varepsilon_{q\tau} = \frac{\beta(1 + \varepsilon)}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} \quad \varepsilon_{q\tau_e} = -\frac{\beta\alpha[1 - \kappa_x(1 + \varepsilon_l)] + \beta(1 - \alpha)\kappa_e(1 + \varepsilon_l)}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)}$$

In turn, the elasticities of income with respect to $1 - \tau$ and $1 - \tau_e$ are as follows.

$$\begin{aligned} \varepsilon_{z\tau} &= \varepsilon_l + (1 + \varepsilon_l) \frac{\beta(1 + \varepsilon)}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} \\ \varepsilon_{z\tau_e} &= -(1 + \varepsilon_l) \frac{\beta\alpha[1 - \kappa_x(1 + \varepsilon_l)] + \beta(1 - \alpha)\kappa_e(1 + \varepsilon_l)}{1 - (\kappa_x + \kappa_e)(1 + \varepsilon_l)} \end{aligned}$$

With the elasticities in hand, we can derive a first-order condition for the optimal tax and education subsidy. There are again three first-order effects of a change the income tax. First, there is the fiscal externality, which takes into account the effect of re-optimization on both government revenue and expenditure on the education subsidy.

$$\begin{aligned} \mathbf{FE}(\tau) &= \tau \bar{\psi} \varepsilon_{z\tau} \frac{\bar{z}}{1 - \tau} - \tau_e \frac{d\bar{k}_e e}{d(1 - \tau)} \\ &= \frac{\tau}{1 - \tau} \bar{\psi} \varepsilon_{z\tau} \bar{z} - \frac{\tau_e}{1 - \tau} \bar{\psi} \varepsilon_{e\tau} \bar{k}_e e \end{aligned}$$

Second, there is the belief externality, which is similar to before.

$$\mathbf{BE}(\tau) = E_i \left[\psi_i \frac{\partial v_i}{\partial \bar{\mu}_x} \frac{\partial \bar{\mu}_x}{\partial \mu_x} \varepsilon_{x\tau} \frac{\mu_x}{1 - \tau} \right]$$

$$= \beta (1 - \alpha) (1 - \tilde{s}) E [\psi_{k_e, k_x} z_{k_e, k_x}] \varepsilon_x$$

Finally, there is the mechanical effect of the transfer.

$$\mathbf{ME}(\tau) = E [\psi_{k_e, k_x} z_{k_e, k_x}] - \bar{\psi} \bar{z}$$

The three effects of a change in the education subsidy, τ_e , are similar. First, there is the fiscal externality.

$$\mathbf{FE}(\tau_e) = \frac{\tau}{1 - \tau_e} \bar{\psi} \varepsilon_{z\tau_e} \bar{z} - \frac{\tau_e}{1 - \tau_e} \bar{\psi} \varepsilon_{e\tau_e} \overline{k_e e}$$

Then there is the belief externality.

$$\mathbf{BE}(\tau_e) = \beta (1 - \alpha) (1 - \tilde{s}) E [\psi_{k_e, k_x} z_{k_e, k_x}] \varepsilon_{x\tau_e}$$

Finally, there is the mechanical effect.

$$\mathbf{ME}(\tau_e) = \overline{\psi k_e e} - E [\psi_{k_e, k_x} k_e e_{k_e, k_x}]$$

Setting the sum of the three effects equal to zero for each instrument, and using the result that $k_e e_{k_e, k_x} = \kappa_e \left(\frac{1 - \tau}{1 - \tau_e} \right) \bar{z}_{k_e, k_x}$, the first-order conditions for the optimal tax and education subsidy are as follows.

$$\frac{\tau}{1 - \tau} = \kappa_e \left(\frac{\tau_e}{1 - \tau_e} \right) \left(\frac{\varepsilon_{e\tau}}{\varepsilon_{z\tau}} \right) + \frac{1 - \gamma}{\varepsilon_{z\tau}} - \frac{\gamma (1 - s) \varepsilon_{q\tau}}{\varepsilon_{z\tau}} \quad (70)$$

$$\frac{\tau_e}{1 - \tau_e} = \frac{1}{\kappa_e} \left(\frac{\tau}{1 - \tau} \right) \left(\frac{\varepsilon_{z\tau_e}}{\varepsilon_{e\tau_e}} \right) + \frac{1 - \gamma}{\varepsilon_{e\tau_e}} - \frac{\gamma (1 - s) \varepsilon_{q\tau}}{\varepsilon_{e\tau_e}} \quad (71)$$

where:

$$s = \frac{\beta \alpha \varepsilon_{e\tau} + \beta (1 - \alpha) \tilde{s} \varepsilon_{x\tau}}{\beta \alpha \varepsilon_{e\tau} + \beta (1 - \alpha) \varepsilon_{x\tau}} = 1 - \beta (1 - \alpha) (1 - \tilde{s}) \frac{\varepsilon_{x\tau}}{\varepsilon_{q\tau}}.$$

The statistic s is the fraction of the social return to higher productivity that workers fails to capture due to employers' imperfect information about x , when they re-optimize in response to changes in τ .

Solving the simultaneous equations above yields the first-order conditions for the optimal tax and education subsidy shown in the proposition.

$$\frac{\tau}{1 - \tau} = M_\tau \left[\frac{1 - \gamma}{\varepsilon_{z\tau}} - \frac{\gamma (1 - s) \varepsilon_{q\tau}}{\varepsilon_{z\tau}} \right] \quad (72)$$

$$\frac{\tau_e}{1 - \tau_e} = M_{\tau_e} \left[\frac{1 - \gamma}{\varepsilon_{z\tau}} - \frac{\gamma(1 - s)\varepsilon_{q\tau}}{\varepsilon_{z\tau}} \right] \quad (73)$$

The constants are the following functions of the elasticities.

$$M_{\tau} = \frac{\kappa_e \left(\frac{\varepsilon_{e\tau}}{\varepsilon_{e\tau e}} \right) + 1}{1 - \left(\frac{\varepsilon_{z\tau e}}{\varepsilon_{e\tau e}} \right) \left(\frac{\varepsilon_{e\tau}}{\varepsilon_{z\tau}} \right)} \quad M_{\tau_e} = \frac{\frac{1}{\kappa_e} \left(\frac{\varepsilon_{z\tau e}}{\varepsilon_{e\tau e}} \right) + \frac{\varepsilon_{z\tau}}{\varepsilon_{e\tau e}}}{1 - \left(\frac{\varepsilon_{z\tau e}}{\varepsilon_{e\tau e}} \right) \left(\frac{\varepsilon_{e\tau}}{\varepsilon_{z\tau}} \right)}$$

□

Proof of Proposition 5. See Stantcheva (2014). □

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

Finally, if $G(\beta(\pi|\tau)) > \pi$ for some π^* then the continuity of $\phi(\theta)$ and G combined with the fact that $G(\beta(1|\tau)) = 0$ implies that $G(\beta(\hat{\pi}|\tau)) = \hat{\pi}$ for some $\hat{\pi} > \pi^*$. There are therefore at least two solutions to equilibrium condition 46. □

Proof of Proposition 7. Social welfare is given by:

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

where:

$$\bar{v}_q(\pi|\tau) = \int_0^1 v(\theta|\pi) dF_q(\theta) - k$$

$$\bar{v}_u(\pi|\tau) = \int_0^1 v(\theta|\pi) dF_u(\theta).$$

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

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

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

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

Proof of Proposition 8. Just as in Sections 2 and 3, there are three effects from a fall in τ . First, there is a mechanical effect. For a worker with signal θ , this is as follows.

