

# Technical Change, Wage Inequality and Taxes: ONLINE APPENDIX

By LAURENCE ALES AND MUSAB KURNAZ AND CHRISTOPHER SLEET \*

## I. PROOFS FROM SECTION III

### PROOF OF PROPOSITION 2:

Let  $\{T, l, \{c_k, e_k, \lambda_k\}_{k=1}^K\}$  and  $\{\omega_k\}_{k=1}^K$  denote a tax equilibrium at spending level  $G$ . Since workers of a given type select the highest possible wage, it follows that for each  $k$  there is a  $w_k < \infty$  such that for every  $v \in \text{Supp } \Lambda_k$ ,  $\omega(v)a_k(v) = w_k$  and for  $v \notin \text{Supp } \Lambda_k$ ,  $\omega(v)a_k(v) \leq w_k$ . Firm optimality implies that for almost every  $v \in [\underline{v}, \bar{v}]$ :  $\omega(v) = b(v) (Y/l(v))^{1/\varepsilon}$ . If  $v \in [\underline{v}, \bar{v}] \setminus \cup_{k=1}^K \text{Supp}(\Lambda_k)$ , then in equilibrium  $l(v) = \sum_{k=1}^K \lambda_k(v)a_k(v)e_k = 0$ . Since  $\omega$  is finite, almost all tasks must be performed. Without loss of generality, we select versions of tax equilibria in which all tasks are performed. For all  $v$  and  $v'$  in  $\text{Supp } \Lambda_k$  with  $v > v'$ ,

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

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

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

Hence, from the labor market clearing condition:

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

\* Ales: Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. email: ales@cmu.edu. Kurnaz: Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. email: mkurnaz@andrew.cmu.edu. Sleet: Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. email: csleet@andrew.cmu.edu.

And so:

$$(I.2) \quad w_k = B_k(\tilde{v}_{k-1}, \tilde{v}_k) \left( \frac{Y}{\pi_k e_k} \right)^{\frac{1}{\varepsilon}},$$

where  $B_k(\tilde{v}_{k-1}, \tilde{v}_k) := \left[ \int_{\tilde{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv \right]^{\frac{1}{\varepsilon}}$ . Substituting (I.1) into (I.2) gives for  $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$ ,  $\lambda_k(v) = \frac{b(v)^\varepsilon a_k(v)^{\varepsilon-1}}{B_k(\tilde{v}_{k-1}, \tilde{v}_k)^\varepsilon} \pi_k$ . In addition, for  $v < \tilde{v}_{k-1}$  and  $v > \tilde{v}_k$ ,  $\lambda_k(v) = 0$ . Now for  $v \in (\tilde{v}_{k-1}, \tilde{v}_k)$ ,  $w_{k+1} > \omega(v) a_{k+1}(v) = b(v) (Y/l(v))^{\frac{1}{\varepsilon}} a_{k+1}(v)$  and  $w_k = \omega(v) a_k(v) = b(v) (Y/y(v))^{\frac{1}{\varepsilon}} a_k(v)$ . Hence:  $\frac{w_{k+1}}{w_k} > \frac{a_{k+1}(v)}{a_k(v)}$ . Conversely, for  $v \in (\tilde{v}_k, \tilde{v}_{k+1})$ ,  $w_{k+1} = \omega(v) a_{k+1}(v) = b(v) (Y/y(v))^{\frac{1}{\varepsilon}} a_{k+1}(v)$  and  $w_k > \omega(v) a_k(v) = b(v) (Y/y(v))^{\frac{1}{\varepsilon}} a_k(v)$ . Consequently,  $\frac{w_{k+1}}{w_k} < \frac{a_{k+1}(v)}{a_k(v)}$ . Then, 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 (I.2) gives the desired expression in the proposition.

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

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

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

PROOF OF LEMMA 1:

Totally differentiating:  $\frac{a_{j+1}}{a_j}(\tilde{v}_j) = \frac{B_{j+1}(\tilde{v}_j, \tilde{v}_{j+1})}{B_j(\tilde{v}_{j-1}, \tilde{v}_j)} \left( \frac{e_j \pi_j}{e_{j+1} \pi_{j+1}} \right)^{\frac{1}{\varepsilon}}$ , with respect to  $\tilde{v}_{j-1}$  and  $\tilde{v}_j$  holding  $e_j/e_{j+1}$  fixed gives:

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j-1}} = \frac{-\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}}}{\frac{\partial \log a_{j+1}/a_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}} \frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j}}.$$

Let  $L_{j-1} = -\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}} - \frac{\partial \log B_j}{\partial \log \tilde{v}_j} \frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j-1}}$ . It follows that:

$$L_{j-1} = -\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}} \left\{ 1 - \frac{\frac{\partial \log B_j}{\partial \log \tilde{v}_j}}{\frac{\partial \log a_{j+1}/a_j}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} + L_j} \right\}.$$

Thus, if  $L_j > 0$ , then  $L_{j-1} > 0$ . For  $j = K - 1$ ,  $\tilde{v}_{j+1} = \tilde{v}_K = \bar{v}$  and  $\frac{\partial \log \tilde{v}_{j+1}}{\partial \log \tilde{v}_j} = \frac{\partial \log \tilde{v}_K}{\partial \log \tilde{v}_{K-1}} = 0$ . Hence,  $L_{K-1} = -\frac{\partial \log B_K}{\partial \log \tilde{v}_{K-1}} > 0$ . It follows by induction that for all  $j \in \{k, \dots, K - 2\}$ ,  $L_j > 0$  and, hence,

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j-1}} = \frac{-\frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}}}{\frac{\partial \log a_{j+1}/a_j}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} + L_j} > 0.$$

Similarly, for all  $j \in \{1, \dots, k - 1\}$ ,

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}} = \frac{\frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}}}{\frac{\partial \log a_{j+1}/a_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + \frac{\partial \log B_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_j}{\partial \log \tilde{v}_{j-1}} \frac{\partial \log \tilde{v}_{j-1}}{\partial \log \tilde{v}_j}}.$$

