Tasks, Talents, and Taxes*

Laurence Ales† Musab Kurnaz‡ Christopher Sleet§

This version: February 15, 2014

Preliminary!

Abstract

A large positive literature emphasizes the role of technological change in driving the demand for skill and talent. We consider the normative implications of such technical change for policy design.

1 Introduction

Technical change is inherently redistributive, complementing the effort of some while substituting for that of others. However, while the positive literature documenting the redistributive nature of technical change is extensive, normative work exploring the policy implications of such change is not. Our paper fills this gap. We make theoretical and quantitative contributions. On the theoretical side we relate existing models of optimal taxation to the assignment framework used by Acemoğlu and Autor (2011), Autor and Dorn (2013) and Costinot and Vogel (2010) and others to analyze the implications of technical change. In this way, we show how the key technological parameters emphasized by the positive literature shape the (labor demand) elasticities central for optimal tax analysis. We show how a parametric assignment model can be brought to the data; we estimate the key parameters and derive the implications of technical change from the 1970s to the present day for policy. We find, consistent with other positive studies, evidence of increasing comparative advantage of high talent workers in complex tasks and

*We thank participants at the Larry Jones 60th Birthday Conference and, especially, our discussant Marco Bassetto for very useful comments.

†Tepper School of Business, Carnegie Mellon University. Email: ales@andrew.cmu.edu

‡Tepper School of Business, Carnegie Mellon University. Email: mkurnaz@andrew.cmu.edu

§Tepper School of Business, Carnegie Mellon University. Email: csleet@andrew.cmu.edu

of polarizing shifts in demand from middle-ranked to low and high complexity tasks. We show that the optimal policy response is to reduce marginal income taxes on low and middle talent workers, but leave those on higher talent workers intact. Transfers to the lowest talents should be phased out.

To formalize the impact of technical change on the distribution of workers across occupations and wages several recent studies have adopted an assignment framework.[1] In this framework differently talented workers sort themselves across tasks differentiated by complexity; highly talented workers are assumed to have a comparative advantage in complex tasks and, sometimes, an absolute advantage in all tasks.[2] Despite this work's implication that imperfect substitutability of talent across tasks exposes workers to the risk of technical change, it has made little contact with the normative literature on optimal taxation. Moreover, most contributions abstract from workers' intensive effort margin. In contrast, normative work focusses on the incentive to supply effort and largely ignores the role of the demand for talent in shaping income inequality. An exception is Stiglitz (1982) who allows for imperfect substitutability between the effort of different talents. He shows, in a two talent model, that this assumption has important implications for the design of optimal taxes. In particular, he identifies a wage compression motive for subsidizing high and taxing low talents. By doing so the wage of high talents is compressed relative to low and the former's incentive constraints are relaxed.

The starting point of our analysis is a generalization of Stiglitz (1982) from two to K talents with an aggregate production function defined directly upon the (imperfectly substitutable) effort of these talents. Stiglitz (1982)'s wage compression channel remains operable, but now takes a more complex form: the motive to tax a given talent type k at the margin depends, in part, on the elasticity of the relative wages of all pairs of adjacent talent types (ordered by wages) with respect to k's effort. This analysis suggests that the implications of an assignment framework for policy will depend upon that framework's implications for the interaction of the effort of different talents. We next enrich an assignment model with various elements from the normative literature: an intensive effort margin, standard concave preferences over consumption and effort and a Paretian gov-

[1] Costinot and Vogel (2010), Acemoğlu and Autor (2011) and Autor and Dorn (2013) are prominent examples.

[2] The assignment framework is not new, originating with Roy (1950) and being extended by Sattinger (1975) and Teulings (1995). It has proven to be a rich laboratory for analyzing the role of talent-task distributions and the productivity of task-talent matches in shaping the wage distribution. Over the last few years there has been a renewed emphasis on comparative statics in this framework and on its use as a lens to view the implications of technical change. Acemoğlu and Autor (2011) emphasize the value of the assignment framework for analyzing many issues over what they call the canonical model in which workers are locked into imperfectly substitutable tasks.

ernment that can tax incomes. In this model workers sort themselves efficiently across tasks *conditional on the effort of other workers.* This induces an indirect production function over the effort of different talents and, hence, micro-founds such a function. The wage distribution and demand elasticities necessary for tax analyses are, thus, related to the technological parameters emphasized in the positive literature. In particular, a perturbation of one talent type's effort raises the supply of labor in that type's current assigned task. This depresses the associated (shadow) task price, causing this type to migrate into neighboring tasks. This in turn raises the labor input in these tasks, causing the talent types occupying them to migrate. Thus, a ripple effect is created and since the relative wages of adjacent talents depend upon their task assignments,[3] an adjustment in one talent type's effort can affect relative wages of all talent types, in the process relaxing and tightening many incentive constraints. The degree to which this migration occurs and the degree to which it affects relative wages when it does depends upon the pattern of comparative advantage of talents across tasks and the pattern of demand for different tasks.

Analytical characterization of the implications of technical change for policy in such settings is complicated. We present some simple two talent examples and use them to highlight such implications analytically and numerically.

We then take the model of the data. Our main data source is the Current Population Survey (CPS). This dataset contains detailed information on the occupation, income and hours worked of respondents. We treat this data as if it were generated by an equilibrium of our assignment model and use parametric assumptions and equilibrium restrictions to recover estimates of key technological parameters for the 1970s and the 2000s. In particular, we order occupations by the average wage paid and, hence, relate them to sets of tasks ordered by complexity. We recover an empirical proxy for the assignment of tasks to talents from the distribution of workers across occupations (ordered by wages). We separate the estimation of parameters determining the demand for tasks from those determining the comparative advantage of talents across tasks by assuming a Cobb-Douglas technology for final goods as a function of tasks. This enables us to identify the demand parameters with occupational compensation shares. The parameters determining the productivity of talent-task matches and, hence, comparative advantage of high talents for complex tasks are then recovered from the empirical assignment function and the distribution of wages across tasks. After obtaining these estimates and supplementing them with calibrated preference parameters, we calculate optimal tax policies for the 1970s and 2000s

[3]More precisely, it depends on their relative comparative advantage in the task threshold that demarcates them.

and, hence, evaluate the implications of 40 years of technical change for policy design.

LITERATURE As noted above, our paper bridges the normative optimal taxation literature and a positive literature that analyzes the role of technical change in driving the wage distribution. Both literatures are large. In addition to the papers referenced above, we mention two that are quite related to ours. Rothschild and Scheuer (2013) also develop implications for optimal tax policy in an assignment framework. In contrast to us, they focus on the two task-many talent case with, in their theory, essentially no restrictions on the distribution of talent across tasks. In this setting they provide an interesting characterization of optimal taxes and contrast such taxes with those obtained in self-confirming equilibria in which governments do not recognize their ability to influence the wage distribution. We focus on many task-many talent cases, but place much stronger restrictions on the task-talent distribution. This allows us to relate our model to those used to analyze technical change in the positive literature and to undertake comparative static exercises which Rothschild and Scheuer (2013) do not do. In addition, Rothschild and Scheuer (2013) draw a sharp distinction between their conclusions and those of Stiglitz (1982) by identifying Stiglitz (1982)'s model with one in which talents cannot be substituted across tasks. In fact, Stiglitz (1982) is silent on tasks and his model is consistent with the (indirect) aggregate production function over effort derived from Rothschild and Scheuer (2013)'s (or our) framework.

Salanié (2011) suggests applying the logic of Stiglitz (1982) to capital taxation. This is done by Slavík and Yazici (2012) who introduce two sorts of capital, buildings and machines. Following the skill premium literature, they assume a machine-skill (or machine-talent) complementarity. Thus, machines raise the marginal product of the talented relative to the untalented and, as in Stiglitz (1982), this dilutes incentives. It is socially desirable to deter the accumulation of machines. In quantitative work, Slavík and Yazici (2012) show that this creates a rationale for quite high rates of (machine) capital taxation.

2 FACTS ON TECHNICAL CHANGE AND TAXATION

Before illustrating our theoretical framework we document some of the key facts that motivate our analysis. Our many data source is the Current Population Survey (CPS)[4] Our sample is from 1975 to 2011. In the bulk of the paper we group observations into two time periods. We will call "the 70s" observations relating to the years 1975-1979 and we will call the "00s" observation relating to the years 2000-2011.

[4]Refer to Appendix E for additional details on the data and our sample selection.

The CPS reports information regarding both the sector and occupation of each worker. We begin by looking at the behavior of sectors over time. Original information on sector is provided at the 3-digit level, for a total of approximately 300 sectors. In Figure 1 we display the changes in the relative size of sectors (in terms of share of national income) aggregated at the 1-digit level of detail.[5] What we observe is a well documented fact:

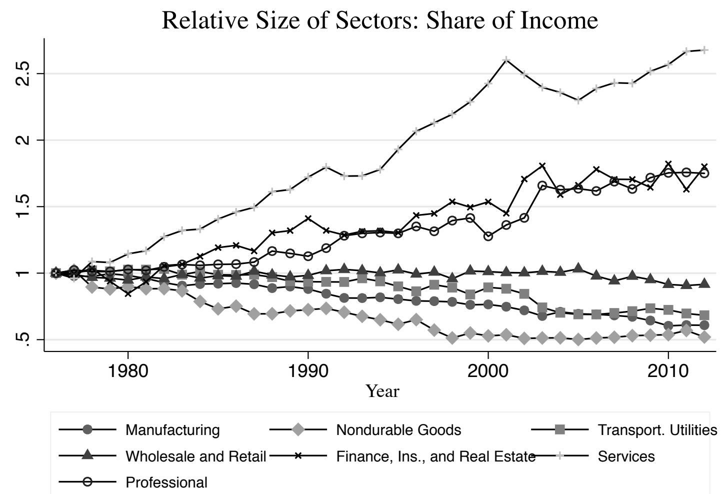
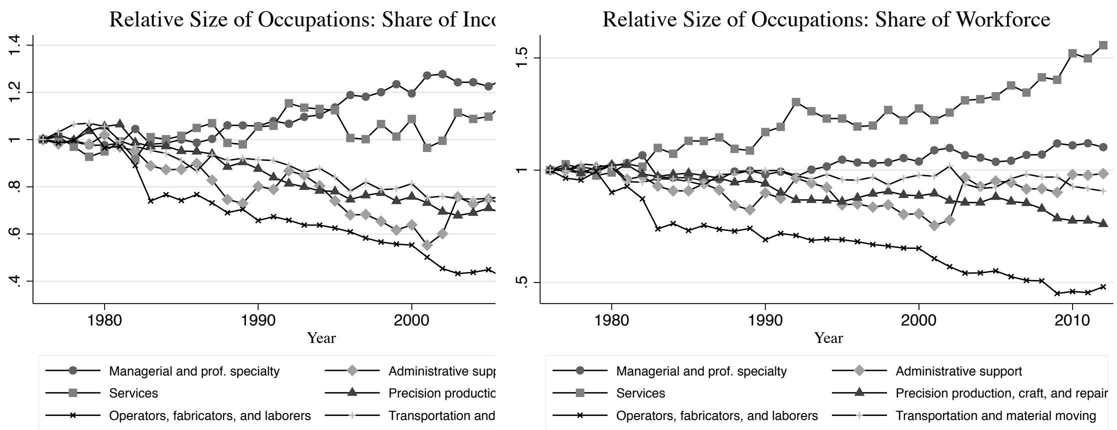


Figure 1: Changes in Share of National Income by Sector Over Time. Source: CPS.

over time there has been a large shift in the relative importance of sectors. In particular we see sectorial "polarization" occurring. A groups of sectors (Services, Financial and Professional) is growing over time while another group of sectors (Manufacturing, Non-durable goods, Wholesale and Transportation) is shrinking over time.

Since these two groups of sectors use different occupations with different relative intensities we would expect to observe some degree of polarization also occurring when looking at occupations. In the CPS information on occupations is also provided at the 3-digit level. In Figure 2a, we display the changes in the share of national income of different occupations over time. In this case also we see the evidence of "polarization" with managerial and service related occupations growing in importance over time while others, for example: operators and fabricators (mostly employed in manufacturing) are shrinking over time. What is interesting is the fact that the two occupations growing over time are positioned, on average, at the opposite end of the income distribution. Hence occupations that are growing over time are both those with very high and with very low income per capita. This peculiarity of technical change has been pointed out by Acemoğlu

[5]Not shown are primary sectors and public administration.



(a) Share of National Income. (b) Share of Employment.

Figure 2: Changes in Size of Occupations Over Rime. Source: CPS

and Autor (2011) when looking at the share of employment (shown in Figure 2b).[6]

The picture that emerges from the CPS is one in which high wage and low wage occupations are growing relative to middle ones. This phenomenon raises a potential additional redistributive motive from the part of the policy makers. A natural channel for this redistribution is income taxation which we analyze next. Our goal is to determine the degree of "progressivity" of the tax code over time. We begin by looking at the profile of average tax rates. Within CPS we can compute effective tax rates for the entire income distribution. To compute effective average tax rate we encode the information provided on CPS into the NBER TAXSIM simulator.[7] For each individual we use information on the year, state of residence, dependents, marital status, total and individual income. In Table 1, we display the effective tax rate of state and federal tax, with and without payroll taxes (note that we are assuming that the burden of the payroll tax falls entirely on the worker).

[6] In Autor and Dorn (2013) this feature of the data is mapped to the behavior of preferences. Over time a higher weight in the utility function is given to those low skilled service jobs and high skilled managerial occupations.

[7] Gouveia and Strauss (1994) emphasize the usage of effective tax rates as opposed to statutory tax rates to correctly capture the differences between market and taxable income as well as the worker's economic response to statutory tax rates.

Table 1: Average Tax Rates on Real Labor Income. Source: CPS And TAXSIM 9.6.

Taxes Included	**Decade**	**Percentiles of Income**					
		10th	*25th*	*50th*	*75th*	*90th*	*99th*
State + Federal	70s	10.8	15.9	19.5	21.9	25.6	35.4
	00s	2.6	11.8	18.2	22.2	26.1	35.0
State + Federal + FICA	70s	26.1	30.9	34.0	35.0	35.6	40.9
	00s	23.7	33.1	38.9	41.7	41.9	41.3

Between the two time periods we observe that the changes in effective tax rates are quantitatively small especially for people above the median income. Overall, the reduction in tax rates seems to point to a small increase in progressivity.[8] To provide an objective measure of the changes in progressivity over time we look at the aggregate *residual income elasticity* (formally defined below). This measure was first proposed by Musgrave and Thin (1948) and was later axiomatized by Jakobsson (1976) and Pfingsten (1986). It has been shown to have numerous desirable properties relative; in particular, Jakobsson (1976) establishes that:

[...] if one tax code is, according to the measure, more progressive than another, then it should also be unambiguously more redistributive than the other.

Thus, the measure establishes a partial order over tax codes, with one code ranked more progressive than another if it induces after-tax Lorenz curves that dominate those induced by the other code. Denote with $y_{i,t}$ the income of individual i at time t and $T(y_{i,t})$ the amount of tax paid by individual i at time t. We define the individual residual income elasticity of i at t as:

$$\eta_{i,t} := \frac{d[y_{i,t} - T_t(y_{i,t})]}{dy_{i,t}} \frac{y_{i,t}}{y_{i,t} - T_t(y_{i,t})}. \tag{1}$$

Re-expressing (1) in terms of the marginal (τ_i) and average ($\bar{t}_i$) tax rate of individual i gives:

$$\eta_{i,t} = \frac{1 - \tau_{i,t}}{1 - \bar{t}_{i,t}}.$$

[8]This contrasts with the conclusions ofPiketty and Saez (2007) who emphasize a reduction in progressivity. The difference stems from their focus on marginal tax rates at the very top of the income distribution within the top percentile.

Let there be N_t individuals at time t, then the *residual income elasticity* is defined as

$$H_t := \frac{1}{N_t} \sum_{i=1}^{N_t} \eta_{i,t}. \tag{2}$$

The progressivity of the tax code is decreasing in H_t with a value less than one denoting a progressive code. We compute the value of H_t for the US over the period 1975-2011 using CPS data and plot the results in Figure 3.

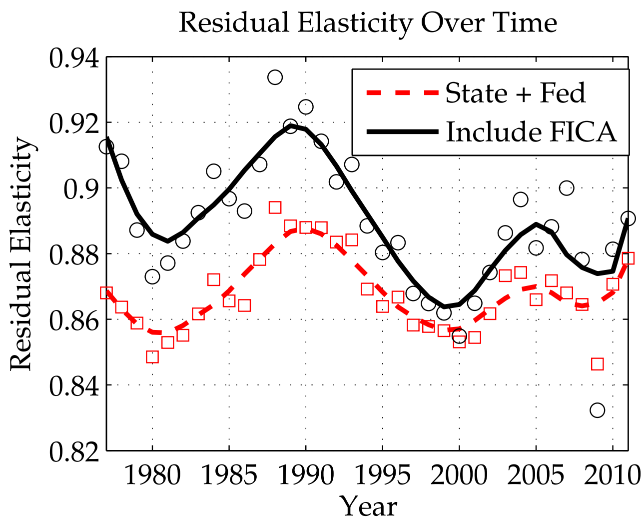


Figure 3: Residual Elasticity Over Time: Source CPS and TAXSIM.

These calculations indicate that the joint state and federal tax code is overall progressive and that, as expected, payroll taxes reduce the degree of progressivity. They show that the degree of progressivity implied by state and federal taxes shows very little trend over the last 40 years, with, perhaps, a modest increase in progressivity after accounting for payroll taxes. Thus, successive tax reforms have tended to only temporarily increase or decrease the degree of progressivity around a stable level.[9]

In summary, we have documented a significant occupational restructuring of the US economy. This is suggestive of technical change that has enhanced the demand for labor and talent in some activities and suppressed it in others. Despite the potential redistributive implications of these long run, secular changes, we find little evidence of any sustained policy response in the direction of either more or less redistribution in the US.

[9]This is consistent with Kasten, Sammartino, and Toder (1994) who find little evidence of changes in the progressivity of the US tax code between 1980 and 1993 despite the enactment of five major tax reforms.

3 A GENERALIZED STIGLITZ MODEL OF TAXATION

Mirrlees (1971)'s classic model of optimal taxation assumes that final output is a weighted sum (or integral) of agent effort, with the weights given by private productivities. Stiglitz (1982) allows for a more general production function that is constant returns to scale in agent effort. He restricts attention to workers of two types and assumes directly that both before and after the implementation of taxes one type has a higher marginal product than the other. In this section, we generalize Stiglitz (1982) to K-types.[10]

3.1 PHYSICAL ENVIRONMENT

AGENTS A continuum of agents has identical preferences over consumption $c \in \mathbb{R}_+$ and effort $e \in [0, \bar{e}]$ described by a utility function $U : \mathbb{R}_+ \times [0, \bar{e}] \to \mathbb{R}$. The function U is assumed to be concave, twice continuously differentiable on the interior of its domain, with for each $e \in [0, \bar{e}]$, $U(\cdot, e)$ increasing and for each $c \in \mathbb{R}_+$, $U(\cdot, c)$ decreasing and strictly concave. First and second partial derivatives of U are denoted U_l and U_{lm} with $l, m \in \{c, e\}$. U satisfies the Inada conditions: for all $c > 0$, $\lim_{e \downarrow 0} U_e(c, \cdot) = 0$ and $\lim_{e \uparrow \bar{e}} U_e(c, \cdot) = \infty$. In addition, U satisfies the Spence-Mirrlees single crossing property: $-U_e(c, y/w)/\{wU_c(c, y/w)\}$ is decreasing in w.[11] Agents are partitioned across $K \in \mathbb{N} \backslash \{1\}$ types with a fraction π_k of agents in type group $k \in \{1, \dots, K\}$. $\Pi_k = \sum_{j=1}^{k} \pi_k$ is the fraction of agents with type less than or equal to k.

Agents sell their labor to firms and pay taxes on the income that they earn. Let $T : \mathbb{R}_+ \to \mathbb{R}$ denote an income tax function. Throughout this section, income tax functions are assumed to be piecewise-differentiable with directional derivatives less than one. An agent of type k receiving wage w_k solves the problem:

$$\sup_{\mathbb{R}_+ \times [0, \bar{e}]} U(c, e) \quad \text{s.t.} \quad c \leq w_k e - T(w_k e). \tag{3}$$

TECHNOLOGY A representative competitive firm hires agents of all types. The firm uses a production function $G : \mathbb{R}_+^K \to \mathbb{R}$ defined directly on the labor inputs of the different

[10] Much of the optimal tax literature is cast in terms of a continuum of types. This literature maintains the linear production function assumption. Although versions of the results that we give below are available for continuum (of type) economies, for general constant returns to scale production functions, their derivation requires stepping outside of the framework of optimal control and maximizing an infinite-dimensional Lagrangian directly. To avoid technical complications that do not generate additional economic insight we do not do this.

[11] If, given U, consumption is a normal good, then this condition is assured.

types. The firm solves:

$$\max_{\mathbb{R}_+^K} G(e_1\pi_1, \ldots, e_K\pi_K) - \sum_{k=1}^{K} w_k \pi_k e_k,$$

where e_k is the common effort level of agents of type k. G is assumed to be a continuously differentiable, linearly homogeneous function with $G(0) = 0$ and k-th partial derivative G_k.

TAX EQUILIBRIUM Let P be a fixed public spending amount. Given P, a *tax equilibrium* is an income tax function $T : \mathbb{R}_+ \to \mathbb{R}$, an allocation $\{c_k, e_k\}_{k=1}^K$ and a wage profile $\{w_k\}_{k=1}^K$ such that (i) for each $k = 1, \ldots, K$, (c_k, e_k) solves (3), (ii) for each $k = 1, \ldots, K$, $w_k = G_k(e_1\pi_1, \ldots, e_K\pi_K)$ and (iii) the goods clearing condition holds: $P + \sum_{k=1}^K c_k \leq G(e_1\pi_1, \ldots, e_K\pi_K)$. We keep P fixed throughout and vary T. Let $\mathscr{E}$ denote the set of tax equilibria (for the fixed spending amount P).

Remark 1. This definition restricts attention to equilibria in which all agents of a given type make the same effort and consumption choice given the tax schedule. In addition, we restrict attention to deterministic tax schedules. The literature has considered richer tax mechanisms that randomize over tax rates, see Stiglitz (1982) and Brito, Hamilton, Slutsky, and Stiglitz (1995). Recently, Hellwig (2007) has given sufficient conditions on preferences for deterministic tax mechanisms to be optimal. We assume these conditions hold in our setting. Finally, we make the usual assumption in public economics that the government cannot directly communicate with firms about a worker's type.