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

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

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

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

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

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

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

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

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

□

Proof of Proposition 9. Fixing a value of $T_{A \rightarrow D}$, the proof that condition 48 must hold at the optimum is analogous to the proof of Proposition 8. A similar perturbation argument can be used to establish that condition 49 must hold. An increase in $T_{A \rightarrow D}$ leads to the following gain in welfare for type A and D individuals:

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

The welfare gain, $\lambda_D \Delta_D - \lambda_A \Delta_A$, must be zero at interior optima if $\pi^\dagger(T)$ is locally continuous, implying condition 49. □

Proof of Proposition 10. Assume – subject to verification – that productivity and investment are log-normally distributed.

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

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

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

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

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

For convenience, define $\ln \tilde{\xi} = B \ln \xi$ and let $\ln \tilde{\theta}$ be the following linear transformation of the 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 $s = \sigma_q^2 / (\sigma_q^2 + \sigma_\xi^2) = \sigma_x^2 / (\sigma_x^2 + \sigma_\xi^2)$. A worker's expected level of productivity is therefore a geometric weighted average of A , x , ξ and μ_q .

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The elasticity of productivity follows directly.

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

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

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

Proof of Proposition 11. Using the results from Proposition 10, a worker's expected utility, \bar{v}_n , can be derived in the same way as in Proposition 2.

$$\begin{aligned} \bar{v}_n &= n^{\frac{s(1 + \varepsilon_l)[\alpha + \beta(1 - \alpha)]}{1 - \beta s(1 - \alpha)(1 + \varepsilon_l)}} \left[(1 - \tau)^{(1 + \varepsilon_l)} \mu_q^{(1 - s)(1 + \varepsilon_l)} E \left(\varepsilon^{s(1 + \varepsilon_l)} \right) \right]^{\frac{1}{1 - \beta s(1 - \alpha)(1 + \varepsilon_l)}} [sB]^{\frac{\beta s(1 + \varepsilon_l)(1 - \alpha)}{1 - \beta s(1 - \alpha)(1 + \varepsilon_l)}} \\ &\quad \times \left[\frac{1 - (1 + \varepsilon_l)sB}{1 + \varepsilon_l} \right] + \tau \bar{z} \end{aligned}$$

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.

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

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

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

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

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

Next, we can calculate the belief externality. This is again captured by the effect via μ_q . Using the elasticities from Proposition 10 and the expression for \bar{v}_n , the effect on the welfare of a worker with ability n is:

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

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

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

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

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

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

be optimal, which yields the result.

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

□

H Empirical Evidence

(For Online Publication)

A. CALCULATING THE WEDGE BETWEEN PRIVATE AND SOCIAL RETURNS

In this appendix, I show how to use an estimate of the speed of employer learning to calculate the implied wedge between private and social returns. To do so, I build on the empirical model developed by Lange (2007), which fits the empirical evidence patterns well. The procedure outlined here is how the statistics in Table 1 were calculated.

Following Lange (2007), I assume that the experience profile of log productivity is independent of other factors. An individual's productivity can be written as:

$$\ln q_{i,t} = \tilde{\chi}(e_i, m_i, \eta_i, a_i) + \tilde{H}(t_i) \quad (74)$$

where: e_i is information available to both employers and the researcher (e.g., schooling); m_i is available to employers but not researchers (e.g., job interview performance); a_i is available to the researcher only (e.g., the skills measured by the AFQT); η_i captures factors that are initially hidden from both the researcher and employers; and t_i is experience.

Further assuming that $\tilde{\chi}$ is linear (and suppressing the subscript for individual i from now on) allows us to write:

$$\ln q_t = re + \alpha_1 m + \rho a + \eta + \tilde{H}(t) \quad (75)$$

Then letting (e, q, a, η) be jointly normally distributed, we can write a and η as:⁷

$$\begin{aligned} a &= E[a | e, m] + v = \gamma_1 m + \gamma_2 e + v \\ \eta &= E[\eta | e, m] + u = \alpha_2 e + u \end{aligned}$$

In turn, this allows us to express log productivity as a linear function of the information available to employers at experience level $t = 0$:

$$\begin{aligned} \ln q &= (r + \rho\gamma_2 + \alpha_2) e + (\alpha_1 + \rho\gamma_1) m + (\rho v + u) + \tilde{H}(t) \\ &= E(\tilde{\chi} | e, m) + (\rho v + u) + \tilde{H}(t) \end{aligned}$$

Still following Lange (2007), the process of employer learning is modeled as follows. After each period, a noisy measure of θ_t of $\tilde{\chi}$ becomes available to all employers.

$$\theta_t = \tilde{\chi} + \varepsilon_t \quad (76)$$

⁷As in Lange (2007), we can normalize the coefficient vector and suppress m from the equation for η .

The noise in this signal, ε_τ , is i.i.d. normal with variance σ_ε^2 . It is uncorrelated with all the other variables. Thus, after t years of experience, a t -dimensional vector of measurements $\theta^t = (\theta_0, \theta_1, \dots, \theta_{t-1})$ has become available to employers.

Given this employer learning process, an employer's rational belief about a worker's productivity after t years of experience is characterized by the posterior distribution:

$$\mu_t = (1 - \lambda_t) E(\tilde{\chi} | e, m) + \lambda_t \left(\frac{1}{t} \sum_{\tilde{t}=0}^{t-1} \theta_{\tilde{t}} \right) \quad \sigma_t^2 = \frac{\sigma_0^2 \sigma_\varepsilon^2}{t\sigma_0^2 + \sigma_\varepsilon^2} \quad (77)$$

where σ_0^2 is the variance of $\tilde{\chi}$ conditional on (e, m) – i.e., the variance of the initial expectation error ($\lambda v + e$).

In equation 77, the weight λ_t is given by the following equation.

$$\lambda_t = \frac{tK}{1 + (t-1)K} \quad (78)$$

Here, $K_1 = \sigma_0^2 / (\sigma_0^2 + \sigma_\varepsilon^2)$ is the speed of employer learning, which measures the information content of initial information relative to subsequent measurements. From here, it is straightforward to show that a worker's competitive wage is given by:

$$\ln W(e, m, \theta^t) = (1 - \lambda_t) E(\tilde{\chi} | s, m) + \lambda_t \left(\frac{1}{t} \sum_{\tilde{t}=0}^{t-1} y_{\tilde{t}} \right) + H(t) \quad (79)$$

where $H(t) = \tilde{H}(t) + \frac{1}{2}\sigma_t^2$.

Lange (2007) shows how to use this result to estimate the speed of employer learning, K . However, our focus here is different. Given an estimate of K , I will examine the implications for the discounted return to an increase in a worker's productivity.