Let  $M_{j+1} = \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} \frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}}$ . It follows that:

$$M_{j+1} = \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}} \left\{ 1 - \frac{\frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j}}{\frac{\partial \log a_{j+1}/a_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + M_j} \right\}.$$

Thus, if  $M_j > 0$ , then  $M_{j+1} > 0$ . For  $j = 1$ ,  $\tilde{v}_{j-1} = \tilde{v}_0 = \underline{v}$  and  $\frac{\partial \log \tilde{v}_{j-1}}{\partial \log \tilde{v}_j} = \frac{\partial \log \tilde{v}_0}{\partial \log \tilde{v}_1} = 0$ . Hence,  $M_1 = \frac{\partial \log B_1}{\partial \log \tilde{v}_1} > 0$ . It follows by induction that for all  $j \in \{1, \dots, k - 1\}$ ,  $M_j > 0$  and, hence,

$$\frac{\partial \log \tilde{v}_j}{\partial \log \tilde{v}_{j+1}} = \frac{\frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_{j+1}}}{\frac{\partial \log a_{j+1}/a_j}{\partial \log \tilde{v}_j} - \frac{\partial \log B_{j+1}}{\partial \log \tilde{v}_j} + M_j} > 0.$$

Next, taking logs and totally differentiating  $\frac{a_{k+1}}{a_k}(\tilde{v}_k) = \frac{B_{k+1}(\tilde{v}_k, \tilde{v}_{k+1})}{B_j(\tilde{v}_{k-1}, \tilde{v}_k)} \left( \frac{e_k \pi_k}{e_{k+1} \pi_{k+1}} \right)^{\frac{1}{\varepsilon}}$  with respect to  $\log e_k$  gives:

$$\frac{\partial \log \tilde{v}_k}{\partial \log e_k} = \frac{1}{\varepsilon} \left\{ \frac{1}{\frac{\partial \log a_{k+1}/a_k}{\partial \log \tilde{v}_k} + M_k + L_k} \right\} > 0.$$

Similarly, taking logs and totally differentiating  $\frac{a_k}{a_{k-1}}(\tilde{v}_{k-1}) = \frac{B_k(\tilde{v}_{k-1}, \tilde{v}_k)}{B_j(\tilde{v}_{k-2}, \tilde{v}_{k-1})} \left( \frac{e_{k-1} \pi_{k-1}}{e_k \pi_k} \right)^{\frac{1}{\varepsilon}}$

with respect to  $\log e_k$  gives:

$$\frac{\partial \log \tilde{v}_{k-1}}{\partial \log e_k} = -\frac{1}{\varepsilon} \left\{ \frac{1}{\frac{\partial \log a_k / a_{k-1}}{\partial \log \tilde{v}_{k-1}} + M_{k-1} + L_{k-1}} \right\} < 0.$$

The implications for the elasticities  $\phi_{k,j}$  described in the lemma then follow immediately from Equation (17).

Complete characterization of the sensitivity of task thresholds to the perturbation of a given talent's effort is provided in the next lemma.

LEMMA I.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:

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

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

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

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,  $i = K-2, \dots, 1$ ,

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

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  $m_K = 1$ .

PROOF:

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

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

with  $\tilde{v}_0 = \underline{v}$  and  $\tilde{v}_K = \bar{v}$ . 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:

$$\Gamma \Delta \mathbf{v}_k = \mathbf{E}_k,$$

where:

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

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

$$\Delta \mathbf{v}_k = \Gamma^{-1} \mathbf{E}_k.$$

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  $\Gamma^{-1}$ . Since  $\Gamma$  is a tridiagonal matrix, explicit formulas for its inverse are available. Applying these formulas gives the expression in the text.

## II. CONTINUOUS TALENT-CONTINUOUS TASK MODEL

In this appendix, we briefly describe the continuous talent (and continuous task) assignment model and its optimal control formulation. In our quantitative work, we treat the data as a discrete approximation to this model and solve it using the open-source numerical optimal control software GPOPS-II. Code is available on request. Workers are now distributed across an interval of talents  $k \in [k, \bar{k}]$  according to a distribution function  $\Pi : [k, \bar{k}] \rightarrow [0, 1]$  with strictly positive and continuously differentiable density  $\pi$ . As before there is a continuum of tasks ranked by complexity  $v \in [v, \bar{v}]$ . The productivity of talent-task combinations is given by a function  $a : [k, \bar{k}] \times [v, \bar{v}] \rightarrow \mathbb{R}_{++}$  satisfying the following assumption.

**ASSUMPTION 1:** (i)  $a$  is twice continuously differentiable on the interior of  $[k, \bar{k}] \times [v, \bar{v}]$  with first derivatives  $a_i$ ,  $i \in \{k, v\}$  and second derivatives  $a_{ij}$ ,  $i, j \in \{k, v\}$ . (ii) (strict absolute advantage)  $a_k > 0$ , (iii) (strict comparative advantage, log supermodularity)  $\frac{\partial^2 \log a}{\partial k \partial v} > 0$ .

Otherwise technologies and preferences are as in the main text. An allocation is a triple of measurable functions  $c : [k, \bar{k}] \rightarrow \mathbb{R}_+$ ,  $v : [k, \bar{k}] \rightarrow [v, \bar{v}]$  and  $e : [k, \bar{k}] \rightarrow \mathbb{R}_+$  describing the consumption, task and effort assignments of each

talent type.<sup>1</sup> As before, task output is linear in labor input. The task output density  $y : [\underline{v}, \bar{v}] \rightarrow \mathbb{R}_+$  satisfies for all  $k$ ,

$$(II.1) \quad \int_{\underline{v}}^{v(k)} y(v) dv = \int_{\underline{k}}^k a[k', v(k')] e(k') \pi(k') dk'.$$

If  $v$  is differentiable with derivative  $v_k$ , then (II.1) can be re-expressed as, for all  $k$ :

$$(II.2) \quad y(v(k)) = a[k, v(k)] e(k) \frac{\pi(k)}{v_k(k)},$$

Heuristically, the numerator is total output of type  $k$ , while the denominator gives the tasks over which the type  $k$  workers are "spread". The shadow wage is given by:

$$w[k, v] = b(v) \left( \frac{y(v)}{Y} \right)^{-\frac{1}{\epsilon}} a[k, v].$$

We restrict planners and policymakers to smooth allocations and mechanisms. This permits the application of optimal control techniques.