3.2 OPTIMAL POLICY

A government attaches Pareto weight g_k to type group k, $\sum_{k=1}^K g_k = 1$, and selects a tax equilibrium to solve:

$$\sup_{\mathscr{E}} \sum_{k=1}^{K} U(c_k, e_k) g_k. \tag{PP}$$

Let T^* and $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$ denote an optimal tax equilibrium. If $e_k^* > 0$, then we call:

$$\tau_k^* = -\frac{U_e(c_k^*, e_k^*)}{w_k^* U_c(c_k^*, e_k^*)} - 1$$

the marginal tax rate at income $q_k^* = w_k^* e_k^* > 0$.[12] An optimal tax equilibrium satisfies a no binding downward incentive constraint condition if:

$$U(c_k^*, e_k^*) > U(c_{k+1}^*, q_{k+1}^*/w_k^*). \tag{NDIC}$$

The following proposition extends results in Stiglitz (1982) to the K-type setting.

Proposition 1. *Let T^* and $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$ denote an optimal tax equilibrium with agent types indexed so that $w_k^* = G_k(e_1^*\pi_1, \ldots, e_K^*\pi_K)$ is increasing in k.*[13] *Assume that $\{c_k^*, e_k^*, w_k^*\}_{k=1}^K$ satisfies the no downward binding condition (NDIC). Optimal marginal tax rates satisfy:*

$$\frac{\tau_k^*}{1-\tau_k^*} = \frac{\Phi_k^*}{\pi_k} U_c(c_k^*, e_k^*) + \mathcal{H}_k^* \frac{\Delta w_{k+1}^*}{w_{k+1}^*} e_k^* \frac{1-\Pi_k}{\pi_k} \Psi_k^*, \tag{4}$$

where $\Delta w_{k+1} := w_{k+1} - w_k$, Φ_k^* is given by:

$$\Phi_k^* := -\sum_{j=1}^{K-1} \Psi_j^* (1-\Pi_j) \frac{U_e\left(c_j^*, \frac{q_j^*}{w_{j+1}^*}\right)}{U_e(c_k^*, e_k^*)} \left\{\phi_{k,j}^* - \phi_{k,j+1}^*\right\} \frac{e_j^* w_j^*}{e_k^* w_{j+1}^*}.$$

with $\phi_{k,j}^* = \frac{e_k^*}{w_j^*} \frac{\partial w_j}{\partial e_k}(e_1^*, \ldots, e_K^*)$,

$$\mathcal{H}_k^* = \left\{ -\frac{U_c(c_k^*, e_k^*) - U_c\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{\frac{\Delta w_{k+1}^*}{w_{k+1}^*} e_k^* U_c(c_k^*, e_k^*)} + \frac{U_e(c_k^*, e_k^*) - U_e\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{\frac{\Delta w_{k+1}^*}{w_{k+1}^*} e_k^* U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k+1}^*} + 1 \right\}$$

and

$$\Psi_k^* = \sum_{j=k}^{K} \left\{ 1 - \frac{g_j U_c(c_j^*, e_j^*)}{\pi_j \chi^*} \right\} \left(\frac{U_c(c_k^*, e_k^*)}{U_c(c_j^*, e_j^*)} \right) \prod_{i=k}^{j-1} \frac{U_c(c_i^*, q_j^*/w_{j+1}^*)}{U_c(c_i^*, e_i^*)} \frac{\pi_j}{1-\Pi_k},$$

with χ^* equal to the shadow price of resources (i.e. the multiplier on the goods market clearing constraint).

Proof. See Appendix C. □

[12] The optimal program (PP) only determines T^* at the points $\{q_k^*\}$. However, T^* may be chosen so that τ_k^* is the left derivative of T^* at q_k^*; if T^* is differentiable at q_k^*, then it is the derivative of T^*. The Inada conditions on agent utilities ensure that absent taxes agents would always choose to work. Consequently, in a tax equilibrium, a wedge between an agent's marginal rate of substitution and its wage is due to taxation and not to a boundary condition on effort.

[13] The ordering of agents types' marginal products is not given exogenously. The labeling of agent types can always be modified to reflect their marginal product ranking at the optimal allocation.

The second term in (4) is the discrete analogue of the terms found in continuous-type models with exogenous productivity (e.g Mirrlees (1971) and Saez (2001)). If preferences are quasi-linear in consumption, $U(c,e) = c + h(e)$, this term reduces to:

$$\left\{ \frac{h_e(e_k^*) - h_e\left(\frac{q_k^*}{w_{k+1}^*}\right)}{\frac{\Delta w_{k+1}^*}{w_{k+1}^*} e_k^* h_e(e_k^*)} \frac{w_k^*}{w_{k+1}^*} + 1 \right\} \frac{\Delta w_{k+1}^*}{w_{k+1}^*} e_k^* \frac{1 - \Pi_k}{\pi_k} \sum_{j=k}^{K} \left\{ 1 - \frac{g_j}{\pi_j} \right\} \frac{\pi_j}{1 - \Pi_k}. \tag{5}$$

For small variations in wages, the first component of (5) is approximately the compensated elasticity of labor supply, the second component is the rate of wage growth across types (this equals the reciprocal of the agent's productivity in models in which types are continuous and wages exogenous). The next component is the hazard function of the type distribution, while the final component captures the government's redistributive motive. Interpretations of these components in the continuous setting are given in Salanié (2011), while Saez (2001) relates the continuous version of the more general second term in (4) to labor supply elasticities, the type distribution and the government's redistributive motive.

The first term in (4), however, does not appear in standard optimal tax equations that are derived from models with linear production functions and exogenous wages. In settings with non-linear production functions, such as ours, the effort of the k-th agent type can affect the marginal rate of transformation and, hence, the ratio of wages between the j and $j+1$-th types. In doing so it can relax or tighten incentive constraints. To mimic the j-th productivity type, the $j+1$-th type must reproduce the j-th type's income. The effort cost of doing this depends on the relative wage ratio between the types: more compressed wage ratios relax incentive constraints and to the extent that the effort of a given type leads to such compression it should be encouraged through taxation. Stiglitz (1982) identifies this wage compression channel in a two type model. In that case there is only one binding incentive constraint and Φ_1^* and Φ_2^* reduce to:

$$\Phi_1^* := -\Psi_1^*(1-\pi_1) \frac{u_e\left(c_1^*, \frac{q_1^*}{w_2^*}\right)}{u_e(c_1^*, e_1^*)} \frac{e_1^* w_1^*}{e_1^* w_2^*} \frac{1}{\mathcal{E}_{1,2}^*}$$

$$\Phi_2^* := -\Psi_1^*(1-\pi_1) \frac{u_e\left(c_1^*, \frac{q_1^*}{w_2^*}\right)}{u_e(c_2^*, e_2^*)} \frac{e_1^* w_1^*}{e_2^* w_2^*} \frac{1}{\mathcal{E}_{1,2}^*},$$

where $\mathcal{E}_{1,2}^* = \frac{1}{\phi_{1,1}^* - \phi_{1,2}^*} = \frac{1}{\phi_{2,1}^* - \phi_{2,2}^*}$ is the elasticity of substitution between the two labor types (i.e. $\frac{w_2/w_1}{e_2/e_1} \frac{\partial e_2/e_1}{\partial w_2/w_1}$). Assuming this is positive, compression of wages between the two types, requires that the effort of the high (resp. low) type should be relatively encouraged (resp. discouraged). Since the second term in (4) is zero for $k = K = 2$, this translates into an optimal marginal income subsidy for high types and an enhanced marginal income tax for low types.[14] If G were *assumed* to be a CES production function, related results would obtain. In this case, marginal rates of transformation between the adjacent types are affected only by the efforts of these types. Consequently, the Φ_k^* terms would have the form:

$$\Phi_k^* := -\Psi_{k-1}^*(1-\Pi_{k-1})\frac{u_e\left(c_{k-1}^*, \frac{q_{k-1}^*}{w_k^*}\right)}{u_e(c_k^*, e_k^*)}\frac{e_{k-1}^* w_{k-1}^*}{e_k^* w_k^*}\frac{1}{\mathcal{E}} - \Psi_k^*(1-\Pi_k)\frac{u_e\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{u_e(c_k^*, e_k^*)}\frac{e_k^* w_k^*}{e_k^* w_{k+1}^*}\frac{1}{\mathcal{E}},$$

with $\mathcal{E}$ the constant elasticity of substitution. More generally, however, the relative wages of adjacent types could depend on other types and elasticities could be non-constant.[15]

Remark 2. Proposition 1 reports optimal tax formula in the absence of binding downward incentive constraints; in our later numerical calculations in Section 5 we verify that such constraints do not bind. Here we follow many contributors who focus on tax formulas in the absence of such binding constraints and verify their absence in numerical calculations, see, e.g. Saez (2001), Rothschild and Scheuer (2013). However, the proof of Proposition 1 indicates how the optimal tax formula would generalize if such constraints were to be binding. Binding downward incentive constraints give rise to bunching with agents who earn different wages working and consuming the same amount. Proposition 1 permits the opposite phenomenon of agents earning the same wage (but being of different talent types) and working and consuming different amounts. This may occur because the efforts of differently talented agents have the same marginal product, but different implications for relative wage elasticities.

4 Allocations with Assignment

We now consider optimal taxation in a framework with optimal task assignment. As noted in the introduction, assignment-based frameworks have been used in the positive

[14]The motive for this is to compress the relative wage. Moreover, this policy does not imply that high types have lower average taxes.

[15]Salanié (2011) makes a similar point. He argues that a CES in effort production function is a poor assumption, because the substitutability of similar agent types is most likely quiet different from that of dissimilar types. Consequently, he restricts his analysis to the two type case.

literature to formalize the impact of technical change on the wage and occupation distribution. As we show below, however, they imply an indirect production function over agent efforts. Consequently, they micro-found such production functions and relate their parameters to deeper, structural parameters that capture the impact of technical change on tasks.

4.1 Physical Environment

As before agents are partitioned across types $1, \ldots, K$ with a fraction π_k being of type k. We now explicitly label types as talents. In addition, there is a continuum of tasks $v \in [\underline{v}, \overline{v}]$ differentiated by complexity. A k-th talent agent faces a wage schedule $\omega_k : [\underline{v}, \overline{v}] \to \mathbb{R}_+$. The agent solves the problem:

$$\sup_{\mathbb{R}_+ \times [0,\overline{e}] \times [\underline{v},\overline{v}]} U(c, e) \quad \text{s.t.} \quad c \leq \omega_k(v)e - T(\omega_k(v)e). \tag{6}$$

Technology The productivity of a talent k in task v is given by $a_k(v) \in \mathbb{R}_+$. The productivity functions $\{a_k\}$, $a_k : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ satisfy the following condition.

Assumption 1. *The functions $a_k : [\underline{v}, \overline{v}] \to \mathbb{R}_+$, $k \in \{1, \ldots, K\}$, are continuously differentiable and satisfy for each $k \in \{1, \ldots, K-1\}$, $\frac{\partial \log a_{k+1}}{\partial v} \geq \frac{\partial \log a_k}{\partial v}$.*

The latter assumption implies that a is weakly log super-modular function of talent and task and that higher talents have a weak comparative advantage in more complex tasks. We sometimes strengthen this assumption by requiring that a is strictly log super-modular in task and talent: $k \in \{1, \ldots, K-1\}$, $\frac{\partial \log a_{k+1}}{\partial v} > \frac{\partial \log a_k}{\partial v}$. We also sometimes supplement Assumption 1 with the absolute advantage condition $a_{k+1} > a_k$. Task output is linear in effective labor input. Let Λ_k be the distribution of talent group k across tasks with (i) $\Lambda_k([\underline{v}, \overline{v}]) = \pi_k$ and (ii) density λ_k defined at almost every v. If members of talent group k exert effort e_k, then for almost every task v, task output is:

$$y(v) = \sum_{k=1}^{K} \lambda_k(v) a_k(v) e_k.$$

Final output Y is produced from task output $y : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ using a CES-technology:

$$Y = \begin{cases} A \left\{ \int_{\underline{v}}^{\overline{v}} b(v) y(v)^{\frac{\varepsilon-1}{\varepsilon}} dv \right\}^{\frac{\varepsilon}{\varepsilon-1}} & \varepsilon \in \mathbb{R}_+ \backslash \{1\}, \\ A \exp \left\{ \int_{\underline{v}}^{\overline{v}} b(v) \ln y(v) dv \right\} & \varepsilon = 1, \end{cases} \tag{7}$$

where $A > 0$, $b : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ is a continuous function satisfying $B(\overline{v}) = 1$, $B(v) := \int_{\underline{v}}^{v} b(v')dv'$.

Remark 3 (Interpreting a and b). The function a captures the idea that different workers may be more or less effective at performing specific tasks or using task-specific capital. Specifically, more complex tasks are more talent intensive and more talented agents have a comparative advantage in these tasks. The formulation of production here follows that in the assignment literature, e.g. Costinot and Vogel (2010), with the important addition of an intensive effort margin. Note that analogous to this literature, talent refers to relative ability in complex tasks. Unless the comparative advantage assumption is supplemented with an absolute advantage condition, "talented" agents need not have greater productivity in all tasks and need not earn higher wages in equilibrium.[16]

Later we allow for the possibility that a may change over time. That is we allow for Harrod-neutral technical progress at the level of a talent-task match and we allow this technical progress to depend upon both talent and task complexity. In particular to the extent that $\frac{\partial \log a}{\partial \theta}$ increases, the technical progress is talent-biased; to the extent that $\frac{\partial \log a}{\partial v}$ increases it is complexity-biased. In addition, if this technical change leads $\frac{\partial^2 \log a}{\partial \theta \partial v}$ to increase it is biased towards high talent-high complexity matches. In the latter case, it enhances the comparative advantage of talent in complex tasks and reduces the substitutability of a talent across tasks.

The function b weights task output in the final good aggregator. Variations in b may be interpreted as stemming from technological or preference-based variations in demand for different task outputs. We do not explicitly model capital. However, the model is easily extended in this direction, in which case the production functions in (7), under the assumption $B(\overline{v}) \in (0, 1)$, may be interpreted as indirect production functions for labor across tasks after the substitution of optimal capital. The parameter $b(v)$ is then interpreted as the sensitivity of final output with respect to the labor input in each task. It is influenced not only by variations in demand for different tasks, but also variations in the capital/labor intensity of tasks. Such variations are stressed by Acemoğlu and Autor (2011) who emphasize the automatization of middle complexity tasks. A further possibility is that b captures the extent to which workers purchase task output in the market or produce it at home. Shifts in b for some occupations may reflect the substitution of market for home production as in Buera and Kaboski (2012).

A representative firm hires workers of all talents to perform tasks and combines task

[16] The assignment literature refers to a worker's innate productive attribute as "skill". Since skills are endogenous, we prefer the word talent. Our model could be reinterpreted as one in which agents exert effort partly or wholly in acquiring skills rather than working.

output to produce final output.[17] The firm pays a wage $\omega_k(v)$ per unit of effort to an agent of talent k in task v and solves:

$$\max_{\lambda,e} A\left\{\int_{\underline{v}}^{\overline{v}} b(v)\left\{\sum_{k=1}^{K}\lambda_k(v)a_k(v)e_k\right\}^{\frac{\varepsilon-1}{\varepsilon}} dv\right\}^{\frac{\varepsilon}{\varepsilon-1}} - \int_{\underline{v}}^{\overline{v}}\sum_{k=1}^{K}\omega_k(v)\lambda_k(v)e_k dv. \tag{8}$$

Tax Equilibrium Let P be a fixed public spending amount. Given P, a *tax equilibrium* is an income tax function $T : \mathbb{R}_+ \to \mathbb{R}$, an allocation $\{c_k, e_k, \lambda_k\}_{k=1}^{K}$ and a wage profile $\{\omega_k\}_{k=1}^{K}$ such that (i) for each $k = 1, \ldots, K$, (c_k, e_k) and v in the support of Λ_k solves (6) at T and ω_k, (ii) $\{\lambda_k, e_k\}$ solves (8) at $\{\omega_k\}$, (iii) the final goods market clears:

$$P + \sum_{k=1}^{K} c_k \leq A\left\{\int_{\underline{v}}^{\overline{v}} b(v)\left\{\sum_{k=1}^{K}\lambda_k(v)a_k(v)e_k\right\}^{\frac{\varepsilon-1}{\varepsilon}} dv\right\}^{\frac{\varepsilon}{\varepsilon-1}}, \tag{9}$$

and (iv) the labor markets clear, for all $k = 1, \ldots, K$,

$$\pi_k = \int_{\underline{v}}^{\overline{v}} \lambda_k(v)dv. \tag{10}$$

Again, let $\mathscr{E}$ denote the set of tax equilibria. The government's problem, as before, is to select a tax equilibrium that maximizes a Paretian payoff:[18]

$$\sup_{\mathscr{E}} \sum_{k=1}^{K} U(c_k, e_k)g_k. \tag{PPA}$$

The following proposition characterizes tax equilibria.

Proposition 2. *Let $\{c_k, e_k, \lambda_k\}_{k=1}^{K}$ and a wage profile $\{\omega_k\}_{k=1}^{K}$ denote a tax equilibrium. Then*

[17] The organization of the production process could be further disaggregated into final goods producers who combine task output into final output and intermediate producers who use labor to produce a single task and sell their output on an intermediate tasks market. Alternatively, producers of tasks could sell directly to consumers with aggregation of task output an aspect of agent preferences. These alternatives are irrelevant for the results and questions we focus upon.

[18] As before, we constrain the set of mechanisms available to the government to ones that deterministically condition upon agent incomes. This assumption is standard in the literature and to a first approximation describes current tax codes. In our setting, it implies that the government cannot observe the task an agent performs or the amount of task output. The former may reasonably reflect the inherent difficulties in distinguishing between a worker's formal job description and the tasks that the worker actually performs. The latter assumption is relaxed in Ales, Kurnaz, and Sleet (2013), where taxation of task output is permitted. This introduces a motive for indirect taxation. We also limit attention to tax equilibria in which agents of a given type work the same amount and agents of a given type distribute themselves across tasks according to a density.

there is a tuple of threshold tasks $\{\tilde{v}_k\}_{k=1}^{K-1}$ *such that:*

$$\lambda_k(v) = \begin{cases} 0 & v < \tilde{v}_{k-1} \\ \frac{b(v)^\varepsilon a_k(v)^{\varepsilon-1}}{B_k(\tilde{v}_{k-1},\tilde{v}_k)^\varepsilon}\pi_k & v \in (\tilde{v}_{k-1}, \tilde{v}_k) \\ 0 & v > \tilde{v}_k, \end{cases}$$

where $B_k(\tilde{v}_{k-1}, \tilde{v}_k) := \left[\int_{\tilde{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv\right]^{\frac{1}{\varepsilon}}$. *All workers of talent k earn a common wage* $w_k = \omega_k(v)$, $v \in [\tilde{v}_{k-1}, \tilde{v}_k]$. *Relative wages are given by:*

$$\frac{w_{k+1}}{w_k} = \frac{a_{k+1}(\tilde{v}_k)}{a_k(\tilde{v}_k)} = \frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1}) \Big/ \{\pi_{k+1} e_{k+1}\}^{\frac{1}{\varepsilon}}}{B_k(\tilde{v}_{k-1}, \tilde{v}_k) \Big/ \{\pi_k e_k\}^{\frac{1}{\varepsilon}}}. \tag{11}$$

In particular, conditional on effort, the equilibrium allocation of talent to tasks maximizes output and is efficient.

Proposition 2 implies that, given effort, the assignment of talents to tasks is output-maximizing and, consequently, output is given by the following indirect production function over efforts:

$$G(\pi_1 e_1, \ldots, \pi_K e_K) = \sup \left\{ A \left\{ \sum_{k=1}^{K} B_k(\tilde{v}_{k-1}, \tilde{v}_k) \{e_k \pi_k\}^{\frac{\varepsilon-1}{\varepsilon}} \right\}^{\frac{\varepsilon}{\varepsilon-1}} \middle| \text{s.t. } \underline{v} \leq \tilde{v}_1 \ldots \leq \tilde{v}_{K-1} \leq \overline{v} \right\}. \tag{12}$$

Once G is determined, Proposition 1 continues to hold. Thus, the government's problem can be decomposed into two steps. In the inner step, G is determined in assignment problem (12); in the outer step a generalized Stiglitz tax problem as in Section 3 is solved using this production function. The inner assignment problem is similar to those considered in Rosen (1978), Teulings (1995) and the more recent papers of Costinot and Vogel (2010) and Acemoğlu and Autor (2011) (with the important difference that the effort and, hence, the labor allocation is determined in the outer step rather than being pinned down parametrically). Solving for G at an effort allocation $\{e_k\}_{k=1}^K$ requires obtaining an optimal sequence of task thresholds $\{\tilde{v}_k^*\}_{k=1}^K$ satisfying the discrete boundary value problem:

$$\frac{a_{k+1}(\tilde{v}_k^*)}{a_k(\tilde{v}_k^*)} = \frac{B_{k+1}(\tilde{v}_k^*, \tilde{v}_{k+1}^*) \Big/ \{\pi_{k+1} e_{k+1}\}^{\frac{1}{\varepsilon}}}{B_k(\tilde{v}_{k-1}^*, \tilde{v}_k^*) \Big/ \{\pi_k e_k\}^{\frac{1}{\varepsilon}}}, \tag{13}$$

with $\tilde{v}_1^* = \underline{v}$ and $\tilde{v}_{K-1}^* = \overline{v}$. In the assignment setting, the elasticity terms $\phi_{k,j} - \phi_{k,j+1}$ from the optimal tax formulas in Proposition 1 are given by:

$$\phi_{k,j} - \phi_{k,j+1} = -\frac{\partial \log(w_{j+1}/w_j)}{\partial \log e_k} = -\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j}\frac{\partial \log \tilde{v}_j}{\partial \log e_k}.$$

Thus, these elasticities depend upon the local comparative advantage of talent $j+1$ relative to talent j at the threshold $\tilde{v}_j$ and the sensitivity of the threshold to the effort of the k-th talent. The former depends entirely upon the functions $\{a_k\}$ at the relevant threshold. For example, if, as assumed later, $a_j(v) = D_j \exp \alpha j v$, then $\frac{\partial \log(a_{j+1}/a_j)}{\partial \log \tilde{v}_j} = \alpha \tilde{v}_j$. The sensitivity of the task thresholds to effort is much more complicated. The following lemma characterizes them.