An individual's expected lifetime earnings until retirement in period T is:

$$\sum_{t=0}^T \delta^t E[W(e, m, \theta^t) | s, m, z, \eta]$$

where δ is the discount rate. Combining this equation with the wage given by equation 79 we obtain the following expression for the present value of earnings:

$$\sum_{t=0}^T \delta^t E \left[\exp \left((1 - \lambda_t) E(\tilde{\chi} | s, m) + \lambda_t \left(\frac{1}{t} \sum_{\tilde{t}=0}^{t-1} \theta_{\tilde{t}} \right) + H(t) \mid s, m, z, \eta \right) \right]$$

Finally, we can ask what the return is to an increase in $\tilde{\chi}$ that is *not* initially rewarded by

employers.⁸ This private return is given by:

$$\sum_{t=0}^T \delta^t \lambda_t E [W (e, m, \theta^t) | s, m, z, \eta] . \quad (80)$$

By contrast, the social return is simply given:

$$\sum_{t=0}^T \delta^t E [W (e, m, \theta^t) | s, m, z, \eta] .$$

The share of the social return to higher $\tilde{\chi}$ that is captured by workers is simply given by the ratio of the two expressions.

$$s = \frac{\sum_{t=0}^T \delta^t \lambda_t E [W (e, m, \theta^t) | s, m, z, \eta]}{\sum_{t=0}^T \delta^t E [W (e, m, \theta^t) | s, m, z, \eta]} \quad (81)$$

Equation 81 is equivalent to equation 19, but now shows that three empirical objects are required to estimate of the share of the return to investment, s , that is captured by workers. First, we need an estimate of the speed of employer learning, which implies λ_t . Table 1 provides four such estimates. Second, we need the lifecycle profile of expected wages, which I take from Lagakos, Moll, Porzio, Qian and Schoellman (2018). Finally, we need a discount rate.

Figure H1 shows how the pieces fit together. It displays the impact of higher unobservable productivity on wages at each level of potential experience, assuming that $K = 0.259$ and $\delta = 0.95$. The social impact is normalized to one at zero years of experience. It then rises as wages increase, but becomes discounted heavily in later periods. The private return starts at zero, but then converges to the social return in later years as a worker's skill becomes evident. The ratio of the blue shaded area in panel (b) to the total shaded area is the fraction of the discounted social return that is not captured by workers.

Each estimate in Table 1 was calculated in this way for different values of the discount rate and the speed of employer learning. All of the estimates imply a meaningful distortion of the returns to investment. As Figure H1 shows, this is due to a relatively large gap between private and social returns over the first decade of a worker's career.

B. HETEROGENEITY IN EMPLOYER LEARNING BY PRODUCTIVITY LEVEL

In the next part of this appendix, I build on Arcidiacono, Bayer and Hizmo (2010) to provide more direct evidence on how learning varies over the productivity distribution.

⁸Specifically, this implies that $E(\tilde{\chi} | s, m)$ is unaffected. This calculation is supported by Lange's (2007) results, which indicate that there is no initial return to higher ability as measured by the AFQT.

FIGURE H1: PRIVATE AND SOCIAL IMPACT OF HIGHER PRODUCTIVITY

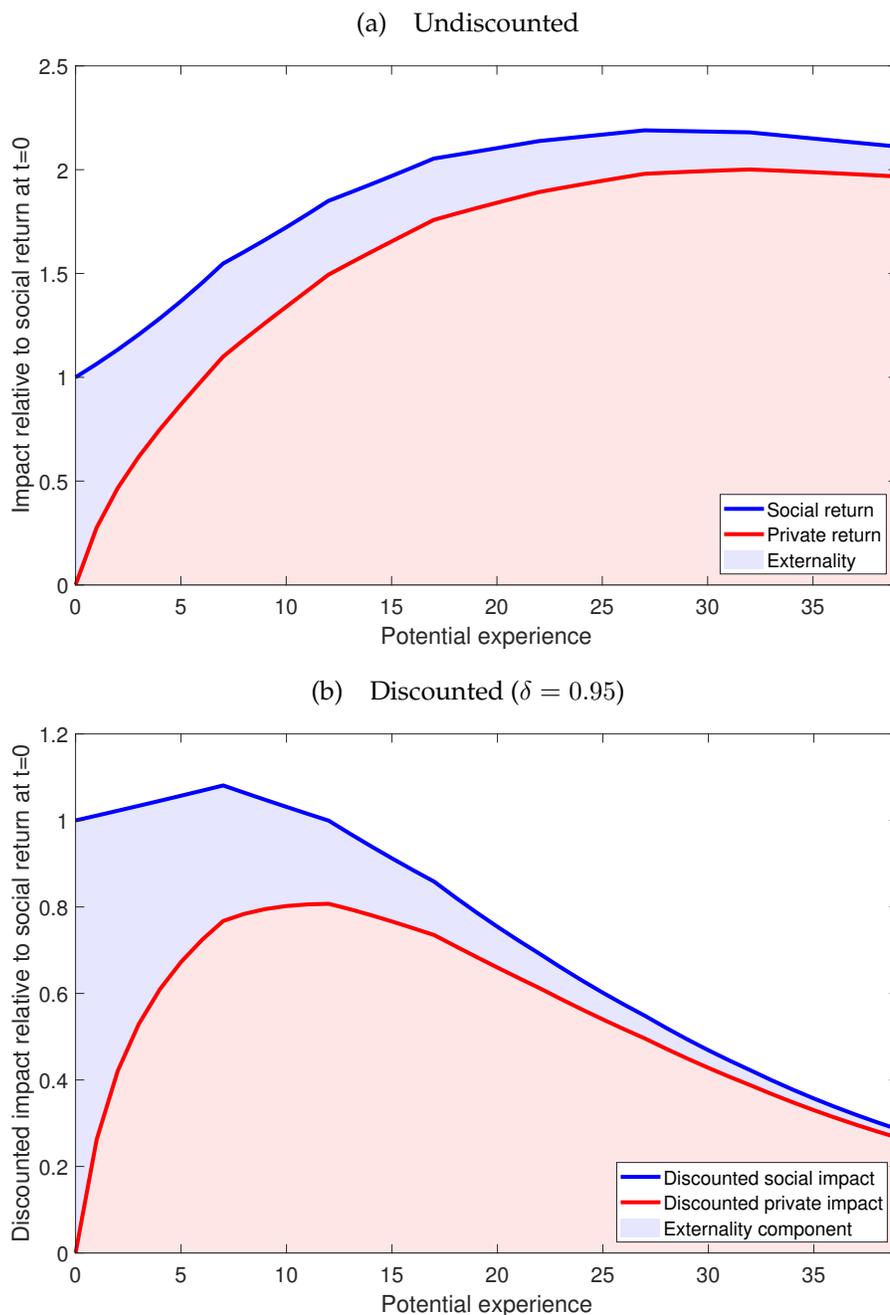


Figure notes. This figure shows the impact of a higher level of initially unobservable productivity on wages at each level of potential experience. The initial social impact is normalized to one. The social impact then rises as wages increase, but becomes discounted heavily in later periods. The private return to higher productivity is initially zero, but converges to the social return in later years as the worker’s skill becomes evident. Panel (a) shows the undiscounted impacts, while panel (b) shows the discounted impacts. The ratio of the blue shaded area to the total shaded area in panel (b) is the fraction of the discounted social return that is not captured by workers. The speed of learning is set to 0.259 as estimated by Lange (2007), and the lifecycle wage profile is taken from Lagakos et al (2018).