#### OPTIMAL CONTROL FORMULATION OF GOVERNMENT'S PROBLEM

We formulate the government's problem as a mechanism design problem and recover optimal taxes from this. Mechanisms are analogous to those considered previously. Each worker reports its talent  $k$  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:

$$\begin{aligned} \psi(k) &= U(c(k), e(k)) \\ \varphi(k) &= w[k, v(k)] / Y^{\frac{1}{\epsilon}} \\ \rho(k) &= \varphi(k) e(k). \end{aligned}$$

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

$$(II.3) \quad U(c(k), e(k)) \geq U \left( c(k'), \frac{\rho(k')}{\omega(k, v')} \right).$$

<sup>1</sup>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.

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  $C : \mathcal{U} \rightarrow \mathbb{R}_+$  be defined according to  $u = U(C[u, e], e)$ . The next proposition gives simpler necessary and sufficient conditions for incentive-compatibility.

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

$$(II.4) \quad \psi_k(k) = -U_e(C[\psi(k), e(k)], e(k)) e(k) \frac{a_k[k, v(k)]}{a[k, v(k)]}$$

$$(II.5) \quad \varphi_k(k) = \varphi(k) \frac{a_k[k, v(k)]}{a[k, v(k)]}.$$

**PROOF:**

(Necessity). Let  $(v, e, c)$  be a smooth incentive-compatible mechanism. Incentive compatibility implies that  $w[k, v(k)] \geq w[k, v(k')]$  and  $w[k', v(k')] \geq w[k', v(k)]$ . Since  $w[k, v] = b(v) (y(v)/Y)^{-\frac{1}{\varepsilon}} a[k, v]$  and  $a$  is strictly log supermodular, it follows that if  $k > k'$ , then  $v(k') > v(k)$ . Hence,  $v$  is increasing. To verify (II.5), we apply (envelope) Theorem 4.3 of Bonnans and Shapiro (2000) to:  $\max_{v \in [\underline{v}, \bar{v}]} U(c(k'), \rho(k')/\omega(k, v'))$ . This requires  $U$  to be continuously differentiable,  $\omega(\cdot, v)$  to be continuously differentiable and  $\omega(k, \cdot)$  to be continuous. The first two properties hold by assumption (and the definition of  $\omega$  and  $w$ , see II), 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  $v_n \rightarrow v$ . Let  $k_n = v^{-1}(v_n)$ , then for  $n$  large enough,  $w[k_n, v_n] < w[k_n, v]$ , 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  $\varphi$ ,  $\varphi(k) = \max_{v \in [\underline{v}, \bar{v}]} w[k, v]$ , is differentiable with  $\varphi_k = \varphi \frac{a_k}{a} > 0$ .

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

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

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

$$\psi_k(k) = -U_e(C[\psi(k), e(k)], e(k)) e(k) \frac{\varphi_k(k)}{\varphi(k)} = -U_e(C[\psi(k), e(k)], e(k)) e(k) \frac{a_k[k, v(k)]}{a[k, v(k)]}.$$

*Sufficiency.* Let  $(v, e, c)$  be a smooth mechanism satisfying the conditions in the proposition. The definition of  $\varphi$ , the envelope condition for wages (II.5) and the smoothness of  $v$  imply the first order condition:  $w_v[k, v(k)]v_k = 0$ . The smoothness of the various functions also implies that  $w_v$  exists and is given by:

$$(II.6) \quad w_v[k, v] = \left\{ \frac{b_v(v)}{b(v)} - \frac{1}{\varepsilon} \frac{y_v(v)}{y(v)} + \frac{a_v[k, v]}{a[k, v']} \right\} w[k, v]$$

An worker's optimization over  $v$  and  $k'$  is separable: regardless of the report choice of  $k'$ , it is optimal for the worker to select a task  $v$  that maximizes its wage  $w[k, v]$ . Let  $k^*$  denote a non-decreasing measurable selection from  $v^{-1}$ . Then, using (II.6), the first order condition  $w_v[k, v(k)]v_k = 0$  and log supermodularity, for  $\hat{v} > v(k)$ ,

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

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

Let  $k_2 > k_1$ , then by the envelope condition for wages, for  $k' \in [k_1, k_2]$ ,  $\varphi(k') = w[k', v(k')] \geq w[k_1, v(k_1)] = \varphi(k_1)$ . Combined with the monotonicity and concavity of  $U$ , this implies  $-U_e(c(k'), e(k'))e(k') + U_e\left(c(k'), \frac{\varphi(k')}{\varphi(k_1)}e(k')\right)\frac{\varphi(k')}{\varphi(k_1)}e(k') < 0$ . The envelope condition for reports and the smoothness of the mechanisms imply:

$$Y_{\hat{k}}[k, k]\hat{k}_k = \left\{ U_c(c(k), e(k))c_k(k) + U_e(c(k), e(k))e(k)\frac{\rho_k(k)}{\rho(k)} \right\} \hat{k}_k = 0.$$

The definitions of  $Y$  and  $\rho$  and the preceding discussion then imply:

$$\begin{aligned} Y[k_1, k_2] - Y[k_1, k_1] &= \int_{k_1}^{k_2} Y_{\hat{k}}[k_1, k'] dk' \\ &= \int_{k_1}^{k_2} \left\{ U_c(c(k'), e(k'))c_k(k') + U_e\left(c(k'), \frac{\rho(k')}{\rho(k_1)}\right)\frac{\rho_k(k')}{\rho(k_1)} \right\} dk' \\ &= \int_{k_1}^{k_2} \left\{ -U_e(c(k'), e(k'))e(k') + U_e\left(c(k'), \frac{\rho(k')}{\rho(k_1)}\right)\frac{w[k', v(k')]}{w[k_1, v(k_1)]}e(k') \right\} \frac{\rho_k(k)}{\rho(k)} dk' \leq 0. \end{aligned}$$



A similar inequality obtains for  $k_1 > k_2$  and so the mechanism induces a  $k$ -worker to make a truthful report  $k$ .