Lemma 1. *The threshold sensitivities satisfy:*

$$\frac{\partial \log \tilde{v}_j}{\partial \log e_k} = (\delta_{j,k-1} - \delta_{j,k})\frac{1}{\varepsilon},$$

where:

$$\delta_{j,k} = \begin{cases} (-1)^{j+k}\prod_{l=j}^{k-1}\left(-\frac{\tilde{v}_l}{B_l}\frac{\partial B_l}{\partial \tilde{v}_l}\right) n_{j-1}m_{k+1}/n_{K-1} & 1 \le j \le k \le K-1 \\ (-1)^{j+k}\prod_{l=k}^{j-1}\left(\frac{\tilde{v}_l}{B_{l+1}}\frac{\partial B_{l+1}}{\partial \tilde{v}_l}\right) n_{k-1}m_{j+1}/n_{K-1} & K-1 \ge j > k \ge 1, \end{cases}$$

for $j = 1, \ldots, K-1$, $\delta_{j,K} = \delta_{j,0} = 0$, the n_i satisfy the recursion, $i = 2, \ldots, K-1$,

$$n_i = \left\{\frac{\tilde{v}_i}{a_{i+1}/a_i}\frac{\partial(a_{i+1}/a_i)}{\partial \tilde{v}_i} - \frac{\tilde{v}_i}{B_{i+1}}\frac{\partial B_{i+1}}{\partial \tilde{v}_i} + \frac{\tilde{v}_i}{B_i}\frac{\partial B_i}{\partial \tilde{v}_i}\right\} n_{i-1} + \left(\frac{\tilde{v}_i}{B_i}\frac{\partial B_i}{\partial \tilde{v}_i}\right)\left(\frac{\tilde{v}_{i-1}}{B_i}\frac{\partial B_i}{\partial \tilde{v}_{i-1}}\right) n_{i-2}$$

with $n_0 = 1$ and $n_1 = \frac{\tilde{v}_1}{a_2/a_1}\frac{\partial(a_2/a_1)}{\partial \tilde{v}_1} - \frac{\tilde{v}_1}{B_2}\frac{\partial B_2}{\partial \tilde{v}_1} + \frac{\tilde{v}_1}{B_1}\frac{\partial B_1}{\partial \tilde{v}_1}$ and the m_i satisfy the recursion, $i = K-2, \ldots, 1, :$

$$m_i = \left\{\frac{\tilde{v}_i}{a_{i+1}/a_i}\frac{\partial(a_{i+1}/a_i)}{\partial \tilde{v}_i} - \frac{\tilde{v}_i}{B_{i+1}}\frac{\partial B_{i+1}}{\partial \tilde{v}_i} + \frac{\tilde{v}_i}{B_i}\frac{\partial B_i}{\partial \tilde{v}_i}\right\} m_{i+1} + \left(\frac{\tilde{v}_{i+1}}{B_{i+1}}\frac{\partial B_{i+1}}{\partial \tilde{v}_{i+1}}\right)\left(\frac{\tilde{v}_i}{B_{i+1}}\frac{\partial B_{i+1}}{\partial \tilde{v}_i}\right) m_{i+2}.$$

with $m_{K-1} = \frac{\tilde{v}_{K-1}}{a_K/a_{K-1}}\frac{\partial(a_K/a_{K-1})}{\partial \tilde{v}_{K-1}} - \frac{\tilde{v}_{K-1}}{B_K}\frac{\partial B_K}{\partial \tilde{v}_{K-1}} + \frac{\tilde{v}_{K-1}}{B_{K-1}}\frac{\partial B_{K-1}}{\partial \tilde{v}_{K-1}}$ and $m_K = 1$.

To better understand these formula, consider the simplest case in which $K = 2$ and $\frac{\tilde{v}_1}{a_2/a_1}\frac{\partial(a_2/a_1)}{\partial \tilde{v}_1} = \alpha \tilde{v}_1$. In this case, there is only one threshold and its effort sensitivities are given by, for $k = 1, 2$,

$$\frac{\partial \tilde{v}_1}{\partial e_k} = \frac{(-1)^{k+1}}{\varepsilon}\frac{1}{\alpha \tilde{v}_1 + \frac{\tilde{v}_1}{B_1}\frac{\partial B_1}{\partial \tilde{v}_1} - \frac{\tilde{v}_1}{B_2}\frac{\partial B_2}{\partial \tilde{v}_1}}. \tag{14}$$

Consider a small increase in, say, the low talent's (talent 1's) effort. Other things equal, this raises output in the tasks in which low talents work, placing downward pressure on shadow prices for the output of these tasks and, hence, the low talent's wage. Low talents respond by populating more complex tasks. This moderates the effect of their effort increase on the shadow prices of the tasks that they previously worked in and on their wage. However, it does not fully offset the impact and as $\tilde{v}_1$ rises low talents move into tasks in which they have a comparative disadvantage relative to high talents. Thus, their relative wage falls. If their comparative advantage falls sharply with tasks (α is large), then the adjustment in $\tilde{v}_1$ will be small. In the limiting case of $\alpha = \infty$, $\frac{\partial \tilde{v}_1}{\partial e_1} = 0$ and the wage compression elasticity in (4) reaches its lowest value of $1/\varepsilon$: talents are as substitutable as the tasks into which they are locked. At the other extreme, $\alpha = 0$, no talent has a comparative advantage, the allocation of tasks is indeterminate and if low complexity tasks are assigned to low talent agents and high complexity tasks to high talents, while the task threshold is sensitive to agent efforts, relative wages are not: they are pinned down by absolute advantage.

For intermediate values of α, the demand for tasks and the b function are relevant for both task thresholds and relative wages. Suppose that b and, hence, demand for tasks in a neighborhood of $\tilde{v}_1$ is low, then a larger adjustment in $\tilde{v}_1$ is needed to absorb the impact of the effort rise. In terms of (14), $\partial B_1/\partial v_1$ and $-\partial B_2/\partial v_2$ will be close to zero, the right hand expression in the formula will be larger and the task threshold will be more sensitive to effort. Thus, threshold sensitivities and, hence, the wage compression elasticities from (4) depend on the structure of demand across tasks. In the more general many talent framework, an adjustment of one talent's effort causes a ripple effect as first local and then non-local task thresholds adjust, with the size of the adjustment depending in part on demand conditions in the neighborhood of the thresholds. Since relative wages between adjacent talents are determined by productivity ratios at thresholds (i.e. by $a_{j+1}(\tilde{v}_j)/a_j(\tilde{v}_j)$), an effort change by talent k workers can affect relative wages across the whole spectrum of talents and be a motive for encouraging or discouraging that effort's talent.

A REMARK ON ROTHSCHILD AND SCHEUER (2013). This section assumes K talents and a continuum of tasks. Following the assignment literature, we also impose some structure on a given talent's "productivity path" across tasks, it is given by the function $a_k : [\underline{v}, \overline{v}] \to \mathbb{R}_+$, and hence, the distribution of possible productivity profiles. This allows us to relate our analysis to recent positive work on technical change. In Appendix D, we extend this approach to models with a continuum of talents and tasks. In contrast, Rothschild and Scheuer (2013) suppose a continuum of talents and only two tasks. They directly define an agent's type to be its productivity profile, in this case just a pair of productivities $a = \{a_v\}_{v=1,2} \in \mathbb{R}^2_+$, but within this setup, allow for any (well behaved) distribution of productivity profiles. Their model also implies an indirect production function over labor inputs:

$$G(l) = \max_{I:\mathbb{R}^2_+ \to \{0,1\}} H\left(\int a_1 l(a) I(a) da, \int a_2 l(a)\{1 - I(a)\} da\right), \tag{15}$$

where $H : \mathbb{R}^2_+ \to \mathbb{R}_+$ is a constant returns final output production function and, respectively, $l(a) = e(a) f(a)$, $e(a)$ and $f(a)$ are the labor input, effort and density of the a-th productivity type. Rothschild and Scheuer (2013) contrast their results with a model in which the productivity distribution is degenerate at two points, $(a_1, 0)$ and $(0, a_2)$ with $a_k > 0$, and no reallocation of types across tasks is possible. They interpret the latter model as a version of Stiglitz (1982). However, while Stiglitz (1982) is silent on the possibility of labor reallocation, it is not inconsistent with it either. As noted previously, Stiglitz writes down a production function over labor inputs directly. Like us, Rothschild and Scheuer (2013), implicitly micro-found this production function (and extend it to many talents). Stiglitz's analysis is applicable to the indirect production function derived by Rothschild and Scheuer (2013).

4.2 Technical Change in the Two Talent Model

In this section we use a simple two talent-many task assignment framework to highlight implications of technical change for tax policy.

Physical Environment We place some additional restrictions on the environment of the preceding sections. The number of talents is restricted to $K = 2$ which, for concreteness, we label low (L) and high (H). Low and high talents have productivities $a_L(v)$ and $a_H(v)$ in task v, where:

$$a_L(v) = \alpha_L \left(1 - \frac{1}{1 + \exp(-\alpha_0(v - \hat{v}))}\right),$$

and

$$a_H(v) = \alpha_H \left(\frac{1}{1 + \exp(-\alpha_0(v - \hat{v}))}\right) = \frac{\alpha_H}{\alpha_L} \exp(\alpha_0(v - \hat{v})) a_L(v),$$

with $\alpha_H > \alpha_L$, $\alpha_0 \in \mathbb{R}_+ \cup \{+\infty\}$ and $\hat{v} \in [\underline{v}, \overline{v}]$. Consequently, $a_j(v)$ is weakly log supermodular in talent and task: $\frac{\partial \log a_H}{\partial v}(v) - \frac{\partial \log a_L}{\partial v}(v) = \alpha_0 \geq 0$. In particular, if $\alpha_0 > 0$, then high talents have a comparative advantage in more complex tasks. The limiting cases $\alpha_0 = 0$ and $\alpha_0 = \infty$ are important benchmarks. In the first case, high talents have an absolute advantage in all tasks, but no comparative advantage for more complex tasks: $\frac{a_H}{a_L} = \frac{\alpha_H}{\alpha_L}$. Thus, each type is perfectly substitutable across tasks. In the second case,

$$a_L(v) = \begin{cases} \alpha_L & \text{if } v < \hat{v} \\ 0 & \text{if } v > \hat{v} \end{cases} \quad \text{and} \quad a_H(v) = \begin{cases} 0 & \text{if } v < \hat{v} \\ \alpha_H & \text{if } v > \hat{v}, \end{cases}$$

and there is an extreme lack of substitutability of a given talent type across tasks either side of $\hat{v}$. In all other respects, technologies are as in the preceding section. Preferences are restricted to be of the form:

$$U(c, e) = u(c) - \frac{e^{1+\gamma}}{1+\gamma},$$

where u is twice continuously differentiable, increasing and concave and $\gamma > 0$. The government's objective is Paretian with weights g_j, $j \in \{L, H\}$.

GOVERNMENT'S FULL INFORMATION PROBLEM We first consider the implications of technological change for full information optima. This provides a useful benchmark for our subsequent analysis of optimal tax equilibria (in which taxes condition only upon incomes). Full information optima can be attained in a decentralized economy with lump sum taxes that condition directly upon talent. However, we analyze them directly as the outcomes of a full information planning problem in which the social objective $\sum_{k=1}^{K} U(c_k, e_k) g_k$ is maximized subject to the final good and labor market clearing conditions. The following lemma is the analogue of Proposition 2 for this setting.

Lemma 2. *Let $\{c_j^F, e_j^F, \lambda_j^F\}$, $j \in \{L, H\}$ denote a solution to the planner's problem with full information. There is a threshold task $\tilde{v}^F \in (\underline{v}, \overline{v})$ such that:*

$$\lambda_L^F(v) = \begin{cases} \frac{b(v)^\varepsilon a_L(v)^{\varepsilon-1}}{B_L(\tilde{v}^F)^\varepsilon} \pi_L & \textit{if } v < \tilde{v}^F \\ 0 & \textit{if } v > \tilde{v}^F \end{cases} \quad \textit{and} \quad \lambda_H^F(v) = \begin{cases} 0 & \textit{if } v < \tilde{v}^F \\ \frac{b(v)^\varepsilon a_H(v)^{\varepsilon-1}}{B_H(\tilde{v}^F)^\varepsilon} \pi_H & \textit{if } v > \tilde{v}^F, \end{cases} \tag{16}$$

where $B_L(\tilde{v}) := \left[\int_{\underline{v}}^{\tilde{v}} b(v)^{\varepsilon} a_L(v)^{\varepsilon-1} dv\right]^{\frac{1}{\varepsilon}}$, $B_H(\tilde{v}) := \left[\int_{\tilde{v}}^{\overline{v}} b(v)^{\varepsilon} a_H(v)^{\varepsilon-1} dv\right]^{\frac{1}{\varepsilon}}$ *and* $\tilde{v}^F$ *satisfies:*

$$\frac{w_H}{w_L} = \frac{a_H(\tilde{v}^F)}{a_L(\tilde{v}^F)} = \left(\frac{B_H(\tilde{v}^F)}{B_L(\tilde{v}^F)}\right)^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} \left(\frac{g_H}{g_L}\right)^{\frac{1}{1+\gamma\varepsilon}} \left(\frac{\pi_L}{\pi_H}\right)^{\frac{1+\gamma}{1+\gamma\varepsilon}}. \tag{17}$$

Proof. See Appendix C. □

Relative to Proposition 2, Lemma 2 explicitly solves for the optimal effort levels and substitutes them into (17), which then determines the task threshold and, in the decentralization, the relative wage between high and low talents (the talent premium) as a function of technological and individual and societal preferences parameters. We assume in the remainder of this section that this premium is greater than one, i.e. that the parameters a_L, a_H, ε and γ satisfy:[19]

$$1 < \frac{a_H(\tilde{v}^F)}{a_L(\tilde{v}^F)} = \left(\frac{B_H(\tilde{v}^F)}{B_L(\tilde{v}^F)}\right)^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} \left(\frac{g_H}{g_L}\right)^{\frac{1}{1+\gamma\varepsilon}} \left(\frac{\pi_L}{\pi_H}\right)^{\frac{1+\gamma}{1+\gamma\varepsilon}} \tag{18}$$

Figure 4 illustrates the determination of $\tilde{v}^F$ and, hence, the talent premium for particular configurations of parameters. The red curve gives the function $M\left(B_H(\tilde{v})/B_L(\tilde{v})\right)^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}}$, $M := \left(g_H/g_L\right)^{\frac{1}{1+\gamma\varepsilon}}\left(\pi_L/\pi_H\right)^{\frac{1+\gamma}{1+\gamma\varepsilon}}$. The blue curves give the function $a_H(\tilde{v})/a_L(\tilde{v})$ for two values of the comparative advantage parameter α_0: $\underline{\alpha}_0 = 0$ and $\overline{\alpha}_0 > 0$. $\tilde{v}^F$ is determined at the intersection of the red and blue curves (see eq. (17)). The function $a_H(\tilde{v})/a_L(\tilde{v})$ passes through the point $(\hat{v}, \alpha_H/\alpha_L)$ regardless of the value of α_0. Increases in α_0 cause the graph of this function to pivot around $(\hat{v}, \frac{\alpha_H}{\alpha_L})$; it is horizontal for $\alpha_0 = 0$ and upward sloping for $\alpha_0 > 0$. Consequently, increases in the comparative advantage parameter α_0 move the intersection of the functions $a_H(\tilde{v})/a_L(\tilde{v})$ and $M\{B_H(\tilde{v})/B_L(\tilde{v})\}^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}}$ and, hence, the threshold complexity $\tilde{v}^F$ towards $\hat{v}$. In the figure, $M\{B_H(\hat{v})/B_L(\tilde{v})\}^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} > \alpha_H/\alpha_L$ and the intersections of the $M\{B_H(\hat{v})/B_L(\tilde{v})\}^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}}$ and $\frac{\alpha_H}{\alpha_L}$ curves lie to the right of $\hat{v}$. In this case, increases in α_0 imply decreases in the threshold complexity $\tilde{v}^F$. Intuitively, this is a case in which demand (for labor) is biased towards more complex tasks ($v > \hat{v}$). This demand can be met by moving low talents into such tasks or inducing high talents to work harder. At larger α_0 values, the relative productivity of low talents falls off more rapidly with increases in complexity. Consequently, greater levels of high talent effort, rather than low talent participation are relied upon to meet demand in high complex tasks. For config-

[19]Since absolute advantage of high talents in all tasks is not assumed, the model permits a talent premium less than one if demand is biased towards simple tasks and low talents have an absolute advantage in these tasks.

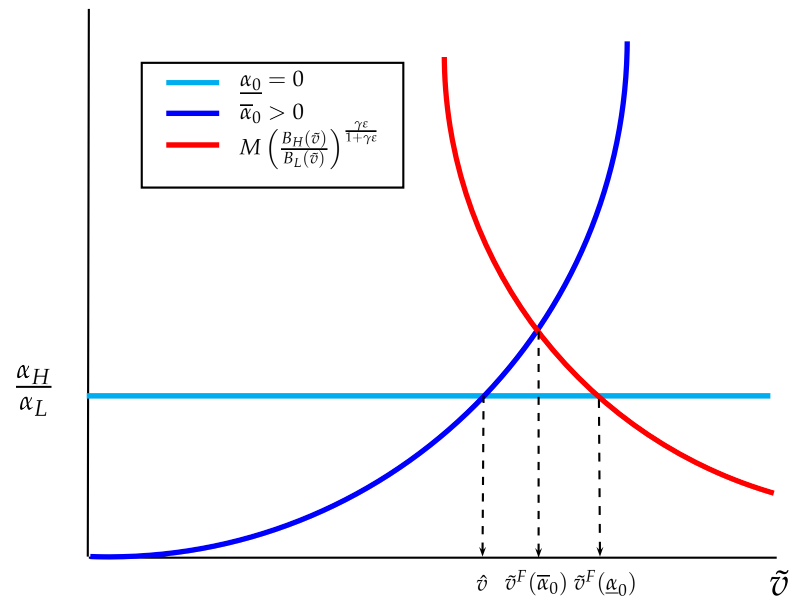


Figure 4: Determination of $\tilde{v}^F$ in the Full Information Planning Problem.

urations of parameters satisfying $M\{B_H(\hat{v})/B_L(\tilde{v})\}^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} > \alpha_H/\alpha_L$, increases in α_0 raise $a_H(\tilde{v}^F)/a_L(\tilde{v}^F)$ (see Figure 4) and, hence, the talent premium. A ("complexity-biased") shift in the b function that everywhere lowers B, leads to an unambiguous increase in $\tilde{v}^F$, the talent premium.

TAX EQUILIBRIA We now consider the implications of technical change for optimal tax equilibria. In the current setting, the optimal marginal tax equations (4) reduce to:

$$\frac{\tau_L^*}{1-\tau_L^*} = \left(\frac{E[1/u_c(c^*)]}{1/u_c(c_L^*)}\frac{g_L}{\pi_L} - 1\right)\left\{1 - \left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma}\left\{1 - \frac{1}{\mathcal{E}(e_H^*/e_L^*)}\right\}\right\} \geq 0 \tag{19}$$

and

$$\frac{\tau_H^*}{1-\tau_H^*} = -\left(1 - \frac{E[1/u_c(c^*)]}{1/u_c(c_H^*)}\frac{g_H}{\pi_H}\right)\left(\frac{q_L^*}{q_H^*}\right)^{1+\gamma}\frac{1}{\mathcal{E}(e_H^*/e_L^*)} \leq 0. \tag{20}$$

with the elasticity of substitution given by:

$$\mathcal{E}(e_H/e_L) := -\frac{\partial \log \frac{e_H}{e_L}}{\partial \log \frac{w_H}{w_L}} = \varepsilon + \frac{1}{\alpha_0}\left[\left(\frac{b_H(\tilde{v}(e_H/e_L))}{B_H(\tilde{v}(e_H/e_L))}\right)^{\varepsilon} + \left(\frac{b_L(\tilde{v}(e_H/e_L))}{B_L(\tilde{v}(e_H/e_L))}\right)^{\varepsilon}\right] \geq \varepsilon. \tag{21}$$

To highlight some implications of technical change for policy we proceed in two steps. First, we impose quasi-linear preferences and analytically derive implications of shifts in the B function; then we numerically calculate the implications of shifts in B and α_0 for non-quasi linear preferences.

TECHNICAL CHANGE UNDER QUASI-LINEAR PREFERENCES Let $U(c,e) = c - e^{1+\gamma}/(1+\gamma)$. Quasi-linearity allows us to pin down the incentive multiplier as a function of parameters and determine the first (redistributive motive) terms in (40) and (41) as $g_L/\pi_L - 1$ and $1 - g_H/\pi_H$ respectively. Implications of complexity-biased shifts in demand for policy are described in Proposition 3.

Proposition 3. *Assume quasi-linear preferences. If $\alpha_0 \in (0,\infty)$, $\varepsilon \geq 1$, $\mathcal{E} \geq 1$ and (locally) constant or if $\alpha_0 = \infty$ and $\varepsilon \geq 1$ then (small) increases in $B_H(v)/B_L(v)$ everywhere raise the talent premium. If $\mathcal{E} > 1$ such increases lead to increases in optimal marginal taxes on low talents. If $\alpha_0 = 0$, then changes in B_H/B_L have no impact on the talent premium or optimal taxes.*

Proof. See Appendix C. □

It follows from Proposition 3 that if tasks are gross substitutes, the elasticity of substitution of talents across tasks locally constant and talents are not perfect substitutes, then complexity-biased shifts in demand raise both the talent premium and optimal marginal taxes on low talents.