Taking AFQT as a proxy for productivity, I adapt equation 20 by interacting the variables of interest with indicators $\mathbf{I}_A = \mathbb{1}(\text{AFQT} > m)$ and $\mathbf{I}_B = \mathbb{1}(\text{AFQT} \leq m)$ for whether a worker's AFQT score is above or below the median, m .

$$\ln w = \sum_{j \in \{A, B\}} \left\{ \begin{aligned} &\rho_{0,j} \text{AFQT} + \rho_{1,j} \text{AFQT} \times \text{Experience} \\ &+ \gamma_{0,j} \text{Education} + \gamma_{1,j} \text{Education} \times \text{Experience} \\ &+ \lambda_{0,j} + \lambda_{1,j} \text{Exper.} + \lambda_2 \text{Exper.}^2 + \lambda_3 \text{Exper.}^3 \end{aligned} \right\} \times \mathbf{I}_j + X' \beta + \varepsilon \quad (82)$$

I estimate equations 20 and 82 using NLSY79 data (Bureau of Labor Statistics 2016).

The sample follows Arcidiacono et al. (2010).⁹ 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.¹⁰ 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.

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; and the less negative interaction between education and experience above the median suggests that differences in learning are driving the results.

⁹Appendix Table I1 provides summary statistics for workers with high and low AFQT scores.

¹⁰The relationship between log wages, AFQT and experience is approximately linear in this region.

TABLE H1: HETEROGENEITY IN EMPLOYER LEARNING

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

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.

I Simulation of the Model

(For Online Publication)

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

A. EVALUATION OF A SINGLE PERTURBATION

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

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

Take the existing wage given each θ but apply T_1 instead of T_0 . Re-optimize labor supply decisions and calculate $v(w(\theta|\pi_0)|T_1)$ for each θ , yielding a candidate vector of utilities $\mathbf{v}_1^{(0)}$. Using $\mathbf{v}_1^{(0)}$, 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 $\mathbf{q}_1^{(1)}$. Use these choices to reconstruct a new candidate $\mathbf{B}_1^{(1)}$ matrix. Then solve for employers' rational productivity inferences at each value of θ , yielding a candidate set of employer beliefs $\pi_1^{(1)}(q)$ and a hypothesized vector of wages $\mathbf{w}_1^{(1)}$.

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

Recalculate utilities to obtain $\mathbf{v}_1^{(1)}$ and adjust workers' investment decisions again, yielding $\mathbf{q}^{(2)}$. 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 θ . Finally, the fiscal externality can be evaluated by subtracting the behavioral effect on tax revenue from all individuals' incomes.

C. SOLVING FOR THE OPTIMAL TAX SCHEDULE

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

D. RECOVERY OF FUNDAMENTALS

To back out fundamentals for the simulation described in Section 4, I begin with the Pareto log-normal approximation of the United States wage distribution provided by Mankiw, Weinzierl and Yagan (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 *fminunc* function to solve for the set of parameters that jointly minimize the Kullback-Leibler divergence between the target wage distribution f_w and the simulated distribution f_w^{sim} .

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

As Figure 8 shows, this process is effective.

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

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

FIGURE I1: OPTIMAL TAX SCHEDULE FROM ALTERNATIVE STARTING POINTS

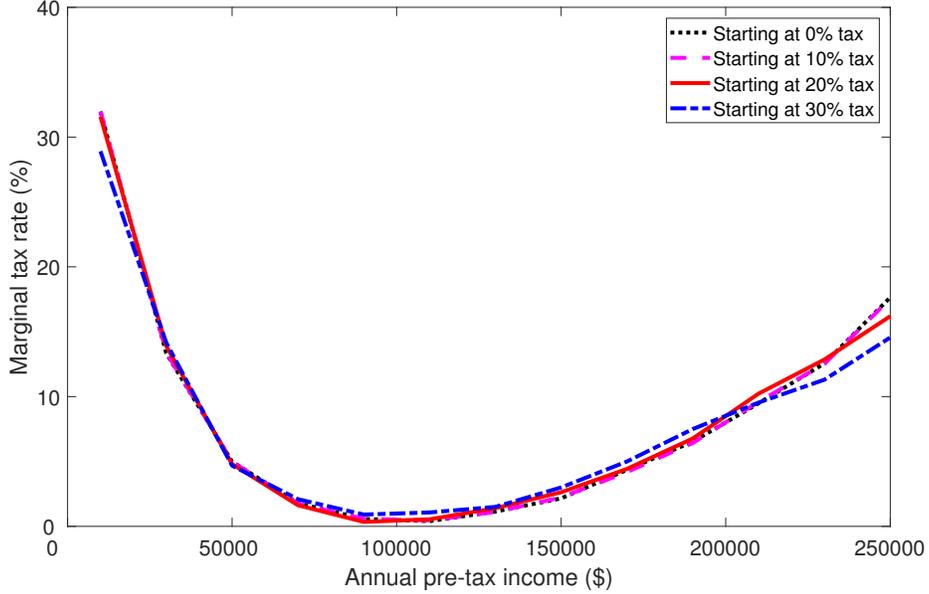


Figure notes. This figure shows the results of simulations as described in Section 4 but starting from different initial tax schedules. Each line shows the optimal tax schedule found using the iterative procedure described here. There remain very minor differences, which could be eliminated only at great computational expense.

where v_0 is a vector that stacks the utility realized at each value of θ and π_0 denotes employers' current and correct beliefs about the distribution of productivity. Combined with individuals' productivity choices and a value of β , this vector of expected utilities can then be used to back out an implied cost distribution.

E. OPTIMAL TAX RATES AT HIGHER INCOME LEVELS

The discussion in Section 4 focuses on the impact of the belief externality on optimal marginal tax rates between \$0 and \$300,000. At higher levels of income, the externality becomes less important for social welfare – and the tax adjustment lower – because it primarily affects individuals with low social welfare weight. Figure I2 shows the impact on the optimal tax schedule at higher levels of income, from \$300,000 up to around \$4 million. Because the extended exercise is very computationally intensive, the size of the tax brackets is larger for this extended exercise.

F. ALTERNATIVE UTILITY AND WELFARE FUNCTIONS

For the simulation introduced in Section 4, I assumed that workers have quasilinear utility, and that social welfare is the average of types' log expected utilities.

$$U = c - l^{1+\frac{1}{\varepsilon_l}} \left/ \left(1 + \frac{1}{\varepsilon_l} \right) - kx \right. \quad \mathcal{W}(T) = \int_K \ln(\text{EU}_k) dG(k) \quad (83)$$