It is convenient to define:

$$(II.7) \quad \bar{\xi}(k) := \int_{\underline{k}}^k \left( \frac{\pi(k')a[k', \nu(k')]e(k')}{\nu_k(k')} \right)^{\frac{\varepsilon-1}{\varepsilon}} b(\nu(k'))\nu_k(k')dk'$$

and

$$(II.8) \quad \zeta(k) = \int_{\underline{k}}^k C[\psi(k'), e(k')] \pi(k') dk'.$$

Together  $\psi$ ,  $\varphi$ ,  $\bar{\xi}$  and  $\zeta$  along with  $\nu$  form a set of state variables for the optimal control formulation of the planning problem with private information. The envelope conditions (II.4) and (II.5) supply laws of motion for  $\psi$  and  $\varphi$ . Equations (II.9) and (II.10) give laws of motion for  $\bar{\xi}$  and  $\zeta$ :

$$(II.9) \quad \bar{\xi}_k(k) = e(k)\varphi(k)\pi(k)$$

$$(II.10) \quad \zeta_k(k) = C[\psi(k), e(k)]\pi(k).$$

Finally, the definition of  $\varphi(k)$  implies a law of motion for  $\nu$ :

$$(II.11) \quad \nu_k(k) = \left( \frac{\varphi(k)}{b(\nu(k))a[k, \nu(k)]} \right)^{\varepsilon} \pi(k)a[k, \nu(k)]e(k),$$

The monotonicity conditions on mechanisms needed to ensure incentive-compatibility are omitted and checked ex post. The effort function  $e$  is the control. The government's problem becomes:

$$(II.12) \quad \max_{\psi, \varphi, \bar{\xi}, \zeta, \nu, e} \int_{\underline{k}}^{\bar{k}} \psi(k)g(k)dk$$

subject to the laws of motion (II.4), (II.5) and (II.9) to (II.11) and the boundary constraints:

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

In this problem there is one control ( $e$ ) and five states ( $\psi, \varphi, \bar{\xi}, \zeta, \nu$ ). Routine manipulation of the first order and co-state equations yields the following ex-

pression for the optimal effort-consumption wedge:

$$\begin{aligned}
 & -w[k, v^*(k)] \frac{U_c^*(k)}{U_e^*(k)} - 1 = \\
 & \underbrace{\mathcal{H}^*(k) \frac{1 - \Pi(k)}{\pi(k)} \frac{\varphi_k^*(k)}{\varphi^*(k)} \int_k^\infty \left( 1 - \frac{g(t) U_c^*(t)}{p^{\zeta^*} \pi(t)} \right) \frac{U_c^*(k)}{U_c^*(t)} \mathcal{N}^*(k, t) \frac{\pi(t)}{1 - \Pi(k)} dt}_{\text{Mirrlees}} \\
 \text{(II.13)} \quad & - \underbrace{\mathcal{I}^*(k) \left[ \frac{p_k^{\varphi^*}(k)}{p^{\varphi^*}(k)} + \frac{a_k[k, v^*(k)]}{a[k, v^*(k)]} \right] \frac{p^{\varphi^*}(k)}{p^{\zeta^*} \pi(k)} \left( -\frac{U_c^*(k)}{U_e^*(k)} \right)}_{\text{Wage Compression}},
 \end{aligned}$$

where  $U_x^*(k) := U_x(c^*(k), e^*(k))$   $x \in \{c, e\}$  and similarly for  $U_{xz}^*(k)$ ,  $\mathcal{I}^*(k) := -\frac{1}{\varepsilon} \frac{w[k, v^*(k)]}{e^*(k)}$  is the elasticity of the  $k$ -talent wage with respect to effort holding the task allocation fixed,  $\mathcal{H}^* := \left\{ -\frac{U_{ec}^*}{U_e^*} + \frac{U_{ee}^*}{U_e^*} \right\} e^* + 1$  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,  $\mathcal{N}^*(k, t) = \exp \left\{ -\int_k^t \frac{e^*(s) U_{ce}^*(s)}{\varphi^*(s) U_c^*(s)} \right\}$ ,  $p^{\zeta^*} = E \left[ \int_k^{\bar{k}} \frac{1}{U_c^*(t'')} \pi(t'') dt'' \right]^{-1}$  is the optimal shadow resource multiplier and  $p^{\varphi^*}$  is the optimal co-state on the shadow wage  $\varphi$ . The Mirrlees and wage compression components are labeled.

### III. EMPIRICAL IMPLEMENTATION

In this appendix, we discuss details of the empirical implementation. We describe the data set used and illustrate robustness of our estimated parameters to alternative sample selection criteria. We discuss the issues raised by dropping the Cobb-Douglas production assumption used in the main text and show how our results are modified by alternative values for the elasticity of final output with respect to occupational output. We show that our empirical results are robust to reordering occupations in different decades according to the average wage within the decade and provide supporting evidence for our functional form restrictions on  $a$ .