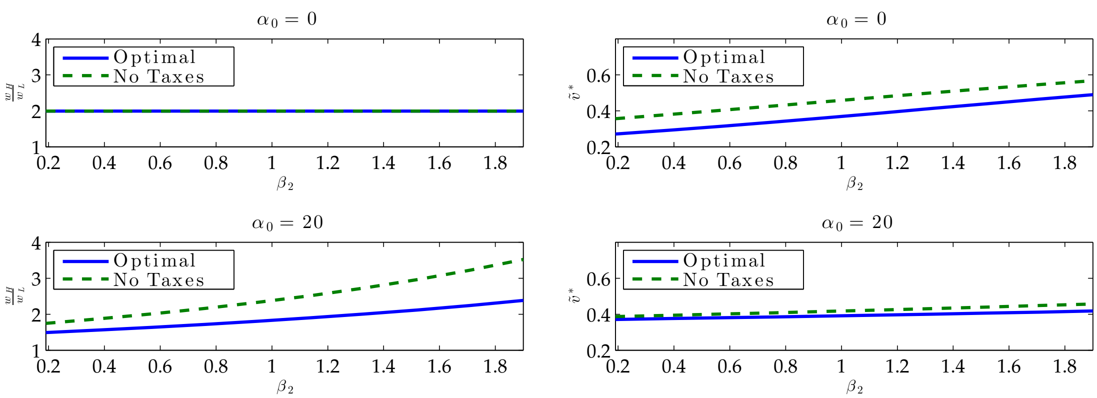
The impact of changes in the comparative advantage of high talents in complex tasks on optimal taxes is more subtle. The direct impact of a rise in α_0 is to reduce the elasticity of substitution $\mathcal{E}(e_H/e_L)$ at all effort ratios. This is a force for higher marginal taxes on low talents. However, there may be an offsetting adjustment in the task threshold $\tilde{v}^*$. Moreover, our previous analysis of the first best case implies that, absent incentive considerations, increases in α_0 can raise or lower the talent premium depending on the degree of complexity-bias in demand. We present a numerical example.[20]

A NUMERICAL EXAMPLE. Figures 5 to 7 illustrate some of the effects discussed above. They show optimal talent premia, threshold tasks, taxes and components of tax formulas for economies in which $u = \log$, $\gamma = 1$, $\underline{v} = 0$, $\overline{v} = 1$, $\hat{v} = 0.5$, $A = \varepsilon = 1$, $\alpha_0 = 0$ or 20 and:

$$b(v) = \beta_1 + \beta_2 v,$$

[20] One case in which more definitive answers are available is the comparison of the $\alpha_0 = 0$ and $\alpha_0 = \infty$ cases. Here the talent premium is greater if $\alpha_0 = \infty$ provided demand is sufficiently complexity-biased (a sufficient condition is given in the Appendix C) and the marginal taxes on low talents rises.

with $\beta_1 = -\frac{\beta_2^2}{2}$ so that $B(\overline{v}) = 1$. Increases in β_2 imply reductions in B everywhere and correspond to complexity-biased shifts in demand. For comparison talent premia and thresholds are included for economies with zero taxes.



(a) Talent Premia.

(b) Threshold Tasks.

Figure 5: Optimal talent premia and threshold tasks for economies with low ($\alpha_0 = 0$) and high ($\alpha_0 = 20$) values of α and various values of the parameter b_2 in $b(v) = \beta_1 + \beta_2 v$. Talent premia and threshold tasks for these parameter values in a market equilibria with no taxes are also shown.

Figure 5 shows that when $\alpha_0 = 0$ and talents are highly substitutable complexity-biased shifts in demand induce large adjustments in the allocation of talents across tasks (the threshold task adjusts considerably in Figure 5b), but no adjustment in the talent premium (Figure 5a). When talents are much less substitutable across tasks ($\alpha_0 = 20$) these effects are reversed. In the latter case, they, and particularly the impact on the talent premium, are much more muted under optimal than under no taxation, a reflection of the role of optimal taxes in compressing the talent premium. The smaller adjustment of the threshold task in this case reflects the corresponding stimulus to high talent effort and discouragement of low talent effort under the optimal arrangement.

Figure 6 shows the implications of changes in the technological parameters for optimal marginal taxes. When talents are perfectly substitutable, optimal marginal taxes are 21% on low talents, 0% on high talents and independent of the pattern of task demand. In contrast, when $\alpha_0 = 20$ and talents are less substitutable, optimal marginal tax rates depend on b_2. As b_2 rises and the demand for complex tasks increases, marginal tax rates on low talents rise from 17 to 28%, while marginal subsidies on high talents fall from nearly 4% to close to 0. Inspection of Figures 5b and 7 show that in the former case this rise is mainly associated with a rise in the talent premium and the redistributive motive, with some modest offset from the endogenous increase in the elasticity of substitution. In the latter case, the fall in the subsidy is mainly associated with a fall in the relative income of low talents.

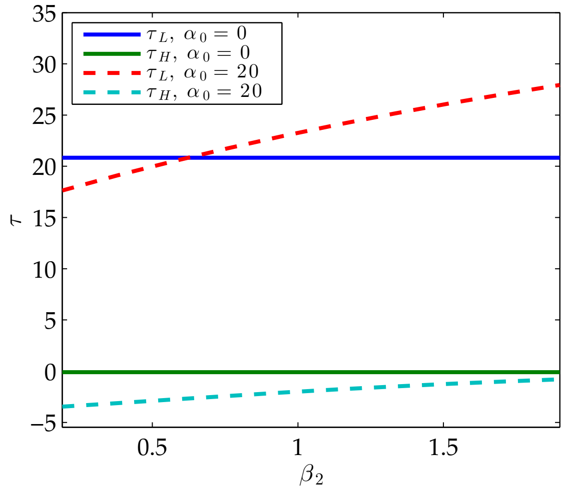


Figure 6: Optimal marginal tax rates on low and high talents for different configurations of α_0 and b_2.

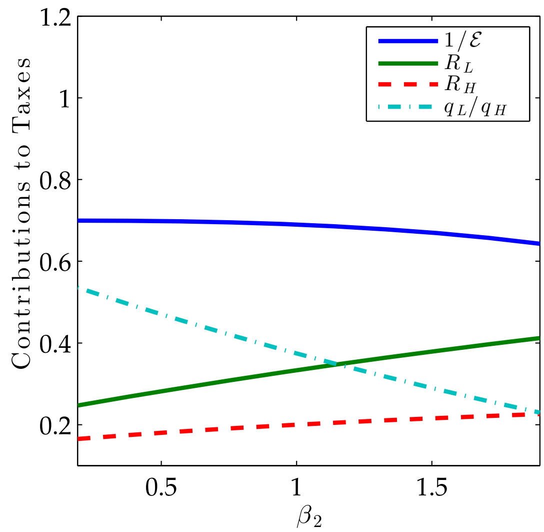


Figure 7: Components of marginal tax formulas (40) and (41) for $\alpha_0 = 20$ and different values of b_2. R_k, $k = L, H$, give the redistributive components of the formulas: $R_L = \frac{E[1/u_c(c^*)]}{1/u_c(c_L^*)} \frac{g_L}{\pi_L} - 1$ and $R_H = 1 - \frac{E[1/u_c(c^*)]}{1/u_c(c_H^*)} \frac{g_H}{\pi_H}$.

SUMMARY Technical change that increases the talent wage premium and reduces the substitutability of talents is associated with higher optimal marginal taxes on low talents. Change that increases the talent income premium and the substitutability of talents is associated with lower marginal subsidies on high talents. However, in general, the technical parameters α_0 and b influence both talent premia and talent substitutability often in, from the perspective of taxation, offsetting ways. This complicates the task of unravelling the implications of technical change for optimal policy. The analysis is still more complicated in settings with multiple talents. Such settings seem essential, however, for exploring the implications of recently documented changes in the patterns of wage premia across talents and skills. In particular, recent analysis in labor economics has emphasized the non-monotonic implications of technical change for the pattern of wage premia across talents and skills: Autor and Dorn (2013) analyze the compression of mid-to-low skill wage premia and the expansion of high-to-mid skill wage premia. To capture the policy implications of such facts a model with more than two talents is required.

5 Quantitative Analysis

We now describe how we bring the model of Section 4 to the data. Our main data source is the Current Population Survey (CPS). Further details of our use and treatment of the data are given in Appendix E. We proceed as if the data were generated by a (sub-optimal) tax equilibrium and use parametric assumptions and equilibrium restrictions from our model to identify and estimate the key technological parameters a and b in the 1970s and 2000s. In the subsequent section, we calculate optimal tax equilibria at these estimated parameters.

5.1 Determining Types and Tasks

Mapping empirical occupations to ordered sets of tasks The CPS categorizes workers into $M = 304$ distinct occupations; it also provides information on worker earnings and hours worked from which a measure of wages can be imputed. Our model involves an interval of tasks ordered by complexity. To relate empirical occupations to modeled tasks, we utilize the model's implication that equilibrium wages are rising in task complexity. We normalize the task space to $[\underline{v}, \overline{v}] = [1, M]$ and sub-divide this interval into M sub-intervals of unit length: $\mathcal{V}_m = [v_{m-1}, v_m]$. We calculate the imputed average wage in each occupation using 1970's data and rank occupations according to this wage. The m-th ranked occupation is then mapped to the m-th subinterval $\mathcal{V}_m$.[21]

Recovering the empirical assignment function $\tilde{v}$ The model in Section 4 featured a finite number of talents; this facilitated the derivation of analytical results. However, for the remainder of the paper we find it convenient to treat agent talent symmet-

[21] This approach is essentially that taken by Acemoğlu and Autor (2011).

rically with task complexity and to assume that agents are distributed uniformly across an interval of talents, $\theta \in \Theta := [\underline{\theta}, \overline{\theta}]$.[22] Thus, an agent's talent should now be an interpreted as an index (and a rank), the implications of which for productivity are captured by a function: $a : [\underline{\theta}, \overline{\theta}] \times [\underline{v}, \overline{v}] \to \mathbb{R}_+$. In particular, we shall restrict a so that higher talent-indexed agents have a comparative advantage in higher complexity-indexed tasks (occupations).[23] The set Θ is normalized to $[0, 1]$.

The continuous analogue of the task thresholds $\{\tilde{v}_k\}$ is a task assignment function: $\tilde{v} : [\underline{\theta}, \underline{\theta}] \to [\underline{v}, \overline{v}]$. This function is strictly increasing in our model; denote its inverse by $\tilde{\theta}$. Under the assumption that agents are distributed uniformly across talent indices, $\tilde{\theta}$ is the distribution of workers across tasks. Consequently, we treat the distribution of workers across ordered occupations as the empirical counterpart of $\tilde{\theta}$ and $\tilde{v}$ to be the inverse of this.

5.2 Estimating b

To facilitate identification of the a and b functions, we restrict the elasticity of substitution between task outputs to be one. Then $b(v)$ is simply the elasticity of final output with respect to task v output and, hence, the share of total compensation paid to workers in task v. Consequently, estimates of b may be be calculated from compensation shares independently of knowledge of the a's.

Precisely, under the Cobb-Douglas restriction, the firm's first order conditions from the continuous-talent version of (8), imply for almost all (θ, v):

$$\omega(\theta, v) = Y \frac{a(\theta, v) b(v)}{y(v)}. \tag{22}$$

In the continuous talent setting, task output is given by $y(v) = a(\tilde{\theta}(v), v) e(\tilde{\theta}(v)) \tilde{\theta}_v(v)$, with $\tilde{\theta}_v$ the derivative of $\tilde{\theta}$. Combining this with (22) and integrating over $\mathcal{V}_m$ gives total labor income in occupation m in terms of the b-function:

$$\int_{\mathcal{V}_m} \omega(\tilde{\theta}(v), v) e(\tilde{\theta}(v)) \tilde{\theta}_v(v) dv = Y \int_{\mathcal{V}_m} b(v) dv.$$

[22] The convenience is two fold. First, since occupational (task) data is discrete, assuming a continuous set of talents avoids having to deal with talent groups that are distributed across adjacent occupations. Second, it allows us to apply numerical optimal control methods to solve the problem.

[23] Note that although the distribution over the (ordinal) talent index is uniform, the distribution over (cardinal) productivities is not. It is induced endogenously by a and by the assignment of talent to tasks. It is, however, natural to ask how this talent index relates to other observables such as education and experience in the data. We discuss this briefly in Appendix G.

Average income in occupation m is then obtained by dividing both sides by the mass of workers in the occupation, S_m:

$$i_m \equiv \frac{1}{S_m} \int_{\mathcal{V}_m} w(\tilde{\theta}(v), v) e(\tilde{\theta}(v), v) \tilde{\theta}_v(v) dv = \frac{Y}{S_m} \int_{\mathcal{V}_m} b(v) dv.$$

Consequently, the average value of b in occupation m, $b_m := \int_{v_{m-1}}^{v_m} b(v)dv$, may be measured as:

$$b_m = \frac{S_m i_m}{Y}, \qquad \forall\, m = 1, \ldots, M. \tag{23}$$

We identify Y with average per capita labor income per decade.[24] A smooth estimate of the b-function is obtained by fitting a LOWESS model to $\{v_m, \log b_m\}$ data.[25] Figure 8 displays estimates of b for the 1970's and the 2000's. The figure shows that b rises for low

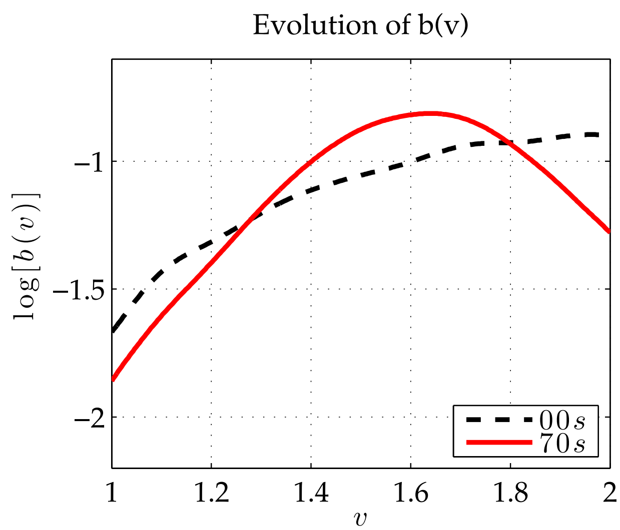


Figure 8: Evolution of $\log(b(v))$ across decades.

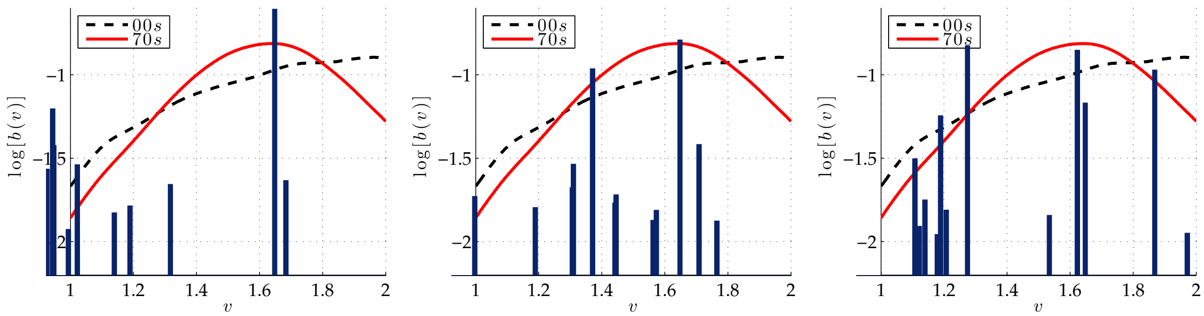
and high v-occupations, but falls for intermediate ones. The picture is consistent with the phenomenon of *job polarization* as discussed in Section 2. This polarization feature is robust to different sample selection assumptions, see Appendix F for further details.

Figure 9 sharpens intuition concerning the relation of different v's to the data by showing the location of different sectoral occupations in the space of v's. The figure overlays the values of $b(v)$ with a bar graph displaying the employment shares of occupations belonging to particular sectors. Figure 9a does this for services, Figure 9b for manufac-

[24]In 2005 dollars we have $Y_{70s} = \$\,34,206$ and $Y_{00s} = \$\,39,204$. M is 304.

[25]LOWESS is an acronym for locally weighted scatterplot smoothing. The method builds up a smooth curve through a set of date points by fitting simple linear or quadratic models to localized subsets of data. We use a smoothing parameter of 0.4.

turing and Figure 9c for finance insurance and real estate.[26] It emerges that the service sector is associated mostly with "low" v occupations (the bar on the right in Figure 9a refers to managers and administrative support). While manufacturing is mostly middle v occupations (although with a wider range), finance insurance and real estate is comprised mostly of "high" v occupations.



(a) Services. (b) Manufacturing. (c) Finance/Ins./Real E.

Figure 9: Smoothed values for $\log(b(v))$ over time (left axis). Share of occupations over v (right axis).

5.3 Estimating a

It is convenient to re-write the productivity function as $a(\theta, v) = e^{\alpha(\theta,v)}$. In our benchmark quantitative environment we set:[27]

$$\alpha(\theta, v) = \alpha_1\theta + \alpha_2\theta \cdot v. \tag{24}$$

Our goal will be to determine a structure of a by estimating the parameters $\{\alpha_i\}_{i=1}^{2}$. Recall that the degree of comparative advantage is given by:

$$\frac{\partial^2 \ln a(\theta, v)}{\partial\theta\partial v} = \alpha_2.$$

Let $w(\theta) = \omega(\theta, \tilde{v}(\theta))$ denote the wage received by talent θ given the (equilibrium) assignment function $\tilde{v}$. Combining (24) with the continuous talent analogue of the efficient

[26] Not shown are occupations that constitute less than 2% of the workforce of each sector.

[27] Note that under our Cobb-Douglas assumption multiplicative terms depending only on v are absorbed into TFP. Thus, if $a(\theta, v) = D(v)\exp^{\alpha(\theta,v)}$ is assumed in place of α, then the only change is rescaling of TFP by $\exp^{\int b(v)\ln D(v)dv}$.

assignment condition (11) gives:

$$\frac{d\ln w(\tilde{\theta}(v))}{d\theta} = \frac{\partial \ln \omega(\tilde{\theta}(v), v)}{\partial\theta} = \frac{\partial \ln a(\tilde{\theta}(v), v)}{\partial\theta} = \alpha_1 + \alpha_2 \cdot v. \tag{25}$$

We construct an empirical counterpart for $\frac{d\ln w(\tilde{\theta}(v))}{d\theta}$ in three steps. First, we use information on weeks and usual hours worked in the previous year and self reported yearly labor income in the CPS to impute workers' average hourly wages. Second, averaging over m and over decade we construct empirical counterparts of $w(\tilde{\theta}(v))$. Third, we utilize this, our previously constructed estimate of $\tilde{\theta}$ and a LOWESS smoother to obtain a smooth approximation to w. From this $\frac{d\ln w(\tilde{\theta}(v))}{d\theta}$ is obtained and, hence, estimates for α_1 and α_2; these are reported in Table 2.

Table 2: Estimation of α_1 and α_2 from Equation (25). Standard Errors in Parenthesis.

Decade	Parameters	
	α_1	α_2
70s	0.59	0.79
	(0.27)	(0.18)
00s	-3.00	2.93
	(0.15)	(0.10)

To complete the parametrization of the production function it remains only to determine TFP, A. This is obtained by dividing aggregate per capita income by the approximation to $\exp\{\int_{\underline{v}}^{\overline{v}} b(v) \ln\{a(\tilde{\theta}(v), v)e(\tilde{\theta})\theta_v(v)\}dv\}$:

$$A = \frac{Y}{\exp\{\sum_{m=1}^{M} b_m \ln\{e^{(\alpha_1+\alpha_2 v_m)\tilde{\theta}_m} e_m S_m\}\}},$$

where e_m are usual hours worked by workers in occupation m and α_1 and α_2 are our previously estimated productivity parameters.

5.4 Parameters Determined Externally

We assume the following parametrization for preferences:

$$u(c) + h(e) = \frac{c^{1-\sigma}}{1-\sigma} - \frac{e^{1+\gamma}}{1+\gamma}.$$

We set $\sigma = 2$ and $\gamma = 2$. Note that the choice of σ and γ has no impact in the estimation of $b(v)$ and $a(\theta, v)$.

6 (Preliminary) Results

In this section we discuss our quantitative results. We begin by displaying the optimal income tax codes for the 1970s and the 2000s. We then compare these to actual taxes in the US over the corresponding decades. Next, we compare the optimal distributions of workers across wages and occupations with their empirical counterparts. Finally, we use optimal tax formulas and counterfactuals to decompose the impact of technical change on taxes.

6.1 Optimal Taxation

We begin by looking at the optimal income tax codes for both the "70s" and the "00s". Figure 10 displays optimal average rates over these decades.

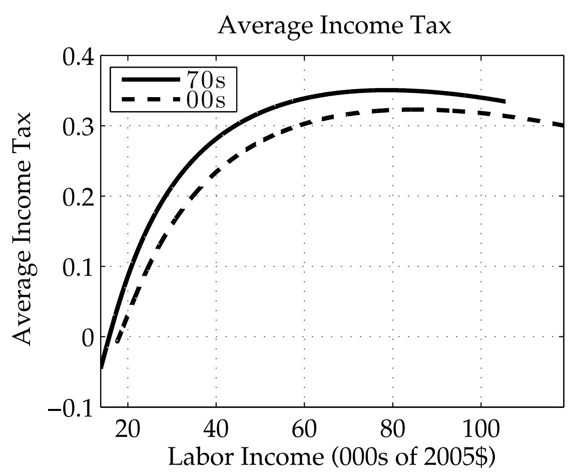


Figure 10: Average Tax Rates.

Figure 11 supplements Figure 10 by displaying optimal marginal tax rates. We observe that the optimal tax codes are progressive for low incomes (approximately for values below thirty thousand dollars.)[28] Above this value they become regressive, with diminishing marginal tax rates. Small transfers available to the lowest incomes in the 1970's are phased out in the 2000's. Marginal income taxes are cut for incomes below the median in the 2000's and overall workers pay lower average taxes by the 2000s'.

[28] All dollar denominated values are in \$ 2005 dollars.

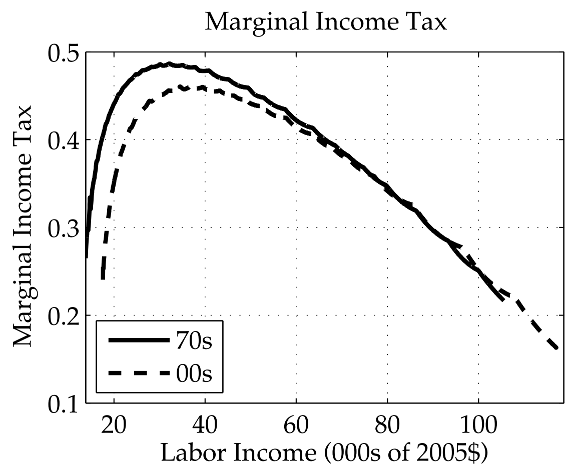


Figure 11: Marginal Tax Rates.

We now compare the optimal tax code with the empirical one. For this comparison we exclude payroll taxes as this gives the closest proxy to income taxes in our model. Table 3 gives optimal end empirical average tax rates across percentiles of the income distribution. and decades

Table 3: Average Tax Rates on Real Labor Income.