FIGURE I2: OPTIMAL TAX RATES AT HIGHER INCOME LEVELS

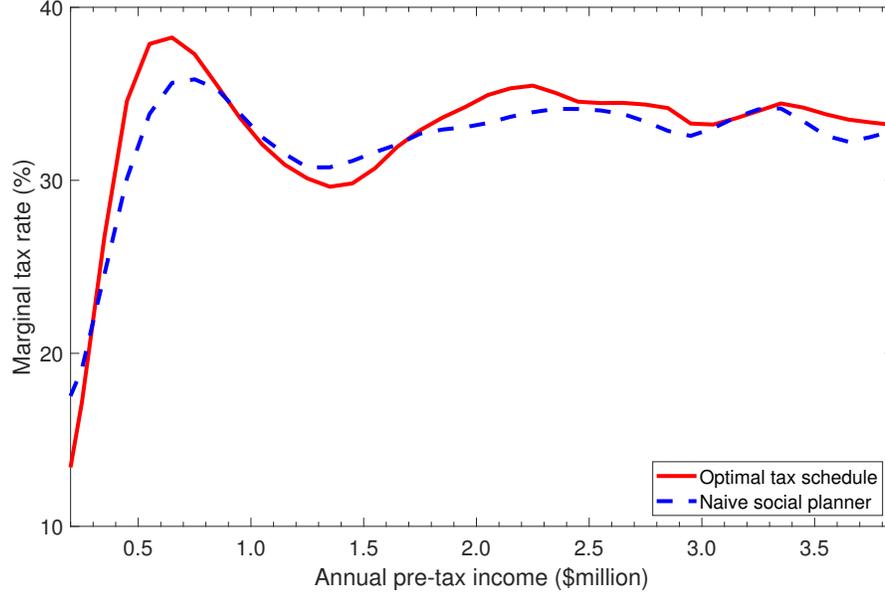


Figure notes. This figure shows the results of a simulation as described in Section 4 when it is extended to higher levels of income. The graph starts at \$300,000 and extends to around \$4 million. The solid red line shows the optimal tax schedule, while the dashed blue line shows a tax schedule that would be accepted by a naïve social planner who sets the sum of the mechanical effect and fiscal externality equal to zero. For computational reasons, the tax function is discretized into \$100,000 brackets rather than \$20,000 brackets.

An alternative is to assume that agents are risk averse over realized consumption, and that the social welfare function is linear.

$$U = \ln \left(c - l^{1 + \frac{1}{\varepsilon_l}} \right) / \left(1 + \frac{1}{\varepsilon_l} \right) - kx \quad \mathcal{W}(T) = \int_{\mathcal{K}} EU_k dG(k) \quad (84)$$

Results with this specification are shown in Figure I3. The results are qualitatively similar to those from the simulation in Section 4. As before, marginal tax rates are generally lower under the optimal than the naïve tax schedule, and the “U” shape of the tax schedule is amplified when the belief externality is taken into account.

There are, however, important quantitative differences. First, marginal tax rates are higher with risk aversion under both the optimal and naïve tax schedules. In part, this is because risk aversion lowers the elasticity of taxable income to around 0.6. In addition, the specification with risk aversion implies a larger benefit to redistribution; the reason for this is that marginal social welfare weights decline more steeply with income, because the planner seeks to equalize *realized* rather than *expected* utilities.

Second, the downward adjustment when the belief externality is taken into account is shifted toward lower incomes. In part, this is again because marginal social welfare

FIGURE I3: UTILITARIAN NON-LINEAR TAXATION WITH RISK AVERSION

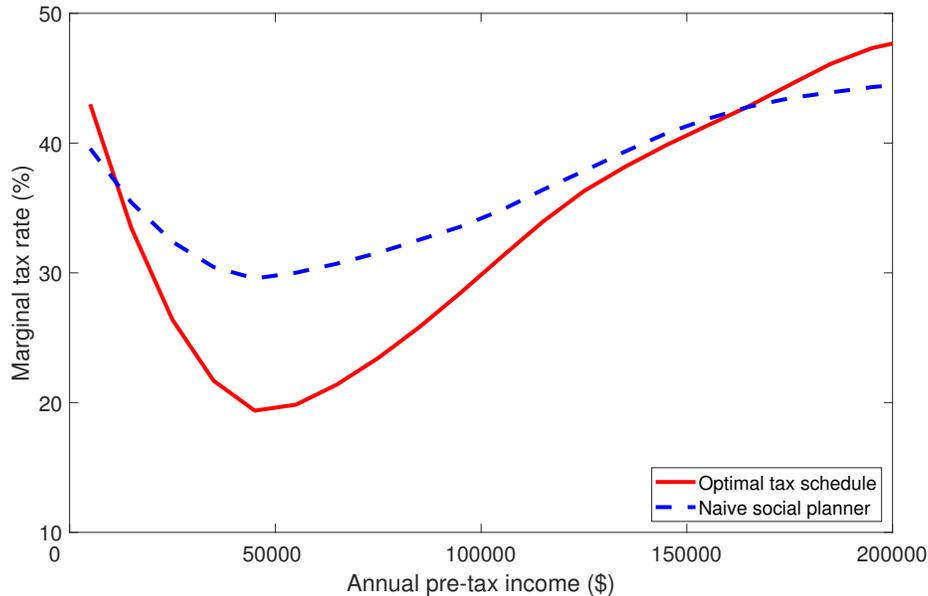


Figure notes. This figure shows the results of a simulation as described in Section 4 except that agents are risk averse and the planner is utilitarian. The solid red line shows the optimal tax schedule, while the dashed blue line shows a tax schedule that would be accepted by a naïve social planner who sets the sum of the mechanical effect and fiscal externality equal to zero. The tax function is discretized into \$20,000 brackets.

weights decline more sharply with realized income; this implies that a given wage impact from the externality matters more at lower incomes than before, and less at higher incomes. In addition, the adjustment to marginal tax rates is slightly smaller because the ratio of the elasticity of productivity to the elasticity of taxable income is lower.

G. ADDING NON-DISCRETIONARY EXPENDITURES

The quantitative exercises throughout the paper focus on redistributive taxation. They therefore assume that there is no separate non-discretionary expenditure requirement. Mathematically, the results are changed if the product of any such expenditure enter individual utilities directly: in this case, the income tax schedule simply needs to be thought of as including the entire system of taxes and expenditures. However, the analysis does change slightly if any such expenditures do not enter individual utilities.

Figure I4 shows the results with a \$5,000 per person exogenous revenue requirement. They are largely unchanged except that marginal tax rates are slightly higher, especially at the low end. Most importantly, the downward adjustment between the naïve and optimal tax schedules is similar with and without the revenue requirement.

FIGURE I4: NON-LINEAR TAXATION WITH NON-DISCRETIONARY EXPENDITURES

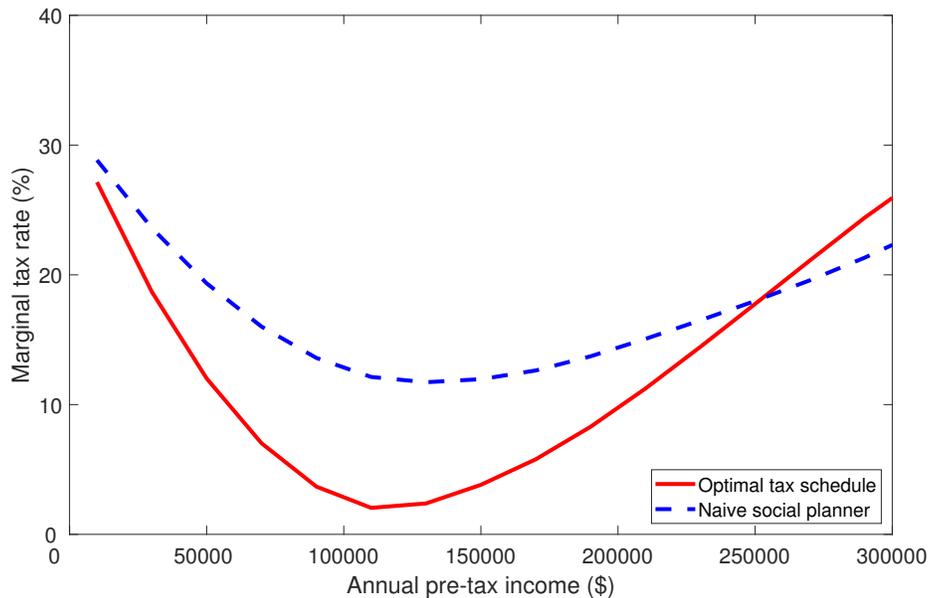


Figure notes. This figure shows the results of a simulation as described in Section 4 except that the government is required to set aside \$5,000 to cover non-discretionary expenditures that do not add to individual utilities. The solid red line shows the optimal tax schedule, while the dashed blue line shows a tax schedule that would be accepted by a naïve social planner who sets the sum of the mechanical effect and fiscal externality equal to zero. The tax function is discretized into \$20,000 brackets.