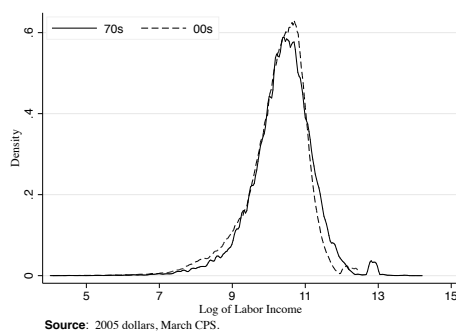
#### A. DATA SET AND SAMPLE SELECTION

Our main data source is the Current Population Survey (CPS) administered by the US Census Bureau and the US Bureau of Labor Statistics. We focus on the March release of the survey.<sup>2</sup> Data is available continuously from 1968 to 2012. On average each year of data contains about 150,000 observations, from 2001 the

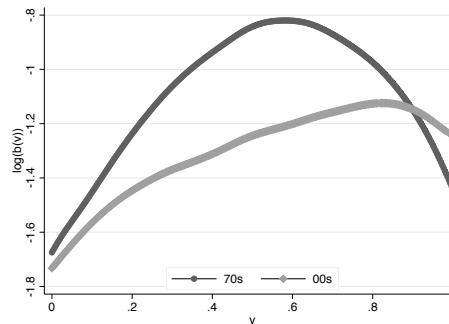
<sup>2</sup>Data is taken from: Miriam King, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

sample size has increased to approximately 200,000. The CPS contains detailed information on the demographic and work characteristics of each individual. For additional details on the CPS refer to Heathcote, Perri and Violante (2010) and Acemoglu and Autor (2011). The CPS data includes a self-reported estimate of hours worked from 1976 onwards. This question as well as questions on income are for the previous calendar year. Hence our sample covers the years 1975 to 2011 (interviews from 1976 to 2012). In the body of the paper we group observations in two groups. We call “the 70s” observations relating to years 1975-1979 (i.e interviewed in years 1976-1980), we call the “00s” observation relating to years 2000-2011 (interviews in 2001-2012).

The model analyzed is highly stylized. In order to make data and model compatible (and to reduce the likelihood of measurement error) we further restrict our sample. We drop individuals for whom income, age, sex, education, sector, occupation is not reported. We consider individuals of working age, i.e. between the ages of 25 and 65. We drop individuals with no formal education and the unemployed. Following Heathcote, Perri and Violante (2010), we also drop underemployed individuals: those working less than 250 hours per year or earning less than \$100 per year (dropping an additional 196,684 observations). Our final sample comprises of 2,039,123 individual/year observations. All variables are weighted with the provided weights and dollar denominated variables are deflated using CPI to 2005 dollars.<sup>3</sup> In Figure III.1a 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”.



(a) Distribution of log-labor income.



(b) Values of  $\log(b(v))$  over time.

Figure III.1. : Estimates on entire sample.

We briefly explore the impact of our sample selection on the estimated values of  $b$ . Figure III.1b shows  $b$  estimates for both decades obtained from the

<sup>3</sup>CPI for all urban consumers, all goods.

CPS sample before applying our sample selection (but after removal of individuals with missing information or an unclassifiable occupation). As can be seen polarization is still apparent. However for low  $v$  occupations we observe little change between the two decades. Note that a similar result would appear using the sample selection of Acemoglu and Autor (2011). This is because the authors only remove individuals who worked less than one week in the previous year or are less than 16 years of age.

### B. BEYOND COBB-DOUGLAS

In this section we extend our empirical strategy beyond the Cobb-Douglas assumption ( $\varepsilon = 1$ ) adopted in the main body. In this direction, it is important to recognize that the estimation of  $a$  is independent of  $\varepsilon$  and  $b(\cdot)$ . Estimation of  $b(\cdot)$  does, however, depend on  $\varepsilon$  and, outside of the Cobb-Douglas case,  $a$  as well. The firm's first order condition is:

$$(III.1) \quad \omega(k(v), v) = b(v)A^{\frac{\varepsilon-1}{\varepsilon}} \left( \frac{Y}{y(v)} \right)^{\frac{1}{\varepsilon}} a(k(v), v).$$

Let  $b^*(v)$  denote the share of output paid to occupation  $v$  and, hence, the estimate of  $b(v)$  in the Cobb-Douglas case. Then from (III.1):

$$(III.2) \quad b(v) = A^{\frac{\varepsilon-1}{\varepsilon}} \left( \frac{\omega(k(v), v)}{a(k(v), v)} \right)^{\frac{\varepsilon-1}{\varepsilon}} b^*(v)^{\frac{1}{\varepsilon}}.$$

Thus, to determine  $b$  outside of the Cobb-Douglas case values for  $a(\cdot)$  and  $\varepsilon$  are needed. Given values for these (III.2) determines  $b(\cdot)$  up to the constant  $A$ . The latter is pinned down by the restriction:  $\int_0^1 b(v)dv = 1$ .

It is well known that the elasticity of substitution between goods and factor augmenting technical progress cannot be separately identified from data on outputs, inputs and marginal products - an observation that goes back to McFadden, Diamond and Rodriguez (1978). The same logic implies that  $\varepsilon$  and  $b$  are not separately identified from this data. A typical response is to restrict the elasticity of substitution or the bias of factor augmenting technical change. In the main text, we proceed similarly by allowing the  $b$  parameter to be arbitrary and the production function to be Cobb-Douglas in occupational output. To assess the implications of this identifying assumption, we perform sensitivity analysis with respect to  $\varepsilon$ : we consider a range of values for  $\varepsilon$  and then re-compute  $b$ 's (and taxes) for each value.