Source	**Decade**	**Percentiles of Income**					
		10^{th}	*25^{th}*	*50^{th}*	*75^{th}*	*90^{th}*	*99^{th}*
Optimal	70s	-2	3.3	16.3	29.9	34.8	33.9
	00s	-.5	0.9	8.8	23.6	31.5	30.5
Data: State + Federal	70s	10.8	15.9	19.5	21.9	25.6	35.4
	00s	2.6	11.8	18.2	22.2	26.1	35.0

For the "70s" the optimal average tax rates implied by the model are lower than actual average tax rates for incomes below the median and higher for those above. Optimal average taxes fall at all income levels in the"00s", with especially striking reductions for mid-level incomes. Actual average tax rates fell, at lower incomes, but were largely unchanged for incomes above the median.

6.2 Accounting for Taxes

In this section we decompose our optimal taxes following the formulas derived in Sections 3 and 4.

`To be Completed.`

6.3 Counterfactuals

In this section we display the impact of polarization and the change in comparative advantage in determining the optimal tax rates. We do this by computing the allocation by either keeping the b function or the a function fixed at it's values in the "70s".

`To be Completed.`

7 Conclusion

`To be Completed.`

References

Acemoğlu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics*, Volume 4B, pp. 1043–1171. Elsevier Press.

Ales, L., M. Kurnaz, and C. Sleet (2013). Optimal taxation with publicly observable tasks. Carnegie Mellon University.

Autor, D. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553–1597.

Bonnans, J. and A. Shapiro (2000). *Perturbation Analysis of Optimization Problems*. Springer-Verlag: Springer Series in Operations Research.

Brito, D., J. Hamilton, S. Slutsky, and J. Stiglitz (1995). Randomization in optimal income tax schedules. *Journal of Public Economics 56*, 189–223.

Buera, F. and J. Kaboski (2012). The rise of the service economy. *American Economic Review 102*(6), 2540–2569.

Costinot, A. and J. Vogel (2010). Matching and inequality in the world economy. *Journal of Political Economy 118*(4), 747–786.

Gouveia, M. and R. Strauss (1994). Effective federal individual income tax functions: An exploratory empirical analysis. *National Tax Journal 47*(2), 317–39.

Heathcote, J., F. Perri, and G. L. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the united states, 1967–2006. *Review of Economic Dynamics 13*(1), 15–51.

Hellwig, M. (2007). The undesirability of randomized income taxation under decreasing risk aversion. *Journal of Public Economics 91*(3-4), 791–806.

Jakobsson, U. (1976). On the measurement of the degree of progression. *Journal of public economics 5*(1), 161–168.

Kasten, R., F. Sammartino, and E. Toder (1994). Trends in federal tax progressivity, 1980–93. In J. Slemrod (Ed.), *Tax Progressivity and Income Inequality*, pp. 9–50. Cambridge, MA: Cambridge University Press.

Kocherlakota, N. (2005). Zero expected wealth taxes: a mirrlees approach to dynamic optimal taxation. *Econometrica 73*(5), 1587–1621.

Linton, O. and J. P. Nielsen (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika 82*, 93–100.

Milgrom, P. and C. Shannon (1994). Monotone comparative statics. *Econometrica 62*(1), 157–180.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies 38*, 175–208.

Musgrave, R. A. and T. Thin (1948). Income tax progression, 1929-48. *The Journal of Political Economy 56*(6), 498–514.

Pfingsten, A. (1986). *The measurement of tax progression*. Springer-Verlag Berlin.

Piketty, T. and E. Saez (2007). How progressive is the us federal tax system? a historical and international perspective. *The Journal of Economic Perspectives 21*(1), 3–24.

Rosen, S. (1978). Substitution and the division of labor. *Economica 45*(179), 235–250.

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the roy model. *Quarterly Journal of Economics 128*(2), 623–662.

Roy, A. (1950). The distribution of earnings and of individual output. *Economic Journal 60*, 489–501.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*(1), 205–229.

Salanié, B. (2011). *Economics of Taxation*. MIT Press, USA.

Sattinger, M. (1975). Comparative advantage and the distributions of earnings and abilities. *Econometrica 43*, 455–468.

Slavík, C. and H. Yazici (2012). Machines, buildings and optimal dynamic taxes. Sabanci University.

Sperlich, S., D. Tjøstheim, and L. Yang (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory 18*(02), 197–251.

Stiglitz, J. (1982). Self selection and pareto efficient taxation. *Journal of Public Economics 17*, 213–240.

Teulings, C. (1995). The wage distribution in a model of the assignment of skills to jobs. *Journal of Political Economy 103*(2), 280–315.

Yatchew, A. (2003). *Semiparametric regression for the applied econometrician*. Cambridge University Press.

Appendix

A Generalized Stiglitz Model

We first formalize a planning problem in a space of simple direct mechanisms and then recover an optimal tax equilibrium from its solution.

Implementation via simple, direct mechanisms A simple, direct mechanism is a tuple $\{c_k, e_k\}_{k=1}^{K}$ of message-contingent consumptions and effort recommendations and a severe penalty. The mechanism implicitly defines a set of shadow wages and incomes: $w_k = G_k(\pi_1 e_1, \ldots, \pi_K e_K)$ and $q_k = w_k e_K$. The planner and the agents play the following game. First, the planner selects a mechanism. Second, each agent sends a message $k \in \{1, \ldots, K\}$, exerts an effort and generates a publicly observable income q. Third, if an agent sends the message k and generates the income q_k, she receives the consumption c_k. Otherwise she receives the severe penalty. The latter is assumed to be sufficient to deter an agent who has sent message k from generating an income $q \neq q_k$. Without loss of generality, attention is restricted to mechanisms that induce agents to report truthfully conditional on almost all other agents doing so. Thus, the mechanism is constrained to satisfy, for all $k \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, K\} \backslash \{k\}$,

$$U(c_k, e_k) \geq U(c_j, q_j / w_k). \tag{26}$$

The planner's problem is then:

$$\sup \sum_{k=1}^{K} U(c_k, e_k) g_k \tag{27}$$

s.t. for all $k \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, K\} \backslash \{k\}$

$$U(c_k, e_k) \geq U\left(c_j, \frac{G_j(e_1\pi_1, \ldots, e_K\pi_K)}{G_k(e_1\pi_1, \ldots, e_K\pi_K)} e_j\right). \tag{28}$$

and

$$P + \sum_{k=1}^{K} c_k \pi_k \leq G(e_1\pi_1, \ldots, e_K\pi_K). \tag{29}$$

Let $\{c_k^*, e_k^*\}_{k=1}^{K}$ denote a solution to (27) with corresponding shadow wages $\{w_k^*\}_{k=1}^{K}$, $w_k^* = G_k(e_1^*\pi_1, \ldots, e_K^*\pi_K)$. Agent types may be ordered according to their shadow wages and relabeled accordingly.[29] Consequently, there is no loss of generality in assuming that a k-th type has the k-th highest wage; we make this assumption below. A (k, j)-th incentive compatibility constraint (28) is local if $j \in \{k-1, k+1\}$. Lemma A.1 below is a well known consequence of the Spence-Mirrlees single crossing property and the structure of

[29] Formally, let $\kappa : \{1, \ldots, K\} \to \{1, \ldots, K\}$ denote a permutation of $1, \ldots, K$ such that for $k = 1, \ldots, K-1$, $w^*_{\kappa(k+1)} \geq w^*_{\kappa(k)}$. Then types may be relabeled according to $k' = \kappa^{-1}(k)$.

the incentive constraints in settings with exogenous wages; it continues to hold in the present setting.[30]

Lemma A.1. *Assuming types are labeled according to their ranking in the wage distribution, then (i) $c_{k+1} \geq c_k$ and $q_{k+1} \geq q_k$, (ii) only "local" incentive constraints bind.*

Proof of Proposition 1. Let χ^* denote the optimal multiplier on the resource constraint and $\eta^*_{k,j}$ the optimal multiplier on the (k,j)-th incentive constraint. In light of the previous lemma only local incentive constraints are potentially binding and, hence, only the $\eta^*_{k,k-1}$ and $\eta^*_{k,k+1}$ are potentially non-zero. The first order condition for e^*_k reduces to:

$$-U_e(c^*_k, e^*_k)$$
$$= \frac{\chi^* w^*_k \pi_k}{g_k + \eta^*_{k,k-1} - \eta^*_{k+1,k} \frac{U_e\left(c^*_k, \frac{q^*_k}{w^*_{k+1}}\right)}{U_e(c^*_k, e^*_k)} \frac{w^*_k}{w^*_{k+1}} + \eta^*_{k,k+1} - \eta^*_{k-1,k} \frac{U_e\left(c^*_k, \frac{q^*_k}{w^*_{k-1}}\right)}{U_e(c^*_k, e^*_k)} \frac{w^*_k}{w^*_{k-1}} + \chi^* \Phi^*_k + \chi^* \Upsilon^*_k},$$

where:

$$\Phi^*_k := -\sum_{j=1}^{K-1} \frac{\eta^*_{j+1,j}}{\chi^*} \frac{U_e\left(c^*_j, \frac{q^*_j}{w^*_{j+1}}\right)}{U_e(c^*_k, e^*_k)} \left\{\phi^*_{k,j} - \phi^*_{k,j+1}\right\} \frac{e^*_j w^*_j}{e^*_k w^*_{j+1}},$$

and

$$\Upsilon^*_k := -\sum_{j=1}^{K-1} \frac{\eta^*_{j-1,j}}{\chi^*} \frac{U_e\left(c^*_j, \frac{q^*_j}{w^*_{j-1}}\right)}{U_e(c^*_k, e^*_k)} \left\{\phi^*_{k,j} - \phi^*_{k,j-1}\right\} \frac{e^*_j w^*_j}{e^*_k w^*_{j-1}}.$$

The first order condition for c^*_k reduces to:

$$U_c(c^*_k, e^*_k) = \frac{\chi^* \pi_k}{g_k + \eta^*_{k,k-1} - \eta^*_{k+1,k} \frac{U_c\left(c^*_k, \frac{q^*_k}{w^*_{k+1}}\right)}{U_c(c^*_k, e^*_k)} + \eta^*_{k,k+1} - \eta^*_{k-1,k} \frac{U_c\left(c^*_k, \frac{q^*_k}{w^*_{k-1}}\right)}{U_c(c^*_k, e^*_k)}}.$$

Define the consumption-effort wedge:

$$\frac{\tau^*_k}{1-\tau^*_k} = -\frac{w^*_k U_c(c^*_k, e^*_k)}{U_e(c^*_k, e^*_k)} - 1.$$

[30] We omit the proof. It can be shown as an application of Theorems 3 and 4 in Milgrom and Shannon (1994).

Combining expressions gives:

$$\frac{\tau_k^*}{1-\tau_k^*} = \frac{U_c(c_k^*, e_k^*)}{\pi_k} \left\{ \Phi_k^* + \Upsilon_k^* + \frac{\eta_{k+1,k}^*}{\chi^*} \left\{ \frac{U_c\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{U_c(c_k^*, e_k^*)} - \frac{U_e\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k+1}^*} \right\} \right.$$
$$\left. + \frac{\eta_{k-1,k}^*}{\chi^*} \left\{ \frac{U_c\left(c_k^*, \frac{q_k^*}{w_{k-1}^*}\right)}{U_c(c_k^*, e_k^*)} - \frac{U_e\left(c_k^*, \frac{q_k^*}{w_{k-1}^*}\right)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k-1}^*} \right\} \right\}.$$

Absent bunching, for all k, $\eta_{k,k+1}^* = 0$, $\Upsilon_k^* = 0$ and the previous expression reduces to:

$$\frac{\tau_k^*}{1-\tau_k^*} = \frac{U_c(c_k^*, e_k^*)}{\pi_k} \left\{ \Phi_k^* + \frac{\eta_{k+1,k}^*}{\chi^*} \left\{ \frac{U_c\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{U_c(c_k^*, e_k^*)} - \frac{U_e\left(c_k^*, \frac{q_k^*}{w_{k+1}^*}\right)}{U_e(c_k^*, e_k^*)} \frac{w_k^*}{w_{k+1}^*} \right\} \right\}.$$

Again absent bunching, the first order equation for c_{k+1}^* implies the following recursion for $\eta_{k+1,k}^*$:

$$\frac{\eta_{k+1,k}^*}{\chi^*} = \left(1 - \frac{g_{k+1} U_c(c_{k+1}^*, e_{k+1}^*)}{\chi^* \pi_{k+1}}\right) \frac{\pi_{k+1}}{U_c(c_{k+1}^*, e_{k+1}^*)} + \frac{\eta_{k+2,k+1}^*}{\chi^*} \frac{U_c\left(c_{k+1}^*, \frac{q_{k+1}^*}{w_{k+2}^*}\right)}{U_c(c_{k+1}^*, e_{k+1}^*)},$$

with $\eta_{K^*+1,K^*} = 0$. Iterating on this recursion, setting $\Psi_k^* := \eta_{k+1,k}^*$ and using the definition of $\mathcal{H}_k^*$ in the proposition gives (4). A standard appeal to the taxation principle establishes that the solution to the planning problem is an optimal tax equilibrium allocation and that τ_k^* is the optimal marginal tax rate for the k-th type. □

B K-Talent Assignment Model

Proof of Proposition 2. Let T, $\{c_k, e_k, \lambda_k\}_{k=1}^K$ and $\{\omega_k\}_{k=1}^K$ denote a tax equilibrium at spending level P. Since agents of a given type select the highest possible wage, it follows that for each k there is a $w_k < \infty$ such that for each $v \in \text{Supp } \Lambda_k$, $\omega_k(v) = w_k$ and for $v \notin \text{Supp } \Lambda_k$, $\omega_k(v) \le w_k$. Without loss of generality, assume that the firm's first order conditions hold at each k and $v \in \Lambda_k(v)$:

$$\omega_k(v) = b(v) \left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v).$$

If $v \in [\underline{v}, \overline{v}] \backslash \cup_{k=1}^k \text{Supp}(\Lambda_k)$, then $y(v) = 0$ and for each k, $\omega_k(v) = b(v) \left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v) = \infty > w_k$. Since this a contradiction, $[\underline{v}, \overline{v}] \backslash \cup_{k=1}^k \text{Supp}(\Lambda_k)$ is measure zero and almost all tasks are performed. For all v and v' in Supp Λ_k with $v > v'$,

$$1 = \frac{\omega_k(v)}{\omega_k(v')} = \frac{b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v)}{b(v')\left(\frac{Y}{y(v')}\right)^{\frac{1}{\varepsilon}} a_k(v')} < \frac{b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_{k+j}(v)}{b(v')\left(\frac{Y}{y(v')}\right)^{\frac{1}{\varepsilon}} a_{k+j}(v')} < \frac{\omega_{k+j}(v)}{\omega_{k+j}(v')}.$$

It follows that $v' \notin \Lambda_{k+j}$ and so $\sup \Lambda_k \leq \inf \Lambda_{k+j}$. Since the supports Λ_k essentially cover $[\underline{v}, \overline{v}]$, it follows that they partition $[\underline{v}, \overline{v}]$ into sub-intervals $[\tilde{v}_0, \tilde{v}_1]$, $[\tilde{v}_1, \tilde{v}_2]$, …, $\tilde{v}_{K-1}, \tilde{v}_K]$, with $\tilde{v}_0 = \underline{v}$, $\tilde{v}_K = \overline{v}$ and $\mathrm{cl}\Lambda_k = [\tilde{v}_{k-1}, \tilde{v}_k]$. By assumption each Λ_k has a density λ_k (concentrated on $[\tilde{v}_{k-1}, \tilde{v}_k]$). Since $w_k = \omega_k(v)$, $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$, we have for all such v:

$$w_k = b(v)\left(\frac{Y}{a_k(v)e_k\lambda_k(v)}\right)^{\frac{1}{\varepsilon}} a_k(v). \tag{30}$$

Hence, from the labor supply condition:

$$\pi_k = \int_{\tilde{v}_{k-1}}^{\tilde{v}_k} \lambda_k(v)dv = \frac{Y}{w_k^{\varepsilon}e_k} \int_{v_{k-1}}^{v_k} b(v)^{\varepsilon} a_k(v)^{\varepsilon-1}$$

And so:

$$w_k = B_k(\tilde{v}_{k-1}, \tilde{v}_k)\left(\frac{Y}{\pi_k e_k}\right)^{\frac{1}{\varepsilon}}, \tag{31}$$

where $B_k(\tilde{v}_{k-1}, \tilde{v}_k) := \left[\int_{\underline{v}_{k-1}}^{\overline{v}_k} b(v)^{\varepsilon} a_k(v)^{\varepsilon-1} dv\right]^{\frac{1}{\varepsilon}}$. Substituting (31) in (31) gives for $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$,

$$\lambda_k(v) = \frac{b(v)^{\varepsilon} a_k(v)^{\varepsilon-1}}{B_k(\tilde{v}_{k-1}, \tilde{v}_k)^{\varepsilon}} \pi_k.$$

And for $v < \tilde{v}_{k-1}$ and $v > \tilde{v}_k$, $\lambda_k(v) = 0$.

Now for $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$, $w_{k+1} > \omega_{k+1}(v) = b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_{k+1}(v)$ and $w_k = \omega_k(v) = b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v)$. Hence:

$$\frac{w_{k+1}}{w_k} > \frac{a_{k+1}(v)}{a_k(v)}.$$

Conversely, for $v \in (\tilde{v}_k, \tilde{v}_{k+1})$, $w_{k+1} = \omega_{k+1}(v) = b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_{k+1}(v)$ and $w_k > \omega_k(v) = b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} a_k(v)$. Hence:

$$\frac{w_{k+1}}{w_k} < \frac{a_{k+1}(v)}{a_k(v)}.$$

By continuity of a_k and a_{k+1},

$$\frac{w_{k+1}}{w_k} = \frac{a_{k+1}(\tilde{v}_k)}{a_k(\tilde{v}_k)}.$$

Combining the last equality with (31) gives the desired expression in the proposition.

Finally, given the effort allocation $\{e_k\}_{k=1}^K$ consider assigning agents so as to maximize output, i.e. solving:

$$\max_{\{\lambda_k\}} \left[\int_{\underline{v}}^{\overline{v}} b(v) \left\{ \lambda_k(v) e_k a_k(v) \right\}^{\frac{\varepsilon-1}{\varepsilon}} dv \right]^{\frac{\varepsilon}{\varepsilon-1}}.$$

subject to for each k, $\pi_k = \int_{\underline{v}}^{\overline{v}} \lambda_k(v) dv$. This is a strictly concave maximization whose unique solution is determined by the first order conditions. Straightforward manipulation of these conditions establishes that the λ_k solved for above attains the solution to this problem. $\square$

Proof of Lemma 1. Given an (equilibrium) effort profile $\{e_k\}_{k=1}^K$, (equilibrium) production maximizing task thresholds $\{\tilde{v}_k\}$ are determined by the conditions, $k = 1, \ldots, K-1$,

$$\frac{B_k(\tilde{v}_{k-1}, \tilde{v}_k)}{\{\pi_k e_k\}^{\frac{1}{\varepsilon}}} = \frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1})}{\{\pi_{k+1} e_{k+1}\}^{\frac{1}{\varepsilon}}} \frac{a_k(\tilde{v}_k)}{a_{k+1}(\tilde{v}_k)}.$$

with $\tilde{v}_0 = \underline{v}$ and $\tilde{v}_K = \overline{v}$. Hence, there are $K-1$ unknowns (and $K-1$ equations). The threshold sensitivities may be computed by taking logs in the preceding equations and totally differentiating with respect to $\log e_k$. This leads to the equations:

$$\Gamma \Delta v_k = E_k,$$

where:

$$\Gamma := \begin{pmatrix} \alpha_1 \tilde{v}_1 - \frac{\tilde{v}_1}{B_2}\frac{\partial B_2}{\partial \tilde{v}_1} + \frac{\tilde{v}_1}{B_1}\frac{\partial B_1}{\partial \tilde{v}_1} & -\frac{\tilde{v}_2}{B_2}\frac{\partial B_2}{\partial \tilde{v}_2} & 0 & 0 & \ldots & 0 \\ \frac{\tilde{v}_1}{B_2}\frac{\partial B_2}{\partial \tilde{v}_1} & \alpha_2 \tilde{v}_2 - \frac{\tilde{v}_2}{B_3}\frac{\partial B_3}{\partial \tilde{v}_2} + \frac{\tilde{v}_2}{B_2}\frac{\partial B_2}{\partial \tilde{v}_2} & -\frac{\tilde{v}_3}{B_3}\frac{\partial B_3}{\partial \tilde{v}_3} & 0 & \ldots & 0 \\ 0 & \frac{\tilde{v}_2}{B_3}\frac{\partial B_3}{\partial \tilde{v}_2} & \alpha_3 \tilde{v}_3 - \frac{\tilde{v}_3}{B_4}\frac{\partial B_4}{\partial \tilde{v}_3} - \frac{\tilde{v}_3}{B_3}\frac{\partial B_3}{\partial \tilde{v}_3} & -\frac{\tilde{v}_4}{B_4}\frac{\partial B_4}{\partial \tilde{v}_4} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$\Delta v_k = (\frac{\partial \log v_1}{\partial \log e_k} \ldots \frac{\partial \log v_k}{\partial \log e_k} \ldots \frac{\partial \log v_{K-1}}{\partial \log e_k})'$ and $E_k = (0 \ldots -\frac{1}{\varepsilon}\ \frac{1}{\varepsilon} \ldots 0)'$ with non-zero elements in the $k-1$ and k-th rows. Thus,

$$\Delta v_k = \Gamma_k^{-1} E_k.$$

and in fact:

$$\frac{\partial \log v_j}{\partial \log e_k} = (\delta_{j,k-1} - \delta_{j,k}) \frac{1}{\varepsilon},$$

where $\delta_{j,k}$ is the (j,k)-th element of Γ_k^{-1}. Since Γ_k is a tridiagonal matrix, explicit formulas for its inverse are available. Applying these formulas gives the expression in the text. $\square$