H. INCIDENCE OF THE BELIEF EXTERNALITY

As I highlighted in Sections 3 and 4, the incidence of the belief externality matters for its effect on welfare. This is because the affected individuals vary in their levels of labor supply, marginal tax rates, and welfare weights. Figure I5 provides a type of decomposition to shed further light on how incidence matters. First, in each tax bracket, I calculate the true impact of the belief externality by following the procedure outlined in part B, above. The solid red line in Figure I5 shows the results, scaled to be relative to the change in average productivity in response to that perturbation.

The remaining lines show what the impact of the belief externality on social welfare would be if we were to abstract from the factors that contribute to the incidence being important. First, the dotted gray line shows what the impact would be if all individuals supplied the average amount of labor, faced the average marginal tax rate, and had the same average welfare weight placed upon them. The dashed blue line takes into account each affected individual’s true level of labor supply to get variation in the pre-tax income impact of the externality. Next, the dashed orange line translates the pre-tax income impact into consumption by allowing marginal tax rates to vary. The remaining difference between the orange and solid red lines is due to variation in welfare weights.

FIGURE I5: INCIDENCE OF THE BELIEF EXTERNALITY

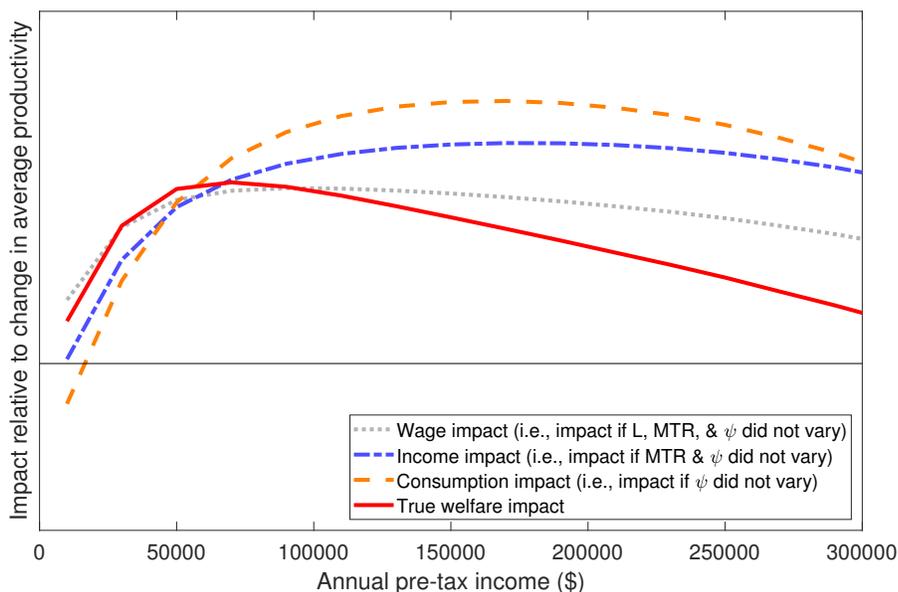


Figure notes. This figure provides a more detailed analysis of the importance of incidence for the impact of the belief externality. The solid red line shows the true impact of the belief externality that arises in response to a small marginal tax cut in each tax bracket, scaled to be relative to the change in average productivity in response to that perturbation. The dotted gray line shows what the impact would be if all individuals supplied the average amount of labor, faced the average marginal tax rate, and had the same average welfare weight placed upon them. The dashed blue line takes into account each affected individual’s true level of labor supply, and the dashed orange line allows the marginal tax rate to vary. The remaining difference between the orange and blue lines is due to variation in welfare weights.

As would be expected based on the equilibrium relationship between productivity and a worker’s expected wage, the wage impact of the externality rises initially and then slowly declines as incomes rise. When differences in labor supply are taken into account, the impact is skewed further toward higher incomes. Differences in marginal tax rates then accentuate the inverse-U shape of the incidence.¹¹ Finally, declining marginal social welfare weights reduce the welfare impact at high incomes. On net, we are left with a strong inverse-U shaped pattern in the impact of the belief externality on welfare.

I. ADDITIONAL FIGURES AND TABLES

Table I1 provides summary statistics for data used to test for employer learning in Section 4. Figures I6 to I8 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 4. Figure I9 shows the expected net transfer from the government for workers of

¹¹At very low incomes, this adjustment amplifies a negative impact on high income individuals enough so that the average consumption impact is negative, although the final impact on social welfare is still positive.

each initial productivity level. Figure I10 plots the utility gain for workers with at each initial productivity level. Finally, Figure I11 shows the change in marginal social welfare weights starting from naïve taxation and transitioning to optimal taxation.

TABLE I1: SUMMARY STATISTICS FOR HIGH AND LOW AFQT WORKERS

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

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

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

FIGURE I6: COMPARISON OF FISCAL EXTERNALITY

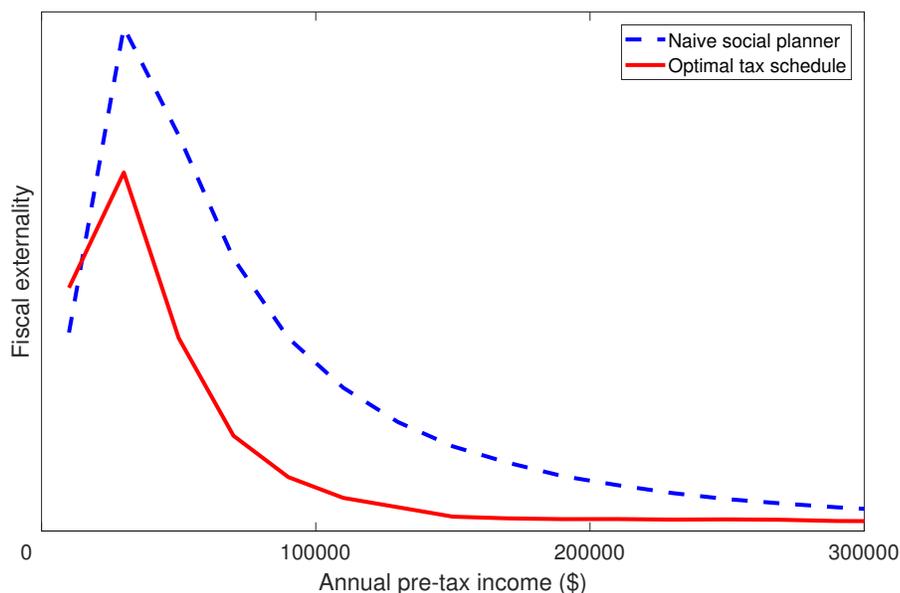


Figure notes. This figure compares the fiscal externality in each tax bracket under naïve and optimal taxation, in the simulation described in Section 4.