Figure III.2 displays the  $b$  function for  $\varepsilon = 0.8$  and  $\varepsilon = 1.3$  (thicker lines show estimated  $b$ 's, thinner lines quadratic approximations to these estimates). The slope of the  $b$  function estimate is significantly impacted by variations in  $\varepsilon$ . However, polarization remains a consistent feature: between the 1970s and the 2000s, the  $b$  function rose in high  $v$  occupations, fell in mid ones and rose or only very

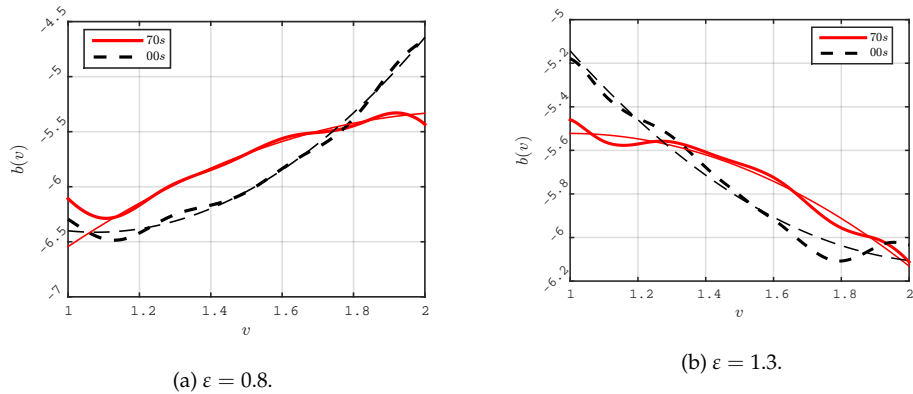
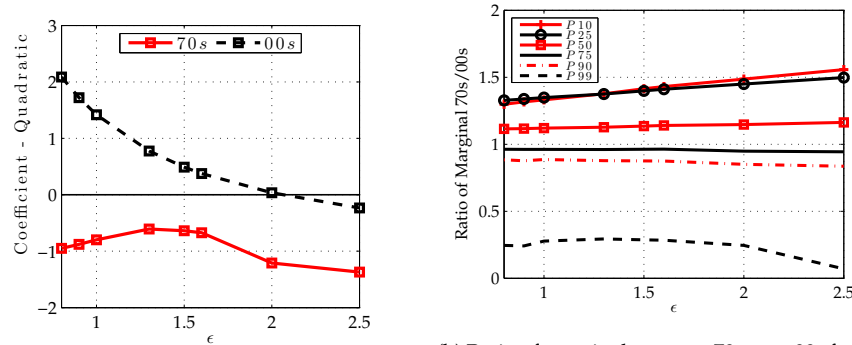


Figure III.2. : Estimates of  $b(v)$  for different  $\epsilon$  values.

marginally fell in low ones. We confirm and re-express this observation by taking a quadratic approximation to the  $b$  function (overlaid in Figure III.2 with thinner lines) for a variety of values of  $\epsilon$  and plotting the quadratic coefficient across decade and  $\epsilon$  value in Figure III.3a. The figure shows that over all the  $\epsilon$  values considered the quadratic coefficient increases between the 1970s and the 2000s and over most it changes sign (as in the benchmark environment considered in the body of the paper).



(a) Values of quadratic coefficient for approximation of  $b(\cdot)$  for different values of  $\epsilon$ .

(b) Ratio of marginal tax rates 70s over 00s for various percentile of income and over different values of  $\epsilon$ .

Figure III.3. : Impact of changes in  $\epsilon$  on  $b$  estimates and optimal taxes.

We recompute the optimal tax functions for different values of  $\epsilon$ . A summary of the impact of these values on optimal taxes is provided in Figure III.3b. Key patterns found in our benchmark case in the main text re-emerge. In particu-

lar, optimal marginal tax rates fall at low to mid incomes between the 1970s and 2000s and rise at higher ones with this effect becoming more pronounced as  $\varepsilon$  rises. Roughly speaking as  $\varepsilon$  rises occupational outputs and, hence, workers become more substitutable. This diminishes the government's ability to influence and, hence, redistribute via relative wages. Consequently, the "wage compression" force is weakened relative to the "Mirrleesian": the government is less motivated to compress wage differentials by moderating marginal tax reductions at the bottom and marginal tax increases at the top.

### C. OCCUPATIONAL ORDERING

In the main body of the paper, we restrict attention to 302 occupations present in both 1970s and 2000s data and order them using wage information from the 1970s. In doing so we follow the approach of Acemoglu and Autor (2011). This approach supposes that the complexity ordering of occupations is time invariant (and is captured by the 1970's wage ordering). To the extent that the occupational wage ordering changes and these changes reflect changes to the relative complexity of occupations estimates of the  $a$  and  $b$  functions are modified. Figure III.4 provides a scatterplot of occupations by their average wage rankings

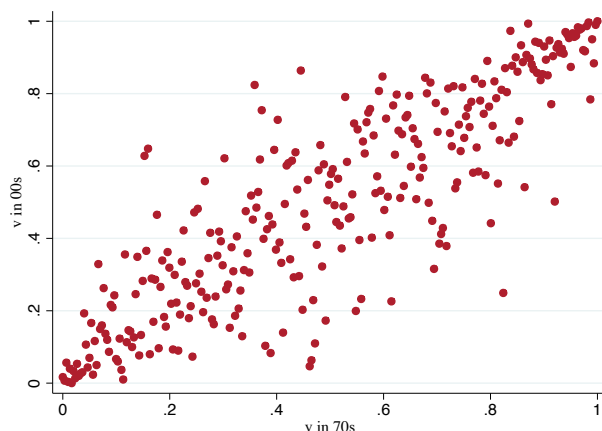


Figure III.4. : Rank order over decade.

in the 1970s and 2000s. The plot shows that while these rankings are not time invariant, they do exhibit stability especially at the bottom and the top. The overall correlation of these rankings over time is 0.8623.

Re-estimating the parameters of the  $a$  function for the 2000s using the 2000s wage ordering gives values of:  $\alpha_1 = 0.64$  and  $\alpha_2 = 2.90$  (compared to our previous estimates of  $\alpha_1 = 0.42$  and  $\alpha_2 = 3.01$ .) Thus, the critical comparative advantage parameter  $\alpha_2$  is only moderately changed under this alternative ordering

and the overall pattern of increasing competitive advantage is preserved. New estimates of the  $b$  function using the 2000s ordering are also only moderately changed and, in particular, they preserve the increase in weights on complex (high wage) occupations found previously.