C Technical Change in the Two Talent Assignment Model

C.1 Analysis of the Full Information Case

Proof of Lemma 2. The planner's full information problem is:

$$\sup \sum_{j\in\{L,H\}} \{u(c_j) - e_j^{1+\gamma}/(1+\gamma)\} g_j \tag{FP}$$

subject to: $\lambda_j(v) \geq 0$, $e_j \in [0, \bar{e}]$, $c_j \geq 0$,

$$\sum_{j\in\{L,H\}} c_j \pi_j \leq A \left\{ \int_{\underline{v}}^{\overline{v}} b(v) \left\{ \sum_{j\in\{L,H\}} \lambda_j(v) a_j(v) e_j \right\}^{\frac{\varepsilon-1}{\varepsilon}} dv \right\}^{\frac{\varepsilon}{\varepsilon-1}}, \tag{32}$$

and:

$$\pi_j = \int_{\underline{v}}^{\overline{v}} \lambda_j(v) dv. \tag{33}$$

Let χ and κ_j denote the multipliers on the resource and j-th labor supply constraint; let δ_j denote the multiplier on the constraint $\lambda_j \geq 0$. The first order condition for $\lambda_j^F(v)$ is:

$$\chi b(v) a_j(v) e_j^F \left(\frac{Y}{\sum_{j'\in\{L,H\}} \lambda_{j'}^F(v) a_{j'}(v) e_{j'}^F} \right)^{\frac{1}{\varepsilon}} - \kappa_j + \delta_j(v) = 0.$$

For almost every v such that $\lambda_j^F(v) > 0$:

$$\chi b(v) a_j(v) e_j^F \left(\frac{Y}{\sum_{j'\in\{L,H\}} \lambda_{j'}^F(v) a_{j'}(v) e_{j'}^F} \right)^{\frac{1}{\varepsilon}} = \kappa_j.$$

Hence, for almost every v such that $\lambda_L^F(v) > 0$ and $\lambda_H^F(v) > 0$:

$$\frac{a_L(v) e_L^F}{a_H(v) e_H^F} = \frac{\kappa_L}{\kappa_H}.$$

However, since a_L/a_H is a strictly decreasing function, the previous equality can hold only at a single point. Denote this point $\tilde{v}^F$. For $v < \tilde{v}^F$, $\frac{a_L(v)e_L^F}{a_H(v)e_H^F} > \frac{\kappa_L}{\kappa_H}$ and it is (weakly) optimal to set $\lambda_H^F(v) = 0$. Conversely, for $v < \tilde{v}^F$, it is weakly optimal to set $\lambda_L^F(v) = 0$. In addition, the first order conditions for λ_j^F reduce to:

$$\lambda_L^F(v) = \begin{cases} \left(\frac{\chi}{\kappa_L}\right)^{\varepsilon} b(v)^{\varepsilon}(a_L(v)e_L)^{\varepsilon-1}Y & v < \tilde{v}^F \\ 0 & v > \tilde{v}^F \end{cases}$$

and

$$\lambda_H^F(v) = \begin{cases} 0 & v < \tilde{v}^F \\ \left(\frac{\chi}{\kappa_H}\right)^{\varepsilon} b(v)^{\varepsilon}(a_H(v)e_H)^{\varepsilon-1}Y & v > \tilde{v}^F. \end{cases}$$

Substituting these expressions into the labor supply conditions (33) gives:

$$\kappa_L = \chi Y^{\frac{1}{\varepsilon}} e_L^{\frac{\varepsilon-1}{\varepsilon}} B_L(\tilde{v}^F)\pi_L^{-\frac{1}{\varepsilon}} \quad \text{and} \quad \kappa_H = \chi Y^{\frac{1}{\varepsilon}} e_H^{\frac{\varepsilon-1}{\varepsilon}} B_H(\tilde{v}^F)\pi_H^{-\frac{1}{\varepsilon}}, \tag{34}$$

where $B_L(\tilde{v}) = \left[\int_{\underline{v}}^{\tilde{v}} b(v)^{\varepsilon} a_L(v)^{\varepsilon-1}dv\right]^{\frac{1}{\varepsilon}}$ and $B_H(\tilde{v}) = \left[\int_{\tilde{v}}^{\overline{v}} b(v)^{\varepsilon} a_H(v)^{\varepsilon-1}dv\right]^{\frac{1}{\varepsilon}}$. And so:

$$\lambda_L^F(v) = \begin{cases} \frac{b(v)^{\varepsilon}a_L(v)^{\varepsilon-1}}{B_L(\tilde{v}^F)^{\varepsilon}}\pi_L & \text{if } v < \tilde{v}^F \\ 0 & \text{if } v > \tilde{v}^F \end{cases} \quad \text{and} \quad \lambda_H^F(v) = \begin{cases} 0 & \text{if } v < \tilde{v}^F \\ \frac{b(v)^{\varepsilon}a_H(v)^{\varepsilon-1}}{B_H(\tilde{v}^F)^{\varepsilon}}\pi_H & \text{if } v > \tilde{v}^F. \end{cases}$$

Also, from the definition of $\tilde{v}^F$ and (34):

$$\frac{a_L(\tilde{v}^F)}{a_H(\tilde{v}^F)} = \frac{\kappa_L e_H^F}{\kappa_H e_L^F} = \frac{B_L(\tilde{v}^F)}{(\pi_L e_L^F)^{1/\varepsilon}} \Bigg/ \frac{B_H(\tilde{v}^F)}{(\pi_H e_H^F)^{1/\varepsilon}}.$$

The first order conditions for effort are then:

$$e_L^{F\gamma} g_L = \chi B_L(\tilde{v}^F)\left(\frac{Y}{e_L^F}\right)^{\frac{1}{\varepsilon}} \pi_L^{\frac{\varepsilon-1}{\varepsilon}} \quad \text{and} \quad e_H^{F\gamma} g_H = \chi B_H(\tilde{v}^F)\left(\frac{Y}{e_H^F}\right)^{\frac{1}{\varepsilon}} \pi_H^{\frac{\varepsilon-1}{\varepsilon}}.$$

Using the first order conditions for consumption: $\sum_{j\in\{L,H\}} u_c(c_j^F)g_j = \chi$ and so, for $k \in \{L,H\}$,

$$e_k^F = \left\{ \left(\sum_{j\in\{L,H\}} u_c(c_j^F)g_j\right) B_k(\tilde{v}^F)\frac{Y^{1/\varepsilon}}{g_k}\pi_k^{\frac{\varepsilon-1}{\varepsilon}} \right\}^{\frac{\varepsilon}{1+\gamma\varepsilon}}.$$

Combining the first order conditions for effort:

$$\frac{e_H^F}{e_L^F} = \left(\frac{B_H(\tilde{v}^F)/g_H}{B_L(\tilde{v}^F)/g_L}\left(\frac{\pi_H}{\pi_L}\right)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{1+\gamma\varepsilon}}.$$

Hence:

$$\frac{a_H(\tilde{v}^F)}{a_L(\tilde{v}^F)} = \left(\frac{B_H(\tilde{v}^F)}{B_L(\tilde{v}^F)}\right)^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} \left(\frac{g_H}{g_L}\right)^{\frac{1}{1+\gamma\varepsilon}} \left(\frac{\pi_L}{\pi_H}\right)^{\frac{1+\gamma}{1+\gamma\varepsilon}}$$

$\square$

C.2 Analysis of the Private Information Case

In this section we solve for an optimal tax equilibrium in an economy with two types. As noted in the text, this problem can be decomposed into an outer "generalized Stiglitz" step and an inner "assignment problem" step. This decomposition is useful relating our model to the existing literature. Here, however, we first combine the steps and simultaneously solve for the consumption-effort allocation and the assignment of talent to tasks. We proceed by solving a planner's problem in the space of simple, direct mechanisms and recovering from this optimal marginal taxes. Under the additional restriction of quasi-linear preferences, we then derive Proposition 3.

The planner's problem is:

$$\sup \sum_{j\in\{L,H\}} \{u(c_j) - e_j^{1+\gamma}/(1+\gamma)\} g_j \tag{A2}$$

subject to: $\lambda_j(v) \geq 0$, $e_j \in [0, \bar{e}]$, $c_j \geq 0$,

$$\sum_{j\in\{L,H\}} c_j \pi_j \leq A \left\{ \int_{\underline{v}}^{\overline{v}} b(v) \left\{ \sum_{j\in\{L,H\}} \lambda_j(v) a_j(v) e_j \right\}^{\frac{\varepsilon-1}{\varepsilon}} dv \right\}^{\frac{\varepsilon}{\varepsilon-1}}, \tag{35}$$

for $j \in \{L, H\}$,

$$\pi_j = \int_{\underline{v}}^{\overline{v}} \lambda_j(v) dv. \tag{36}$$

and for each $j, k \in \{L, H\}$, v'' in the support of λ_k and all v',

$$c_j - \frac{e_j^{1+\gamma}}{1+\gamma} \geq c_k - \left(\frac{\frac{b(v'')}{\{\sum_{l\in\{L,H\}} \lambda_l(v'') a_l(v'') e_l\}^{\frac{1}{\varepsilon}}} a_k(v'')}{\frac{b(v')}{\{\sum_{l\in\{L,H\}} \lambda_l(v') a_l(v') e_l\}^{\frac{1}{\varepsilon}}} a_j(v')} \right)^{1+\gamma} \frac{e_k^{1+\gamma}}{1+\gamma}. \tag{37}$$

Lemma C.1. *Assume $\alpha_0 > 0$. Let $\{c_j, e_j, \lambda_j\}$, $j \in \{L, H\}$ denote an incentive-compatible mechanism. There is a threshold task $\tilde{v} \in [\underline{v}, \overline{v}]$ such that:*

$$\lambda_L(v) = \begin{cases} \frac{b(v)^\varepsilon a_L(v)^{\varepsilon-1}}{B_L(\tilde{v})} \pi_L & v < \tilde{v} \\ 0 & v > \tilde{v}^* \end{cases} \quad \text{and} \quad \lambda_H(v) = \begin{cases} 0 & v < \tilde{v}^* \\ \frac{b(v)^\varepsilon a_H(v)^{\varepsilon-1}}{B_H(\tilde{v})} \pi_H & v > \tilde{v}. \end{cases} \tag{38}$$

The talent premium $\frac{w_H}{w_L}$ is:

$$\frac{w_H}{w_L} = \frac{B_H(\tilde{v})}{\{\pi_H e_H\}^{\frac{1}{\varepsilon}}} \Bigg/ \frac{B_L(\tilde{v})}{\{\pi_L e_L\}^{\frac{1}{\varepsilon}}} = \frac{a_H(\tilde{v})}{a_L(\tilde{v})}. \tag{39}$$

Proof. It is immediate from the incentive constraint that for all v in the support of λ_j and all v':

$$\frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_j(v) \geq \frac{b(v')}{y(v')^{\frac{1}{\varepsilon}}}a_j(v').$$

If v' is not in the support of either $\lambda_{j'}$, then for a given j, $\frac{b(v')}{y(v')^{1/\varepsilon}}a_j(v') = \infty$. But the support of λ_j cannot be empty and so there is a v with $\infty > \frac{b(v)}{y(v)^{1/\varepsilon}}a_j(v) \geq \frac{b(v')}{y(v')^{1/\varepsilon}}a_j(v')$. It follows that, in fact, all v' belong to the support of at least one $\lambda_{j'}$. If v and v' are in the support of λ_L and $\alpha_0 > 0$, then:

$$\frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}\bigg/\frac{b(v')}{y(v')^{\frac{1}{\varepsilon}}} = \frac{a_L(v')}{a_L(v)} < \frac{a_H(v')}{a_H(v)}$$

and so v is not in the support of λ_H. Consequently, any incentive-compatible mechanism admits a threshold task $\tilde{v}$ below (resp. above) which high (resp. low) talents do not work. Since the shadow wage of high (resp. low) talents must equal a common value above (resp. below) $\tilde{v}$, it follows that:

$$\frac{w_H}{Y} = \frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_H(v) = \frac{b(v)a_H(v)^{\frac{\varepsilon-1}{\varepsilon}}}{\{\lambda_H(v)e_H\}^{\frac{1}{\varepsilon}}}, \quad \forall v \in (\tilde{v}, \overline{v}]$$

and

$$\frac{w_L}{Y} = \frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_L(v) = \frac{b(v)a_L(v)^{\frac{\varepsilon-1}{\varepsilon}}}{\{\lambda_L(v)e_L\}^{\frac{1}{\varepsilon}}}, \quad \forall v \in [\underline{v}, \tilde{v})$$

Combining these expressions with the labor supply constraints gives $\lambda_H = \frac{b(v)^{\varepsilon}a_H(v)^{\varepsilon-1}}{B_H(v)(\tilde{v})}\pi_H$ (for $v > \tilde{v}$), $\lambda_L = \frac{b(v)^{\varepsilon}a_L^{\varepsilon-1}}{B_L(\tilde{v})}\pi_L$ (for $v < \tilde{v}$) and

$$\frac{w_L}{Y} = \frac{B_L(\tilde{v})}{\{e_L\pi_L\}^{1/\varepsilon}} \text{ and } \frac{w_H}{Y} = \frac{B_H(\tilde{v})}{\{e_H\pi_H\}^{1/\varepsilon}}.$$

For all $v < \tilde{v}$, $w_H > \frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_H(v)$ and $w_L = \frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_L(v)$ and so:

$$\frac{w_H}{w_L} > \frac{a_H(v)}{a_L(v)}$$

Conversely, for all $v > \tilde{v}$, $w_H = \frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_H(v)$ and $w_L > \frac{b(v)}{y(v)^{\frac{1}{\varepsilon}}}a_L(v)$ and so:

$$\frac{w_H}{w_L} < \frac{a_H(v)}{a_L(v)}.$$

By continuity of a_H and a_L:

$$\frac{w_H}{w_L} = \frac{a_H(\tilde{v})}{a_L(\tilde{v})}.$$

□

Lemma C.2. *Given condition* (18) *and* $g_L \geq \pi_L$, *low talents strictly prefer to reveal their type in any optimal mechanism.*

Proof (Sketch). Recall that under our assumption, the shadow wage of high talents exceeds that of low talents at the first best. Consider imposing only the incentive constraint on the high talents. It is easily verified that if the solution to this problem involves an increase in low talent effort, a reduction in high talent effort and, if $\alpha_0 > 0$, an increase in $\tilde{v}$ and, hence, an increase in the shadow wage relative to the first best. At this solution, if $g_L \geq \pi_L$, low talents are strictly better off reporting their type and so the low talent incentive constraint does not bind. □

Lemma C.3. *Optimal marginal tax rates are given by:*

$$\frac{\tau_L^*}{1-\tau_L^*} = \left(\frac{E[1/u_c(c^*)]}{1/u_c(c_L^*)}\frac{g_L}{\pi_L} - 1\right)\left\{1 - \left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma}\left\{1 - \frac{1}{\mathcal{E}(e_H^*/e_L^*)}\right\}\right\} \qquad \geq 0 \tag{40}$$

and

$$\frac{\tau_H^*}{1-\tau_H^*} = -\left(1 - \frac{E[1/u_c(c^*)]}{1/u_c(c_H^*)}\frac{g_H}{\pi_H}\right)\left(\frac{q_L^*}{q_H^*}\right)^{1+\gamma}\frac{1}{\mathcal{E}(e_H^*/e_L^*)} \qquad \leq 0. \tag{41}$$

Proof. In this proof we simultaneously solve for the optimal effort-consumption allocation and the optimal task threshold. Thus, $\tilde{v}$ is a choice variable and the constraint (39) is explicitly imposed. The first order conditions for consumption are:

$$0 = u_c(c_L^*)g_L - \chi^*\pi_L - \eta u_c(c_L^*) \tag{42}$$
$$0 = u_c(c_H^*)g_H - \chi^*\pi_H + \eta u_c(c_H^*). \tag{43}$$

Hence, $\chi^* = [\sum_{j\in\{L,H\}} \pi_j/u_c(c_j^*)]^{-1}$ and $\eta^* = g_L - \chi^*\pi_L/u_c(c_L^*) = \chi^*\pi_H/u_c(c_H^*) - g_H$. The first order condition for e_L^* is:

$$\begin{aligned} 0 = {} & -e_L^{*\gamma}g_L + \chi^* B_L(\tilde{v}^*)\left(\frac{Y^*}{e_L^*}\right)^{\frac{1}{\varepsilon}}\pi_L^{\frac{\varepsilon-1}{\varepsilon}} \\ & + \eta^*\left(\frac{\varepsilon-1}{\varepsilon}\right)\left(\frac{B_L(\tilde{v}^*)/\{e_L^*\pi_L\}^{\frac{1}{\varepsilon}}}{B_H(\tilde{v}^*)/\{e_H^*\pi_H\}^{\frac{1}{\varepsilon}}}e_L^*\right)^{1+\gamma}\frac{1}{e_L^*} - \frac{\mu^*}{\varepsilon}\frac{B_L(\tilde{v}^*)}{\{e_L^*\pi_L\}^{\frac{1}{\varepsilon}}}\frac{1}{e_L^*}, \end{aligned} \tag{44}$$

where μ^* is the optimal multiplier on (39). The first order condition for $\tilde{v}^*$ is:

$$\begin{aligned} & \mu^*\frac{B_L(\tilde{v}^*)}{\{e_L^*\pi_L\}^{\frac{1}{\varepsilon}}}\left\{\varepsilon\alpha_0 + \left(\frac{b_L(\tilde{v}^*)}{B_L(\tilde{v}^*)}\right)^{\varepsilon} + \left(\frac{b_H(\tilde{v}^*)}{B_H(\tilde{v}^*)}\right)^{\varepsilon}\right\} \\ & \quad + \eta^*\left(\frac{B_L(\tilde{v}^*)/\{e_L^*\pi_L\}^{\frac{1}{\varepsilon}}}{B_H(\tilde{v}^*)/\{e_H^*\pi_H\}^{\frac{1}{\varepsilon}}}e_L^*\right)^{1+\gamma}\left\{\left(\frac{b_L(\tilde{v}^*)}{B_L(\tilde{v}^*)}\right)^{\varepsilon} + \left(\frac{b_H(\tilde{v}^*)}{B_H(\tilde{v}^*)}\right)^{\varepsilon}\right\} = 0. \end{aligned} \tag{45}$$

Combining (44) and (45) and using the definitions of w_H^*, w_L^* and $\mathcal{E}(e_H/e_L)$ gives **??** in the main text:

$$-e_L^{*\gamma} g_L + \chi^* w_L^* \pi_L + \eta^* \left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma} \left\{1 - \frac{1}{\mathcal{E}(e_H^*/e_L^*)}\right\} e_L^{*\gamma} = 0. \tag{46}$$

The first order condition for high talent effort is:

$$\begin{aligned} 0 = &-e_H^{*\gamma} g_H + \chi^* B_H(\tilde{v}^*) \left(\frac{Y^*}{e_H^*}\right)^{\frac{1}{\varepsilon}} \pi_H^{\frac{\varepsilon-1}{\varepsilon}} \quad (47)\\ &+ \eta^* \left\{-e_H^{*1+\gamma} + \frac{1}{\varepsilon}\left(\frac{B_L(\tilde{v}^*)/\{e_L^*\pi_L\}^{\frac{1}{\varepsilon}}}{B_H(\tilde{v}^*)/\{e_H^*\pi_H\}^{\frac{1}{\varepsilon}}} e_L^*\right)^{1+\gamma}\right\} \frac{1}{e_H^*} + \frac{\mu^*}{\varepsilon} \frac{B_H(\tilde{v}^*)}{\{e_H^*\pi_H\}^{\frac{1}{\varepsilon}}} \frac{1}{e_H^*} \frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)}. \end{aligned}$$

Combining (47) with (45) gives:

$$-e_H^{*\gamma} g_H + \chi^* w_H^* \pi_H - \eta^* e_H^{*\gamma} + \eta^* \left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma} \frac{1}{\mathcal{E}(e_H^*/e_L^*)} \frac{e_L^*}{e_H^*} e_L^{*\gamma} = 0. \tag{48}$$

Tax implementation of the optimal second best allocation is achieved if $1 - \tau_j^*$, where τ_j^* is the marginal income tax on the j-th talent, equals the optimal consumption-effort wedge. For low talents this is given by:

$$1 - \tau_L^* = \frac{e_L^{*\gamma}}{w_L^* u_c(c_L^*)} = \left\{\frac{g_L - \eta^*}{g_L - \eta^* \left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma} \left\{1 - \frac{1}{\mathcal{E}(e_H^*/e_L^*)}\right\}}\right\}$$

Hence:

$$\frac{\tau_L^*}{1 - \tau_L^*} = \frac{\eta^*}{g_L - \eta^*} \left\{1 - \left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma} \left\{1 - \frac{1}{\mathcal{E}(e_H^*/e_L^*)}\right\}\right\} > 0.$$

Substituting for η^* from (42) then gives the formula in the proposition. The optimal consumption-effort wedge for high talents is:

$$1 - \tau_H^* = \frac{e_H^{*\gamma}}{w_H^* u_c(c_H^*)} = \frac{g_H + \eta^*}{g_H + \eta^* + \eta^* \left(\frac{q_L^*}{q_H^*}\right)^{1+\gamma} \frac{1}{\mathcal{E}(e_H^*/e_L^*)}}.$$

Hence:

$$\frac{\tau_H^*}{1 - \tau_H^*} = -\frac{\eta^*}{g_H + \eta^*} \left(\frac{q_L^*}{q_H^*}\right)^{1+\gamma} \frac{1}{\mathcal{E}(e_H^*/e_L^*)} < 0.$$

Substituting for η^* from (43) then gives the formula in the proposition. □

Proof of Proposition 3. Under quasi-linear preferences ($u(c) = c$), the first order conditions for consumption and the fact that the lower (zero) bounds on consumption are non-binding give: $\chi^* = 1$ and

$$\eta^* = g_L - \pi_L = \pi_H - g_H.$$

Case 1: $\alpha_0 = \infty, \varepsilon \geq 1$. Substituting the previous expressions into (46) and (48) and using $\mathcal{E}(e_H^*/e_L^*) = \varepsilon$ when $\alpha_0 = \infty$ gives:

$$e_L^{*\gamma} = \frac{w_L^* \pi_L}{g_L - (g_L - \pi_L)\frac{\varepsilon-1}{\varepsilon}\left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma}}$$

and

$$e_H^{*\gamma} = \frac{w_H^* \pi_H}{\pi_H - (\pi_H - g_H)\frac{1}{\varepsilon}\left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma}\left(\frac{e_L^*}{e_H^*}\right)^{1+\gamma}}.$$

If $\alpha_0 = \infty$, then $\tilde{v}^*$ is fixed at $\hat{v}$ and writing B_j for $B_j(\hat{v})$, the talent premium is given by $\frac{w_L^*}{w_H^*} = \frac{B_L}{B_H}\left(\frac{\pi_H e_H^*}{\pi_L e_L^*}\right)^{\frac{1}{\varepsilon}}$. Dividing the previous two equations for $e_j^*\gamma$ into one another and using this expression for the talent premium gives:

$$\left(\frac{e_L^*}{e_H^*}\right)^{\frac{\gamma\varepsilon+1}{\varepsilon}} = \left(\frac{B_L}{B_H}\right)\left(\frac{\pi_L}{\pi_H}\right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{\pi_H - (\pi_H - g_L)\frac{1}{\varepsilon}\left(\frac{B_L}{B_H}\right)^{1+\gamma}\left(\frac{e_L^*}{e_H^*}\right)^{(1+\gamma)\frac{\varepsilon-1}{\varepsilon}}}{g_L - (g_L - \pi_L)\frac{\varepsilon-1}{\varepsilon}\left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma}}$$

After further rearrangement and substitution:

$$\left(\frac{w_H^*}{w_L^*}\right)^{\gamma\varepsilon+1}\left(\frac{B_L}{B_H}\right)^{\gamma\varepsilon}\left(\frac{\pi_H}{\pi_L}\right)^{\frac{\gamma\varepsilon+1}{\varepsilon}}$$
$$= \left(\frac{\pi_L}{\pi_H}\right)^{\frac{\varepsilon-1}{\varepsilon}} \frac{\pi_H - (\pi_H - g_L)\frac{1}{\varepsilon}\left(\frac{B_L}{B_H}\right)^{(\varepsilon-1)(1+\gamma)}\left(\frac{w_H^*}{w_L^*}\right)^{(\varepsilon-1)(1+\gamma)}\left(\frac{\pi_H}{\pi_L}\right)^{(1+\gamma)\frac{\varepsilon-1}{\varepsilon}}}{g_L - (g_L - \pi_L)\frac{\varepsilon-1}{\varepsilon}\left(\frac{w_L^*}{w_H^*}\right)^{1+\gamma}} \tag{49}$$

The left hand side is increasing in w_H^*/w_L^* and B_L/B_H; with $\varepsilon \geq 1$, the right hand side is decreasing in w_h^*/w_L^* and B_L/B_H. It follows that an increase in B_H/B_L (caused, for example, by a tilting of demand towards more complex tasks) causes the talent premium to rise and, by our optimal tax formulas, an increase in marginal taxes on low talents. Note that the right hand side of the equality must remain positive and so if $B_L/B_H \to \infty$, then $w_H^*/w_L^* \to 0$. On the other hand, as $B_L/B_H \to 0$, $w_H^*/w_L^* \to \infty$. Thus, for the talent premium in this case to exceed that when $\alpha_0 = 0$ (i.e. to exceed α_H/α_L) requires $B_H(\hat{v})/B_L(\hat{v})$ to be large enough.