FIGURE I7: COMPARISON OF BELIEF EXTERNALITY

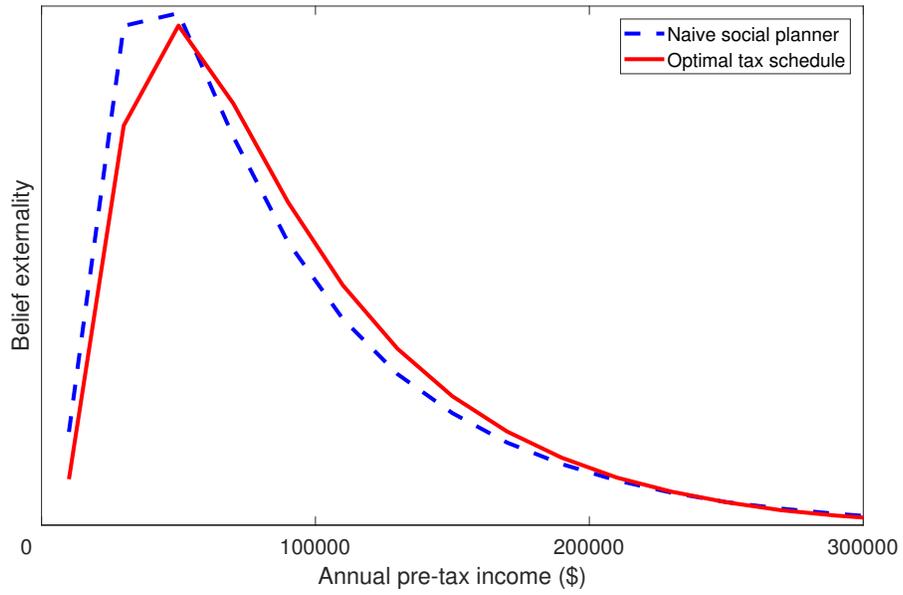


Figure notes. This figure compares the belief externality in each tax bracket under naïve and optimal taxation, in the simulation described in Section 4.

FIGURE I8: COMPARISON OF MECHANICAL EFFECT

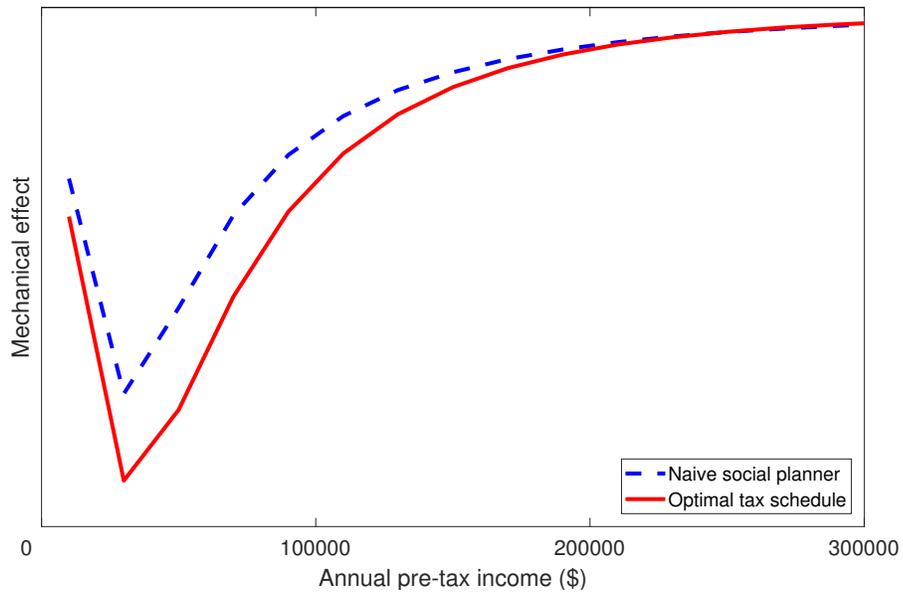


Figure notes. This figure compares the mechanical effect in each tax bracket under naïve and optimal taxation, in the simulation described in Section 4.

FIGURE I9: EXPECTED NET TRANSFER

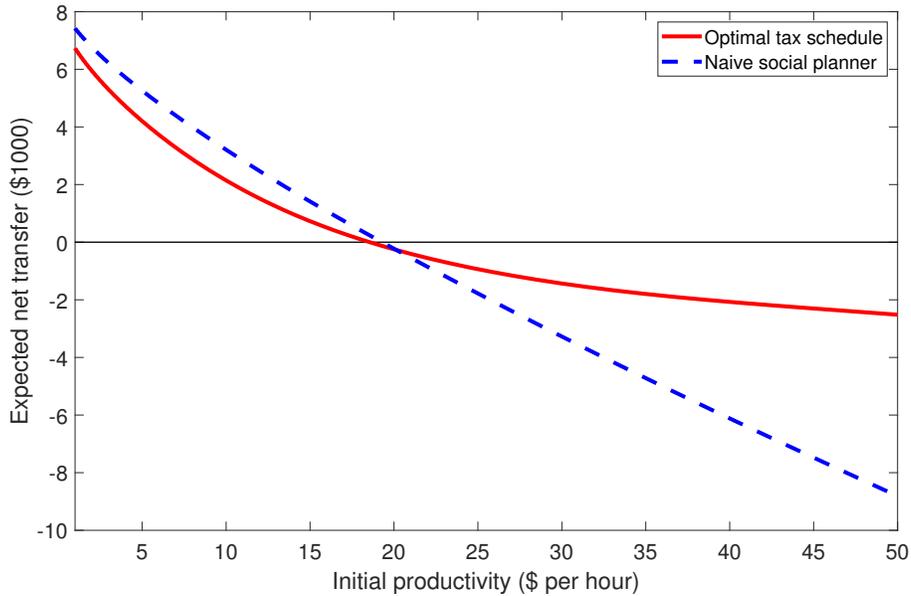


Figure notes. This figure plots the expected net transfer from the government for workers of each initial productivity level for the simulation described in Section 4. The solid red line shows the transfer under the optimal tax schedule, while the dashed blue line shows the transfer under the naïve tax schedule.

FIGURE I10: UTILITY GAIN FROM OPTIMAL TAXATION

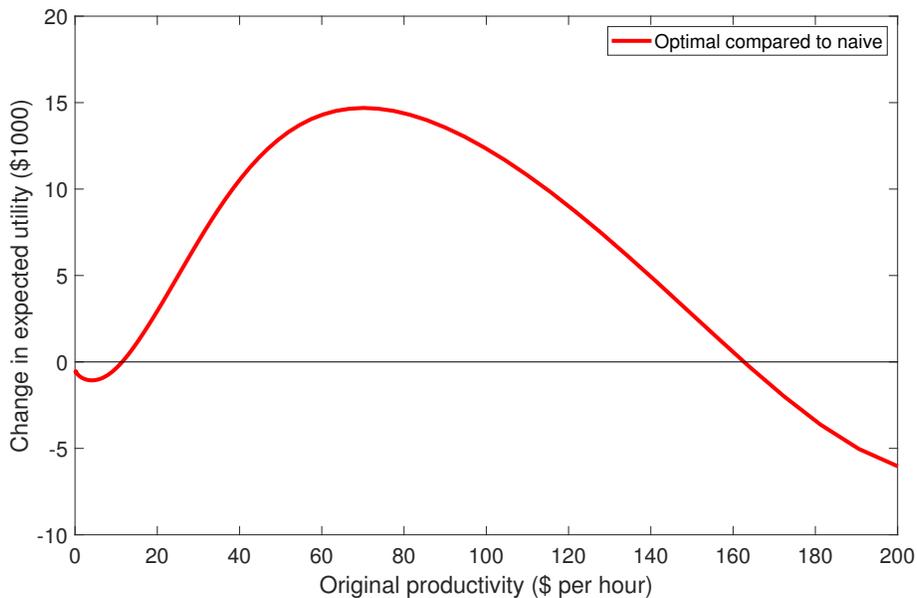


Figure notes. This figure compares the utility levels of agents at each initial productivity level under naïve and optimal taxation in the simulation described in Section 4.