We recompute optimal taxes using the estimates of  $a$  and  $b$  function parameters after reordering and recoding. Although marginal taxes for the 2000's are slightly higher under this alternative parameterization, the overall impact of technical change upon them is little altered. As before, these taxes fall on low (but not the lowest) and rise on high (but not the highest) incomes. The numerical results are available on request.

#### D. THE DEMAND FOR COMPLEX SKILLS ACROSS OCCUPATIONS

In the main text we assume a functional form for  $a$  that attributes the steepening of the profile of average wages across ranked occupations to increases in the comparative advantage of high talents in more complex tasks. However, an alternative scenario is also possible. Under this alternative, highly talented workers accumulated large stocks of general skills between the 1970s and the 2000s. This minority concentrates in high wage occupations where they have a slight comparative advantage, but talent-complexity comparative advantage has not greatly increased. Our account and this second one are difficult to distinguish using CPS data which, of course, does not give (imputed) wages for occupations other than those chosen by a worker. Greater acquisition of general skills by top talents could, in principal, be detected by adding non-linear terms in  $k$  (e.g.  $k^2$ ) to our empirical specification of the log  $a$  function. This, however, is problematic as these terms are highly collinear with the comparative advantage term  $k\nu$  in this function. Consequently, we look outside of the CPS for evidence of increasing talent-complexity comparative advantage. In particular, we use the O\*NET database to assess whether the intensity of complex skills and abilities has increased in high wage occupations relative to low. We treat such evidence as suggestive of increasing comparative advantage: increasing returns to talent in complex occupations.

Our empirical approach is similar to that in Autor, Levy and Murnane (2003), though our focus is distinct.<sup>4</sup> We pair two editions of the O\*NET database with the CPS, merging by occupation. For each of the Census-defined occupations we record the level and the importance of each of the 52 abilities and 35 skills reported by O\*NET. This enables us to create two snapshots of the ability and skill requirement of occupations across time. We use the first (beta) release of the O\*NET database from 1998 and the latest available release (version number 19.0) from July 2014.<sup>5</sup> We multiply the two reported dimensions of skill/ability

<sup>4</sup> This paper documents the mix of routine and non routine tasks performed across occupations and its evolution. It uses the predecessor of the O\*NET database, the Dictionary of Occupational Titles.

<sup>5</sup> Admittedly the time span covered by these two datasets is much shorter than the one covered by our two benchmark time periods in the body of the paper. However we conjecture that the patterns uncovered in our

within an occupation (importance and level) to create a single skill/ability “intensity” index. The numerical range over which these dimensions are measured has not changed across the editions of O\*NET. However, the meaning associated with each score has. To overcome this limitation and allow for a consistent time comparison we follow Autor, Levy and Murnane (2003) and look at the percentile-rank of each occupation by each skill/ability index. In Figure III.5 we display the (smoothed) change of the intensity index across occupations (ranked as in the paper by  $v$ ) for three distinct skill/abilities: *complex problem solving*, *deductive reasoning* and *mathematical reasoning*. The first of these is the skill and the

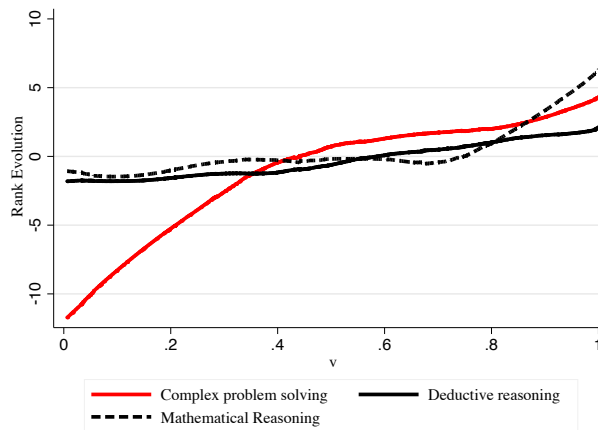


Figure III.5. : Rank evolution of selected skills and abilities intensity across occupations. LOWESS smoothed.

second the ability that most correlate with the task rank  $v$ . Thus, these intensity indices provide measures of task complexity that are consistent with the model. The third index (*mathematical reasoning*) is an ability that is commonly used to describe the complexity of an occupation. From the figure, each of these intensity index changes is negative for lower ranked occupations ( $v \leq 0.5$ ) and positive for higher ranked ones ( $v \geq 0.8$ ). The former is especially marked for complex problems solving and the latter for mathematical reasoning. Thus, consistent with our specification of  $a$ , skills and abilities that are particularly associated with occupational complexity have grown only in more complex occupations.

Table III.1 displays coefficients from regressions of changes in various skill/ability intensity indices on  $v$ . A positive estimate denotes a relative increase in the intensity of a skill or ability in more complex occupations. All of the point estimates in Table III.1 are positive. The majority of them are so with a high degree of confidence.

comparison of these two dataset would be amplified with a longer timespan.



| Skill/Ability                  | Coefficient     |
|--------------------------------|-----------------|
| <i>Complex problem solving</i> | 12.51*** (2.34) |
| <i>Critical Thinking</i>       | 3.51** (1.91)   |
| <i>Deductive reasoning</i>     | 4.70*** (2.15)  |
| <i>Inductive reasoning</i>     | 9.10*** (2.14)  |
| <i>Mathematical Reasoning</i>  | 4.02* (2.52)    |

Table III.1—: Slope of the change of skill/ability intensity index over occupation. Standard Errors in Parenthesis. \*\*\*= 95% confidence, \*\*= 90% confidence, \*= 85% confidence.