Case 2: $\alpha_0 \in (0, \infty)$, $\varepsilon \geq 1$, $\mathcal{E} \geq 1$ and locally constant. In this case, substituting into (39) gives:

$$\frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)} = \left(\frac{B_H(\tilde{v}^*)}{B_L(\tilde{v}^*)}\right)^{\frac{\gamma\varepsilon}{1+\gamma\varepsilon}} \left(\frac{\pi_L}{\pi_H}\right)^{\frac{1+\gamma}{1+\gamma\varepsilon}}$$
$$\times \left\{ \frac{\pi_H - (\pi_H - g_H)\frac{1}{\mathcal{E}} \left(\frac{a_H(\tilde{v}^*)}{a_L(\tilde{v}^*)}\right)^{(1+\gamma)(\varepsilon-1)} \left(\frac{B_H(\tilde{v}^*)}{B_L(\tilde{v}^*)}\right)^{-(1+\gamma)\varepsilon}}{g_L - (g_L - \pi_L)\left(\frac{a_L(\tilde{v}^*)}{a_H(\tilde{v}^*)}\right)^{1+\gamma} \left(\frac{\mathcal{E}-1}{\mathcal{E}}\right)} \right\}^{\frac{1}{1+\gamma\varepsilon}}. \tag{50}$$

Holding $\mathcal{E}$ constant, the left hand side function is increasing in $\tilde{v}^*$, while the right hand side function ins increasing in $\tilde{v}^*$ so there is at most one solution to the equation. If $B_H(v)/B_L(v)$ increases at all v, then the right hand side of the equation must rise at all v and, consequently, holding $\mathcal{E}$ constant, $\tilde{v}^*$ and the talent premium must rise. Thus, if $\mathcal{E}$ is (locally) constant, (small) increases in B_H/B_L at all v lead to increases in $\tilde{v}^*$ and the talent premium. □

D CONTINUOUS TALENT-CONTINUOUS TASK MODEL

In this appendix, we briefly describe the continuous talent (and continuous task) version of the model from Section 4. In our quantitative work, we treat the data as a discrete approximation to a continuous model and use optimal control methods to solve it.[31] Agents are now distributed across talents an interval of talents $\theta \in \Theta := [\underline{\theta}, \overline{\theta}]$ according to a distribution function $F : \Theta \to \mathbb{R}_+$ with strictly positive and continuously differentiable density f. As before there is a continuum of tasks ranked by complexity $v \in V := [\underline{v}, \overline{v}]$. The productivity of talent-task combinations is given by a function $a : \Theta \times V \to \mathbb{R}_{++}$ satisfying the following assumption.[32]

Assumption 2. *(i) a is twice continuously differentiable on the interior of $\Theta \times V$ with first derivatives a_i, $i \in \{\theta, v\}$ and second derivatives a_{ij}, $i, j \in \{\theta, v\}$. (ii) (absolute advantage) $a_\theta > 0$, (iii) (comparative advantage, log supermodularity) $\frac{\partial^2 \log a}{\partial\theta\partial v} > 0$.*

Otherwise technologies and preferences are as in the main text. An allocation is a triple of measurable functions $c : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}_+$, $\nu : [\underline{\theta}, \overline{\theta}] \to [\underline{v}, \overline{v}]$ and $e : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}_+$ describing the consumption, task and effort assignments of each talent type.[33] As before, task output is linear in labor input. The task output density $y : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ satisfies for

[31] Thus, we solve for effort and the task assignment simultaneously rather than utilizing the two step effort-assignment decomposition described in the main text.

[32] Our treatment of talent and task complexity follows that in the assignment literature. As there, we assume that the different attributes of talent can be compressed into a single index that captures the value of talent in production and that the mapping of these attributes into the index is stable over time.

[33] The implicit assumption that all talents are assigned to a specific consumption, task and effort is without loss of generality. It may be shown, along the lines of Proposition 2, that assignment of talents to tasks is strictly increasing in talent given strict comparative advantage.

all θ,

$$\int_{\underline{v}}^{\nu(\theta)} y(u)du = \int_{\underline{\theta}}^{\theta} a[t, \nu(t)]e(t)f(t)dt. \tag{51}$$

If ν is differentiable with derivative ν_θ, then (51) can be re-expressed as, for all θ:

$$y(\nu(\theta)) = a[\theta, \nu(\theta)]e(\theta)\frac{f(\theta)}{\nu_\theta(\theta)}, \tag{52}$$

Heuristically, the numerator is total output of type θ, while the denominator gives the tasks over which the type θ agents are "spread". Throughout the remainder of the paper, we restrict planners and policymakers to smooth allocations and mechanisms. This permits the application of optimal control techniques.

FIRST BEST ALLOCATIONS We briefly describe the first best solution in this case. To permit a simple handling of the resource constraint, define the (state) variables $\xi : \Theta \to \mathbb{R}$ as:

$$\xi(\theta) := \int_{\underline{\theta}}^{\theta} \left(\frac{f(t)a[t, \nu(t)]e(t)}{\nu_\theta(t)}\right)^{\frac{\varepsilon-1}{\varepsilon}} b(v(t))\nu_\theta(t)dt \tag{53}$$

and

$$\kappa(\theta) := \int_{\underline{\theta}}^{\theta} c(t)f(t)dt. \tag{54}$$

The planner's first best (optimal control) problem may then be stated as:

$$\sup_{c,e,\nu,\xi,\kappa} \int_{\underline{\theta}}^{\overline{\theta}} U(c(\theta), e(\theta))g(\theta)d\theta \tag{55}$$

subject to:

$$\xi_\theta(\theta) = \{f(\theta)a[\theta, \nu(\theta)]e(\theta)\}^{\frac{\varepsilon-1}{\varepsilon}} \nu_\theta(\theta)^{\frac{1}{\varepsilon}} b(\nu(\theta)) \tag{56}$$

$$\kappa_\theta(\theta) = c(\theta)f(\theta) \tag{57}$$

and the boundary constraints $\xi(\underline{\theta}) = 0$, $\kappa(\overline{\theta}) \leq \xi(\overline{\theta})^{\frac{\varepsilon}{\varepsilon-1}}$ and $0 \leq \nu(\underline{\theta}) \leq \nu(\bar{\theta}) \leq \overline{v}$. Define the shadow wage of talent θ in task v as:

$$w[\theta, v] = b(v)\left(\frac{y(v)}{Y}\right)^{-\frac{1}{\varepsilon}} a[\theta, v]. \tag{58}$$

The first order and co-state equations from (55) imply:

$$U_c(c(\theta), e(\theta))w[\theta, \nu(\theta)] = -U_e(c(\theta), e(\theta)) \tag{59}$$

$$w_v[\theta, \nu(\theta)] = 0. \tag{60}$$

The first of these conditions is quite standard[34]. The second condition is the efficient assignment condition. It states that each agent's task assignment (locally) maximizes her shadow wage. Re-expressing (60) as $\frac{w_v \nu_\theta}{w} = 0$ and evaluating gives the efficient assignment condition:

$$0 = -\frac{1}{\varepsilon}\Big[\frac{f_\theta(\theta)}{f(\theta)} + \frac{e_\theta(\theta)}{e(\theta)} - \frac{\nu_{\theta\theta}(\theta)}{\nu_\theta(\theta)} + \frac{a_\theta[\theta,\nu(\theta)]}{a[\theta,\nu(\theta)]} + \frac{a_v[\theta,\nu(\theta)]}{a[\theta,\nu(\theta)]}\nu_\theta(\theta)\Big] + \left\{\frac{a_v[\theta,\nu(\theta)]}{a[\theta,\nu(\theta)]} + \frac{b_v(\nu(\theta))}{b(\nu(\theta))}\right\}\nu_\theta(\theta). \tag{61}$$

This is simply the continuous analogue of (13). The right hand side of this condition decomposes the proportional effect of a small adjustment $\nu_\theta(\theta)$ to a θ-talent's task assignment task upon her shadow wage. The first term is an output effect stemming from variation in the effective labor input; the second term captures the direct effect of the adjustment on the agent's productivity. The final term captures variation in the relative valuation of task output induced by the b term.[35]

SECOND BEST ALLOCATIONS We describe direct implementation in a many talent-many task economy in which both talent and task are private information; we then detail the tax decentralization. We strengthen the assumptions on a by assuming that $a_\theta > 0$ (absolute advantage) and $\frac{\partial^2 \log a[\theta,v]}{\partial\theta\partial v} > 0$ (strict log supermodularity). Mechanisms are analogous to the simple case considered previously. Each agent reports its talent θ and, conditional on this, is assigned a consumption c, task v and effort e. The combination of mechanism and truthfully reported talent imply utility and normalized shadow wage and income levels for each type:

$$\psi(\theta) = U(c(\theta), e(\theta))$$
$$\varphi(\theta) = w[\theta,\nu(\theta)]/Y^{\frac{1}{\varepsilon}}$$
$$\rho(\theta) = \varphi(\theta)e(\theta).$$

In addition, let $\omega(\theta, v) = w[\theta, v]/Y^{\varepsilon}$. An agent claiming to be type θ' must reproduce the observable income level $\rho(\theta')$. Incentive-compatibility thus requires for all θ, θ' and v':

$$U(c(\theta), e(\theta)) \geq U\left(c(\theta'), \frac{\rho(\theta')}{\omega(\theta, v')}\right). \tag{62}$$

[34] Though not normally found in assignment models which typically lack an intensive margin.

[35] Further analytical characterization is not possible without strong additional assumptions. One case in which it is occurs when $U(c,e) = u(c) - e^{1+\gamma}/1+\gamma$, $\varepsilon = 1$, b to a constant function and F to the uniform distribution. In this case, $\frac{e_\theta(\theta)}{e(\theta)} = \frac{1}{\gamma}\frac{a_\theta[\theta,v(\theta)]}{a[\theta,v(\theta)]}$ and, after substitution, (61) can be reduced to: $-\nu_{\theta\theta}(\theta) + [\left(\frac{1+\gamma}{\gamma}\right)\alpha_0]\nu(\theta)\nu_\theta(\theta) = 0$ with $v(\underline{\theta}) = \underline{v}$ and $v(\overline{\theta}) = \overline{v}$. This boundary value problem has an analytical solution in which $\nu(\theta) = D_0 \tan(\alpha_0 D_1(\theta - \underline{\theta}))$ for parameters D_j, $j = 0, 1$. Thus, an increase in comparative advantage α_0 or the Frisch elasticity of labor supply $\frac{1}{\gamma}$ leads agents to downgrade their tasks, tasks to upgrade their workers and relative wages to decline.

Let $\mathcal{U} = \{(u,e) \in \mathbb{R} \times [0,\bar{e}] : u = U(c,e) \text{ for some } c \in \mathbb{R}_+\}$ and let $C : \mathcal{U} \to \mathbb{R}_+$ be defined according to $u = U(C[u,e],e)$.

The next proposition gives simpler necessary and sufficient conditions for incentive-compatibility.

Proposition D.1. *Let (ν, e, c) be a smooth mechanism that induces a smooth task output function y. The mechanism is incentive-compatible if and only if: (i) (Monotonicity) $v_\theta \geq 0$ and $\rho_\theta \geq 0$ hold and (ii) (Envelope) the envelope conditions for utility and shadow wages hold:*

$$\psi_\theta(\theta) = -U_e\left(C[\psi(\theta), e(\theta)], e(\theta)\right) e(\theta) \frac{a_\theta[\theta, \nu(\theta)]}{a[\theta, \nu(\theta)]} \tag{63}$$

$$\varphi_\theta(\theta) = \varphi(\theta) \frac{a_\theta[\theta, \nu(\theta)]}{a[\theta, \nu(\theta)]}. \tag{64}$$

Proof. (*Necessity*). Let (ν, e, c) be a smooth incentive-compatible mechanism. Incentive compatibility implies that $w[\theta, \nu(\theta)] \geq w[\theta, \nu(\theta')]$ and $w[\theta', \nu(\theta')] \geq w[\theta', \nu(\theta)]$. Since $w[\theta, v] = b(v)\ (y(v)/Y)^{-\frac{1}{\epsilon}}\ a[\theta, v]$ and a is strictly log supermodular, it follows that if $\theta > \theta'$, then $\nu(\theta') > \nu(\theta)$. Hence, ν is increasing. To verify (64), we apply (envelope) Theorem 4.3 of Bonnans and Shapiro (2000) to the maximization $\max_{[\underline{v},\overline{v}]} U\left(c(\theta'), \rho(\theta')/\omega(\theta, v')\right)$. This requires U to be continuously differentiable, $\omega(\cdot, v)$ to be continuously differentiable and $\omega(\theta, \cdot)$ to be continuous. The first two properties hold by assumption (and the definition of ω and w, see (58)), the latter holds if y is continuous. Suppose that y is discontinuous at v and, without loss of generality assume $y(v) > y(v_n)$ for some sequence $v_n \to v$. Let $\theta_n = \nu^{-1}(v_n)$, then for n large enough, $w[\theta_n, v_n] < w[\theta_n, v]$, which is a contradiction. Then Theorem 4.3 and Remark 4.14, p.273-4 in Bonnans and Shapiro (2000) and the definition of w imply that the function φ, $\varphi(\theta) = \max_{v \in [\underline{v},\overline{v}]} w[\theta, v]$, is differentiable with $\varphi_\theta = \varphi \frac{a_\theta}{a} > 0$.

Let $\rho(\theta) = \varphi(\theta) e(\theta)$ and:

$$\Upsilon[\theta, \theta'] = U\left(c(\theta'), \frac{\rho(\theta')}{\varphi(\theta)}\right).$$

Incentive-compatibility requires that: $\Upsilon[\theta, \theta] \geq \Upsilon[\theta, \theta']$ and $\Upsilon[\theta', \theta'] \geq \Upsilon[\theta', \theta]$. Hence, $U\left(c(\theta), \frac{\rho(\theta)}{\varphi(\theta)}\right) - U\left(c(\theta'), \frac{\rho(\theta')}{\varphi(\theta)}\right) \geq U\left(c(\theta), \frac{\rho(\theta)}{\varphi(\theta')}\right) - U\left(c(\theta'), \frac{\rho(\theta')}{\varphi(\theta')}\right)$. The assumed Spence-Mirrlees condition and the increasingness of φ, then imply that ρ and c are increasing also. Additionally, since (ν, e, c) is continuous by assumption and w is continuous, Theorem 4.3 in Bonnans and Shapiro (2000) can again by applied to show that: $\psi(\theta) = \max_{\theta' \in [\underline{\theta},\overline{\theta}]} \Upsilon(\theta, \theta')$ is differentiable with:

$$\psi_\theta(\theta) = -U_e\left(C[\psi(\theta), e(\theta)], e(\theta)\right) e(\theta) \frac{\varphi_\theta(\theta)}{\varphi(\theta)} = -U_e\left(C[\psi(\theta), e(\theta)], e(\theta)\right) e(\theta) \frac{a_\theta[\theta, v(\theta)]}{a[\theta, v(\theta)]}.$$

Sufficiency. Let (ν, e, c) be a smooth mechanism satisfying the conditions in the proposition. The definition of φ, the envelope condition for wages (64) and the smoothness of ν imply the first order condition: $w_v[\theta, \nu(\theta)]\nu_\theta = 0$. The smoothness of the various functions also implies that w_v exists and is given by:

$$w_v[\theta, v] = \left\{ \frac{b_v(v)}{b(v)} - \frac{1}{\varepsilon}\frac{y_v(v)}{y(v)} + \frac{a_v[\theta, v]}{a[\theta, v']} \right\} w[\theta, v] \tag{65}$$

An agent's optimization over v and θ' is separable: regardless of the report choice of θ', it is optimal for the agent to select a task v that maximizes its wage $w[\theta, v]$. Let θ^* denote a non-decreasing measurable selection from v^{-1}. Then, using (65), the first order condition $w_v[\theta, \nu(\theta)]\nu_\theta = 0$ and log supermodularity, for $\hat{v} > v(\theta)$,

$$\begin{aligned}
w[\theta, v] - w[\theta, \nu(\theta)] &= \int_{\nu(\theta)}^{v} w_v[\theta, v']dv' \\
&= \int_{\nu(\theta)}^{\hat{v}} \left\{ \frac{b_v(v')}{b(v')} - \frac{1}{\varepsilon}\frac{y_v(v')}{y(v')} + \frac{a_v[\theta, v']}{a[\theta, v']} \right\} w[\theta, v']dv' \\
&= \int_{\nu(\theta)}^{\hat{v}} \left\{ -\frac{a_v[\theta^*(v'), v']}{a[\theta^*(v'), v']} + \frac{a_v[\theta, v']}{a[\theta, v']} \right\} w[\theta, v']dv' < 0.
\end{aligned}$$

and similarly for $\hat{v} < v(\theta)$. Consequently, the mechanism induces a θ-agent to choose the task assignment $v(\theta)$.

Let $\theta_2 > \theta_1$, then by the envelope condition for wages, for $\theta' \in [\theta_1, \theta_2]$, $\varphi(\theta') = w[\theta', \nu(\theta')] \geq w[\theta_1, \nu(\theta_1)] = \varphi(\theta_1)$. Combined with the monotonicity and concavity of U, this implies $-U_e\left(c(\theta'), e(\theta')\right)e(\theta') + U_e\left(c(\theta'), \frac{\varphi(\theta')}{\varphi(\theta_1)}e(\theta')\right)\frac{\varphi(\theta')}{\varphi(\theta_1)}e(\theta') < 0$. The envelope condition for reports and the smoothness of the mechanisms imply:

$$\Upsilon_{\hat{\theta}}[\theta, \theta]\hat{\theta}_\theta = \left\{ U_c(c(\theta), e(\theta))c_\theta(\theta) + U_e(c(\theta), e(\theta))e(\theta)\frac{\rho_\theta(\theta)}{\rho(\theta)} \right\} \hat{\theta}_\theta = 0.$$

The definitions of Υ and ρ and the preceding discussion then imply:

$$\begin{aligned}
&\Upsilon[\theta_1, \theta_2] - \Upsilon[\theta_1, \theta_1] = \int_{\theta_1}^{\theta_2} \Upsilon_{\hat{\theta}}[\theta_1, \theta']d\theta' \\
&= \int_{\theta_1}^{\theta_2} \left\{ U_c(c(\theta'), e(\theta'))c_\theta(\theta') + U_e\left(c(\theta'), \frac{\rho(\theta')}{\varphi(\theta_1)}\right)\frac{\rho_\theta(\theta')}{\varphi(\theta_1)} \right\} d\theta' \\
&= \int_{\theta_1}^{\theta_2} \left\{ -U_e\left(c(\theta'), e(\theta')\right)e(\theta') + U_e\left(c(\theta'), \frac{\rho(\theta')}{\varphi(\theta_1)}\right)\frac{w[\theta', \nu(\theta')]}{w[\theta_1, \nu(\theta_1)]}e(\theta') \right\} \frac{\rho_\theta(\theta)}{\rho(\theta)} d\theta' \leq 0.
\end{aligned}$$

A similar inequality obtains for $\theta_1 > \theta_2$ and so the mechanism induces a θ-agent to make a truthful report θ. □

It is convenient to define:

$$\zeta(\theta) = \int_{\underline{\theta}}^{\theta} C[\psi(t), e(t)]f(t)dt. \tag{66}$$

Together ψ, φ and ζ along with ξ and ν form a set of state variables for the optimal control formulation of the planning problem with private information. The envelope conditions (63) and (64) supply laws of motion for ψ and φ. Equations (53) and (66) give laws of motion for ξ and ζ:

$$\xi_\theta(\theta) = e(\theta)\varphi(\theta)f(\theta) \tag{67}$$
$$\zeta_\theta(\theta) = C[\psi(\theta), e(\theta)]f(\theta). \tag{68}$$

Finally, the definition of $\varphi(\theta)$ implies:

$$\nu_\theta(\theta) = \left(\frac{\varphi(\theta)}{b(\nu(\theta))a[\theta,\nu(\theta)]}\right)^{\varepsilon} f(\theta)a[\theta,\nu(\theta)]e(\theta), \tag{69}$$