FIGURE I11: MARGINAL SOCIAL WELFARE WEIGHTS

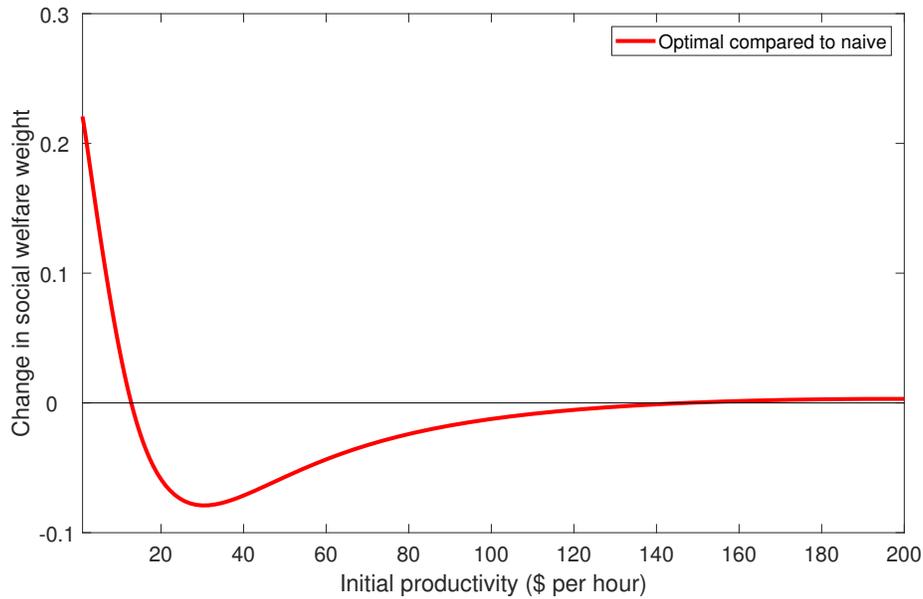


Figure notes. This figure plots the change in marginal social welfare weights starting from naïve taxation and transitioning to optimal taxation, for the simulation described in Section 4.

References

- Aakvik, Arild, Kjell G. Salvanes, and Kjell Vaage,** “Measuring Heterogeneity in the Returns to Education using an Education Reform,” *European Economic Review*, 2010, 54 (4), 483–500.
- Acemoglu, Daron and Jorn-Steffen Pischke,** “Why Do Firms Train? Theory and Evidence,” *Quarterly Journal of Economics*, 1998, 113 (1), 79–119.
- Altonji, Joseph G. and C. Pierret,** “Employer Learning and Statistical Discrimination,” *Quarterly Journal of Economics*, 2001, 116 (1), 313–50.
- Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo,** “Beyond Signaling and Human Capital: Education and the Revelation of Ability,” *American Economic Journal: Applied Economics*, 2010, 2 (4), 76–104.
- Arrow, Kenneth,** “The Theory of Discrimination,” in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, New Jersey: Princeton University Press, 1973, pp. 3–33.
- Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange,** “Signaling and Employer Learning with Instruments,” *NBER Working Paper No. 25885*, 2020.

- Becker, Gary S.**, *The Economics of Discrimination*, Chicago: Chicago University Press, 1957.
- Bedard, Kelly**, "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy*, 2001, 109 (4), 749–775.
- Bureau of Labor Statistics**, *National Longitudinal Survey of Youth - 1979*, United States Department of Labor, 2016.
- Coate, Stephen and Glenn C. Loury**, "Antidiscrimination Enforcement and the Problem of Patronization," *American Economic Review*, 1993, 83 (2), 92–98.
- Fernandez, Luis and Eric Rasmusen**, "Perfectly Contestable Monopoly and Adverse Selection," *Working Paper*, 1993.
- Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow**, "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 2019, 87 (5), 1439–1474.
- Kahn, Lisa B. and Fabian Lange**, "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *Review of Economic Studies*, 2014, 81 (4), 1575–1613.
- Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman**, "Life Cycle Wage Growth across Countries," *Journal of Political Economy*, 2018, 126 (2), 797–849.
- Landers, Renée M., James B. Rebitzer, and Lowell J. Taylor**, "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," *American Economic Review*, 1996, 86, 329–348.
- Lang, Kevin and David Kropp**, "Human Capital Versus Sorting: The Effects of Compulsory Attendance Laws," *Quarterly Journal of Economics*, 1986, 101 (3), 609–624.
- Lange, Fabian**, "The Speed of Employer Learning," *Journal of Labor Economics*, 2007, 25 (1), 1–35.
- **and Robert H. Topel**, "The Social Value of Education and Human Capital," in Eric A. Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 1, Elsevier, 2006, chapter 8, pp. 459–509.
- Mailath, George J. and Ernst Ludwig von Thadden**, "Incentive Compatibility and Differentiability: New Results and Classic Applications," *Journal of Economic Theory*, 2013, 148 (5), 1841–1861.

- Mankiw, N. Gregory, Matthew Weinzierl, and Danny Yagan**, "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 2009, 23 (4), 147–174.
- Meghir, Costas and Marten Palme**, "Educational Reform, Ability, and Family Background," *American Economic Review*, 2005, 95 (1), 414–424.
- Mimra, Wanda and Achim Wambach**, "Contract Withdrawals and Equilibrium in Competitive Markets with Adverse Selection," *Economic Theory*, 2019, 67 (4), 875–907.
- Miyazaki, Hajime**, "The Rat Race and Internal Labor Markets," *The Bell Journal of Economics*, 1977, 8, 394–418.
- Moro, Andrea and Peter Norman**, "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory*, 2004, 114 (1), 1–30.
- Netzer, Nick and Florian Scheuer**, "A Game Theoretic Foundation of Competitive Equilibria with Adverse Selection," *International Economic Review*, 2014, 55 (2), 399–422.
- Oreopoulos, Philip**, "The Compelling Effects of Compulsory Schooling: Evidence from Canada," *Canadian Journal of Economics*, 2006, 39 (1), 22–52.
- Rothschild, Michael and Joseph E. Stiglitz**, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," 1976.
- Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin**, "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *Working Paper*, 2019.
- Spence, Michael**, "Job Market Signaling," *Quarterly Journal of Economics*, 1973, 87 (3), 355–374.
- , "Product Differentiation and Performance in Insurance Markets," *Journal of Public Economics*, 1978, 10, 427–447.
- Stantcheva, Stefanie**, "Optimal Income Taxation with Adverse Selection in the Labour Market," *Review of Economic Studies*, 2014, 81 (3), 1296–1329.
- Wilson, Charles**, "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 1977, 16 (2), 167–207.