The additional monotonicity conditions on mechanisms needed to ensure incentive-compatibility are omitted and checked ex post. The planner's problem becomes:

$$\max_{\psi,\varphi,\zeta,\nu,\xi,e} \int_{\underline{\theta}}^{\overline{\theta}} \psi(\theta)g(\theta)d\theta \tag{70}$$

subject to the laws of motion (63), (64) and (67) to (69) and the boundary constraints:

$$\begin{aligned} &\zeta(\overline{\theta}) \leq \xi(\overline{\theta})^{\frac{\varepsilon}{\varepsilon-1}} \\ &0 = \xi(\underline{\theta}) \quad 0 = \zeta(\underline{\theta}) \quad \underline{v} = v(\underline{\theta}) \quad v(\overline{\theta}) = \bar{v}. \end{aligned}$$

In this problem there is one control (e) and five states $(\psi, \varphi, \zeta, \nu, \xi)$. Routine manipulation of the first order and co-state equations yields the following expression for the optimal effort-consumption wedge:

$$\begin{aligned} -w[\theta,\nu^*(\theta)]\frac{U_c^*(\theta)}{U_e^*(\theta)} - 1 = &-\mathcal{I}^*(\theta)\left[\frac{p_\theta^\varphi(\theta)}{p^\varphi(\theta)} + \frac{a_\theta[\theta,\nu^*(\theta)]}{a[\theta,\nu^*(\theta)]}\right]\frac{p^{\varphi*}(\theta)}{p^{\zeta*}f(\theta)}\left(-\frac{U_c^*(\theta)}{U_e^*(\theta)}\right) \\ &+\mathcal{N}^*(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{\varphi_\theta^*(\theta)}{\varphi^*(\theta)}\int_\theta^\infty \left(1 - \frac{g(t)U_c^*(t)}{p^\zeta f(t)}\right)\frac{U_c^*(\theta)}{U_c^*(t)}\mathcal{M}^*(\theta,t)\frac{f(t)}{1-F(\theta)}dt, \end{aligned} \tag{71}$$

where $U_l^*(\theta) := U_l(c^*(\theta), e^*(\theta))$ $l \in \{c, e\}$ and similarly for $U_{lm}^*(\theta)$, $\mathcal{I}^*(\theta) := -\frac{1}{\varepsilon}\frac{w[\theta,\nu^*(\theta)]}{e^*(\theta)}$ is the elasticity of the θ-talent wage with respect to effort holding the task allocation fixed, $N_u^* := \{-\frac{U_{ec}^*}{U_e^*} + \frac{U_{ee}^*}{U_e^*}\}e^* + 1$ is $(1+\mathcal{E}_u)/\mathcal{E}_c$, where $\mathcal{E}_u$ and $\mathcal{E}_c$ are, respectively, the uncompensated and compensated labor supply elasticities, $\mathcal{M}^*(\theta,t) = \exp\left\{-\int_\theta^t \frac{e^*(s)U_{ce}^*(s)}{\varphi*(s)U_c^*(s)}\right\}$, $p^{\zeta*} = E\left[\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{U_c^*(t'')}f(t'')dt''\right]^{-1}$ is the optimal shadow resource multiplier and $p^{\varphi*}$ is the optimal co-state on the shadow wage φ. In the decentralization of the optimal allocation, this wedge determines the marginal income tax. The second term on the right hand side of (71) is familiar from standard analyses of optimal taxes (see, inter alia, Saez (2001) and Salanié (2011)). The main modification is the incorporation of endogenous wage growth

$\varphi_\theta^*(\theta)/\varphi^*(\theta)$ (which in typical Mirrlees models is given simply by $1/\theta$). The first term on the right hand side of (71) is entirely new. This term captures the net benefit of adjustments to the θ-agent's shadow wage stemming from a small effort change. Absent private information, the law of motion for shadow wages is consistent with the efficient allocation of tasks and attains a zero co-state. In the second best, it is desirable to distort the shadow wage schedule to relax incentive constraints. The second best co-state function $p^{\varphi*}$ reflects this.

IMPLEMENTATION VIA TAXES Under certain conditions the solution to the planner's problem can be established as a market equilibrium in an economy with a smooth income tax function $T : \mathbb{R} \to \mathbb{R}$.

WORKERS A θ-talent worker chooses her task, consumption and effort to solve:

$$\sup U(c, e) \tag{WP$_\theta$}$$

subject to:

$$c \leq w[\theta, v]e - T[w[\theta, v]e], \tag{72}$$

and $w[\theta, v]e \in \mathcal{Q}$, where $\mathcal{Q}$ is an interval of admissible incomes, $w : [\underline{\theta}, \overline{\theta}] \times [\underline{v}, \overline{v}] \to \mathbb{R}_+$ is a measurable wage function and $T : \mathbb{R}_+ \to \mathbb{R}_+$ is a twice differentiable tax function with slope less than one. Since T has a slope less than one, a worker's after tax income is increasing in pre-tax income and, consequently, the worker will choose the occupation that maximizes her wage.[36] Enforcement of incomes within the admissible set $\mathcal{Q}$ may be via the income tax or some unmodeled enforcement technology. Consistent with our treatment of direct mechanisms and without loss of generality, we focus on symmetric equilibria in which (almost all) agents of type θ make the same deterministic choice. For a given wage function w, let $c : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}_+$, $e : [\underline{\theta}, \overline{\theta}] \to [0, \overline{e}]$ and $\nu : [\underline{\theta}, \overline{\theta}] \to [\underline{v}, \overline{v}]$ denote a measurable selection from the policy correspondence describing solutions to the family of problems (WP$_\theta$).

FIRMS A representative firm combines tasks to create final goods. The firm pays piece rates $p : [\underline{v}, \overline{v}] \to \mathbb{R}_+$ to workers for the tasks they perform. The firm chooses task outputs $y \in \mathcal{Y}$ to maximize profits:

$$\sup_{\mathcal{Y}} \left[\int_{\underline{v}}^{\overline{v}} b(v) y(v)^{\frac{\varepsilon-1}{\varepsilon}} dv \right]^{\frac{\varepsilon}{\varepsilon-1}} - \int_{\underline{v}}^{\overline{v}} p(v) y(v) dv. \tag{73}$$

MARKET CLEARING AND EQUILIBRIUM The implied wage of a θ-talent working in task v is $w[\theta, v] = p(v)a[\theta, v]$. Goods market clearing requires that output of final goods

[36] We assume that even workers who exert no effort still associate with the occupation that offers them the greatest wage.

weakly exceeds total consumption plus the total net transfer to the government:

$$\int_{\underline{\theta}}^{\overline{\theta}} c(\theta)f(\theta)d\theta + \int_{\underline{\theta}}^{\overline{\theta}} T[w[\theta,\nu(\theta)]e(\theta)] \leq \left[\int_{\underline{v}}^{\overline{v}} b(v)y(v)^{\frac{\varepsilon-1}{\varepsilon}}dv\right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{74}$$

Labor market clearing requires that for each θ:

$$\int_{\underline{v}}^{v(\theta)} y(v')dv' = \int_{\underline{\theta}}^{\theta} a[\theta',\nu(\theta')]e(\theta')f(\theta')d\theta'. \tag{75}$$

Definition 1. An equilibrium of an assignment market economy with tax function T and admissible income set $\mathcal{Q}$ is a tuple of measurable, non-negative functions $\{y,c,e,v,p,w\}$ such that (i) y solves (73), (ii) (c,e,v) solves (WP_θ), (iii) $\{y,c,e,v\}$ satisfy the final good and labor market clearing conditions (74) and (75) and (iv) $w[\theta,v] = p(v)a[\theta,v]$.

We now show that if the optimal mechanism admits an increasing shadow income function, then it can be implemented as an equilibrium of a market economy with taxes and admissible income set $\rho^*(\underline{\theta}),\rho^*(\overline{\theta})]$.

Proposition D.2. *Let (c^*,e^*,nu^*,y^*) be a smooth solution to the planner's problem with private information over types and tasks satisfying with strictly increasing shadow income function ρ^*. Then (c^*,e^*,nu^*,y^*) is an equilibrium allocation of a market economy with admissible income set: $[\rho^*(\underline{\theta}),\rho^*(\overline{\theta})]$ and smooth taxes on income $T[q] = q - c^*(\rho^{*-1}(q))$.*

Remark 4. Incentive-compatibility in the private types and tasks setting only ensures that ρ is non-decreasing. The strictness ensure that types are measurable with respect to income and, hence, consumption is measurable with respect to income. Alternatively, one can assume the latter directly as is sometimes done, e.g. Kocherlakota (2005).

Proof. Given the allocation (c^*,e^*,ν^*,y^*) and implied output of final good Y^*, define the price function p^* as:

$$p^*(v) = b(v)\left(\frac{y^*(v)}{Y^*}\right)^{-\frac{1}{\varepsilon}}, \quad v \in [\underline{v},\overline{v}]. \tag{76}$$

Since ρ^* is strictly increasing by assumption, an agent's talent θ is measurable with respect to her (shadow) income q. $T : [\rho^*(\underline{\theta}),\rho^*(\overline{\theta})] \to \mathbb{R}$ can then be defined according to:

$$T[q] = q - c^*(\rho^{*-1}(q)), \quad q \in [\rho(\underline{\theta}),\rho(\overline{\theta})].$$

By construction the allocation, prices p and taxes T satisfy market clearing and the agent's budget constraint. It remains to check that the allocation (c^*,e^*,v^*,y^*) is optimal for firms and agents given (p,T). The firm's problem is strictly concave and, by (76), y^* satisfies the firm's first order conditions given p. Consequently, y^* is optimal for the firm.

It is convenient to rewrite the agent's market problem as:

$$\max_{q,v} U\left(q - T[q], \frac{q}{w[\theta,v]}\right), \tag{77}$$

where q gives the agent's labor income and market wages are $w[\theta, v] = p^*(v)a[\theta, v]$. Since eacg $U(c, \cdot)$ is decreasing, the household's utility is increasing in $w[\theta, v]$ *for all* q. It follows that the optimization in (77) can be decomposed into two steps. In the first v is chosen to solve $w^*(\theta) = \sup_v w[\theta, v]$. This maximization is independent of the choice of q (and of the tax function T). In the second step q is chosen to maximize $U\left(q - T[q], \frac{q}{w^*(\theta)}\right)$.

Now, the schedule of market wages for talent θ, $w[\theta, \cdot]$, is exactly equal to the schedule of shadow wages in the incentive problem and since $\nu^*(\theta)$ was optimal for the agent in the latter, it remains optimal for the agent in the latter. Thus, regardless of her income (or effort) choice, a θ-talent will select the task $\nu^*(\theta)$. Given this choice the schedule of consumption and efforts available to the agent via a choice of $q \in [\rho^*(\underline{\theta}), \rho^*(\overline{\theta})]$ is:

$$\left(q - T[q], \frac{q}{w[\theta, \nu^*(\theta)]}\right) = \left(c^*(\rho^{*-1}(q)), \frac{w[\rho^{*-1}(q), \nu^*(\rho^{*-1}(q))]}{w[\theta, \nu^*(\theta)]}\right)$$

This is exactly the same set as in the mechanism and so the optimal income choice for a θ-talent is $\rho^*(\theta)$ or, equivalently, $(c^*(\theta), e^*(\theta))$. Thus, the allocation is implemented in the market economy. $\square$

The equilibrium is characterized by a group of first order conditions. First, the firm's market problem and the market clearing condition for tasks imply:

$$p(v) = b(v)\left(\frac{Y}{y(v)}\right)^{\frac{1}{\varepsilon}} = b(v)\left(\frac{Y}{\frac{f(\theta)a[\theta, v(\theta)]e(\theta)}{v'(\theta)}}\right)^{\frac{1}{\varepsilon}}. \tag{78}$$

It is immediate from the agent's problem in this economy that for $\theta \in (\underline{\theta}, \overline{\theta})$, optimal marginal income taxes satisfy:

$$\frac{T_q[\rho^*(\theta)]}{1 - T_q[\rho^*(\theta)]} = -w[\theta, \nu^*(\theta)]\frac{U_c(c^*(\theta), e^*(\theta))}{U_e(c^*(\theta), e^*(\theta))} - 1.$$

E Dataset

Our many data source is the Current Population Survey (CPS) administered by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics. We focus on the March release of the survey.[37] Data is available continuously from 1968 to 2012. On average each year of data contains about 150,000 observations, from the year 2001 the sample size has increase to approximately 200,000. The CPS contains detailed information on demographics and work characteristic of each individual. For additional details on the CPS refer to Heathcote, Perri, and Violante (2010) and Acemoğlu and Autor (2011). The CPS data includes a self reported estimate of hours worked from 1976 onwards. This question as well as

[37] Data is taken from: Miriam King, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

question on income are for the previous calendar year. Hence our sample will be covering the years 1975 to 2011 (interviews from 1976 to 2012). In the body of the paper we group observations in two groups. We will call "the 70s" observations relating to years 1975-1979 (i.e interviewed in years 1976-1980), we will call the "00s" observation relating to years 2000-2011 (interviews in 2001-2012). The model analyzed is highly stylized. In

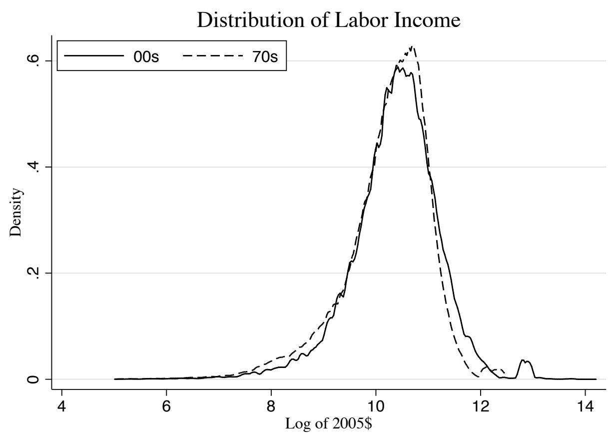


Figure 12: Smoothed Distribution of Log-Labor Income Over Time. Source: CPS.

order to make data and model compatible (and to reduce the likelihood of measurement error) we further restrict our sample. We drop individuals for which income, age, sex, education, sector, occupation data is not reported. We consider individuals of working age: between the ages of 25 and 65 included (dropping 301,144 observations). Given the changing nature of female labor participation over the time period considered, we only focus on male workers. We drop individuals with no formal education (7,180 observations). Our objective is to link individual type and productivity with the task in which he is employed. Given this we drop individuals unemployed (dropping an additional 137,213 observations). Following (Heathcote, Perri, and Violante 2010) we also drop underemployed individuals: with the estimates of average hours worked in the previous year we drop individuals who work less than a total of 250 per year (dropping an additional 35,663 observations). Our final sample comprises of 1,228,933 individual/year observations. All variables will be weighted with the provided weights and dollar denominated variables deflated using CPI to 2005$.[38] In Figure 12 we display the evolution of the distribution of Log labor income between the "70s" and the "00s". The main feature that emerges is the widening of the distribution in the "00s" relative to the "70s". The "00s" are also characterized by a much fatter tail.

[38]CPI for all urban consumers, all goods.

F Robustness

F.1 Sample Selection

In this section we explore the impact of the sample selection on the estimated values of $b(v)$. As a first case we replicate the sample selection of Acemoğlu and Autor (2011). In their paper they consider all observation for individuals who worked at least one week in the previous year and that are of at least 16 years of age. The estimated values of $b(v)$ are in Figure 13a. As it can be seen the polarization behavior is still apparent. However for low v occupations we observe little change between the two decades. In Figure 13b we display the effect on $b(v)$ when looking at the entire CPS sample. In this case the values are not dissimilar to the previous case. This is not a surprise given the fact that the sample selection of (Acemoğlu and Autor 2011) is not restrictive

We next analyze separately the effect of two of hours sample selection: gender and minimum hour worked. In Figure 13c we only consider males (without applying any other sample restriction). In Figure 13d we only impose as a sample selection criterion the requirement on hours worked per year. In both case we observe that the polarization phenomenon is robust.

F.2 Semi-parametric Analysis of a

A key parameter in the analysis is the degree of comparative advantage. We found the degree of comparative advantage to be positive and increasing over time. In this section we follow a Semi-parametric approach to determine the sign of the interaction term and it's evolution over time.[39] From the continuous talent version of equation (22) we have:

$$\ln w(\theta^*(v), v)) - \ln b(v) + \ln y(v) - \ln Y = \alpha_0 + \alpha_1 \theta^*(v) + \alpha_2 v + \alpha_3 \theta^*(v) \cdot v. \tag{79}$$

Integrating between v_{m-1} and v_m and dividing by Δv_m we have

$$\frac{1}{\Delta v_m} \int_{\mathcal{V}_m} [\ln w(\theta^*(v), v) + \ln y(v)] dv - \ln b_m - \ln Y =$$
$$\alpha_0 + \alpha_2 \overline{v}_m + \frac{1}{\Delta v_m} \int_{\mathcal{V}_m} [\alpha_1 \theta^*(v) + \alpha_3 \theta^*(v) \cdot v] dv.$$

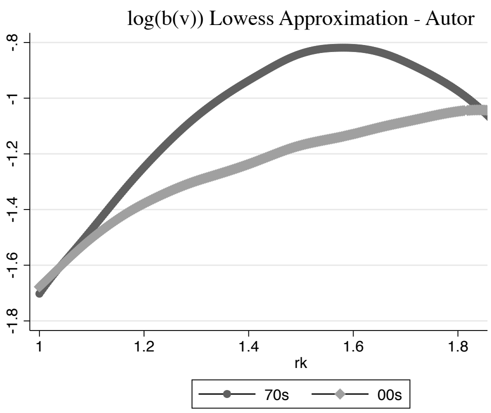
substituting $\theta^*(v) = \overline{\theta}_m$ for all v in $\mathcal{V}_m$ and let the average log-wage rate in occupation m be defined as:

$$E[\ln w | V_m] = \frac{1}{\Delta v_m} \int_{\mathcal{V}_m} \ln w^*(\theta^*(v), v) dv;$$

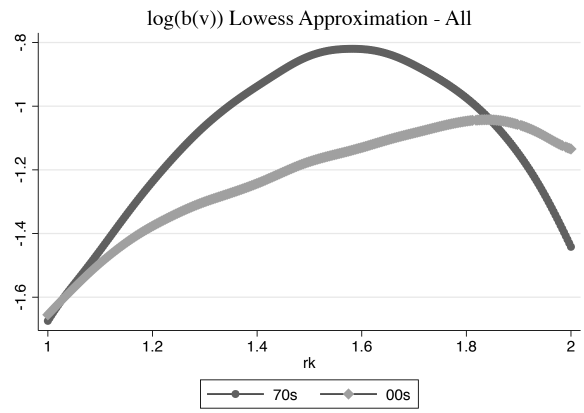
define $f_m = E[\ln w | V_m] - \ln b_m - \ln Y$. Also let

$$g(v_m) = \alpha_2 \overline{v}_m - \frac{1}{\Delta v_m} \int_{\mathcal{V}_m} [\ln y(v)] dv.$$

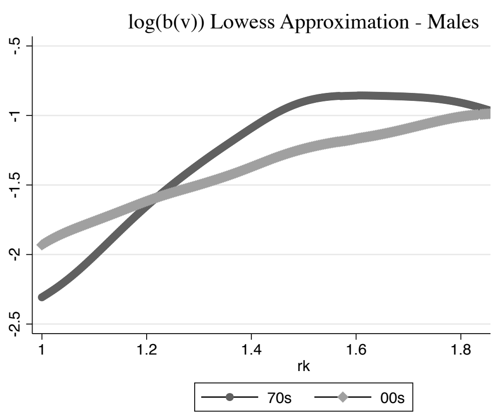
[39] Also refer to Sperlich, Tjøstheim, and Yang (2002) Yatchew (2003) Linton and Nielsen (1995).



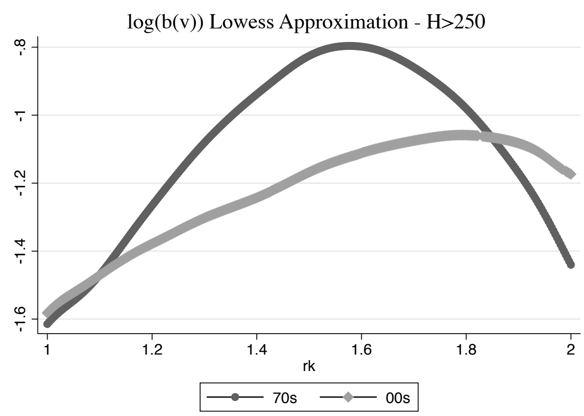
(a) Acemoğlu and Autor (2011) Sample.



(b) Entire Sample.



(c) Sex.



(d) Hours.

Figure 13: Estimated values of $\log(b(v))$ for different sample specifications.

We then have:

$$f_m = \alpha_1 \overline{\theta}_m + \alpha_3 \overline{\theta}_m \cdot \overline{v}_m + g(v_m), \qquad \forall m = 1, \dots, M. \tag{80}$$

To estimate Equation (80) we approximate $g(v_m)$ with a first order fractional polynomial: i.e we approximate $g(v_k)$ with a polynomial comprised of v_m, $1/v_m$ and $\log(v_m)$ and a constant. Results are in Table 4. We observe that as in our previous specification the interaction term is positive, denoting the presence of comparative advantage and growing over time.

G Empirical Talent Types and Observables

It is to connect the imputed talent index with observables available in the data. To do this for each occupation $\mathcal{V}_m$ and each decade t, we compute: the average number of years of

Table 4: Estimation of Equation (80).

Decade	Interaction Term: α_3
70s	4.148
00s	4.706

education completed by each worker: $edu_{t,m}$; and the attained level of experience $exp_{t,m}$. Experience is equal to the age minus 18 for non college graduates and equal to the age minus 22 for college graduates. We then evaluate the following linear relationship:

$$\bar{\theta}_{m,t} = \eta_0 + \eta_1 \cdot edu_{t,m} + \eta_2 \cdot exp_{t,m} + \eta_3 \cdot edu_{t,m} \cdot exp_{t,m} \tag{81}$$

our estimates are in Table 5.

Table 5: Estimation of Equation (81). Robust Regression. Standard Errors in Parenthesis.

Decade	Parameters			
	η_0	η_1	η_2	η_3
70s	0.75	0.12	-0.0006	-0.15
	(0.26)	(0.024)	(0.006)	(0.007)
00s	0.62	0.13	-0.004	-0.011
	(0.3)	(0.031)	(0.006)	(0.007)

From Table 5 we observe that average education attainment is significant in explaining one's type (as opposed to experience). However for both decades the R-squared are low and diminishing over point to education being a poor proxy for productivity type.