We conclude this section by taking a broader look at all 87 skills and abilities. To do this we perform a principal component analysis of the change in intensity indices across occupations. In Figure III.6a we display the first two principal components. The second principal component displays an increasing pattern of variation similar to the profiles plotted in Figure III.5. In Figure III.6b we display the loading factors. We label each skill/ability as being either associated with a high degree of complexity (the ones associated with information processing, problem solving, analytical thinking, managerial abilities) or not (mostly physical and interpersonal abilities).<sup>6</sup> We see that high complexity skill/abilities (in red in the graph) are on average associated with high loading on the second principal component and in some instances with a low loading on the first principal component. This implies that overall the high complexity skill/abilities display a positive sloping profile over the space of occupations. The opposite is true for low complexity skill/abilities.

Overall, we view data on the evolution of the skill and ability composition of tasks as broadly consistent with increasing talent-task complexity comparative advantage and the functional form restrictions that we place on the  $a$  function in our analysis.

#### E. ALTERNATIVE PRODUCTIVITY FUNCTION

As discussed in Subsection IV. C., in our benchmark estimation the increase in the growth rate for log wages at higher talents leads to the identification of a growing comparative advantage over time. In this section we explore an alternative formulation of the productivity function aimed at fitting more closely the high and increasing growth rate of wages for high talents. Specifically, we set  $\frac{\partial a}{\partial k}(k, v) = \alpha_3 \cdot v^2$ . Proceeding as in Subsection IV. C., we find a value of  $\alpha_3 = 0.79$  (0.03) for the 1970's and a value of  $\alpha_3 = 0.91$  (0.04) for the 2000's.<sup>7</sup> As in our benchmark, there is an increase in the degree of comparative advantage

<sup>6</sup>Detailed listing available is upon request.

<sup>7</sup>In addition, to emphasize the behavior of wages for higher talents we estimate  $\alpha_3$  without weighting by the shares of talent in each occupation.

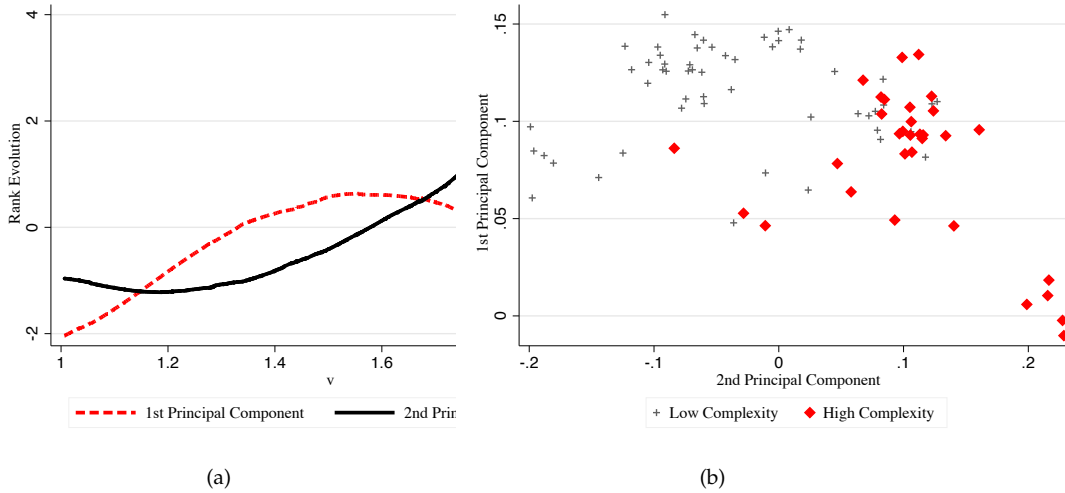


Figure III.6. : Principal component analysis of the rank evolution of skill and ability intensity across occupations.

over time. Given the quadratic nature of the productivity function the change in overall top to bottom talent inequality in wages is greater than in our benchmark setting. In Table III.2 we display the resulting optimal behavior of average and marginal tax rates over percentiles of the income distribution. Relative to

Table III.2—: Optimal Tax Rates on Real Labor Income, Alternate Case.

|                  | Decade | Percentiles of Income |                  |                  |                  |                  |                  |
|------------------|--------|-----------------------|------------------|------------------|------------------|------------------|------------------|
|                  |        | 10 <sup>th</sup>      | 25 <sup>th</sup> | 50 <sup>th</sup> | 75 <sup>th</sup> | 90 <sup>th</sup> | 99 <sup>th</sup> |
| <b>Averages</b>  | 70s    | -3.3                  | 0.9              | 7.8              | 20.3             | 24.7             | 20.8             |
|                  | 00s    | -12.5                 | -8.7             | 4.0              | 21.6             | 28.3             | 23.7             |
| <b>Marginals</b> | 70s    | 17.2                  | 28.8             | 40.0             | 38.9             | 24.3             | -1.8             |
|                  | 00s    | 20.0                  | 32.5             | 44.5             | 44.8             | 29.5             | -1.5             |

Note: Estimates of tax rates determined using  $\frac{\partial a}{\partial k} = \alpha_3 \cdot v^2$ .

the benchmark, the striking difference is the sharp fall in average and increase in marginal rates over the time period. Now, the rise in comparative advantage dominates. It strengthens the wage compression channel and increases wage growth across talents. This increases the motive for redistribution towards the bottom.

## IV. COUNTERFACTUALS

In this appendix we separately evaluate the impact of change in the  $a$  and  $b$  functions on optimal policy. To do so, we first hold the parameters of  $a$  fixed at their 1970s values, while allowing those of  $b$  to change to their 2000s values; we compute the corresponding optimal tax equilibrium. We then repeat the exercise holding the parameters of  $b$  fixed, while allowing those of  $a$  to change. We compare the resulting tax equilibria to those in which both functions are at their 1970s or 2000s levels.

## ASSIGNMENT

Figure IV.1 shows the impact of the empirical  $a$  and  $b$  changes together and in isolation on the density of workers across tasks. Changes in  $b$  alone lead to quite large changes in the relative “number” of workers performing tasks. In particular, the polarizing adjustments in task demand (growth at the extremes relative to the middle) occurring between the 1970s and the 2000s induce growth in the density of workers at the extremes and, hence, job polarization in the associated optimal tax equilibrium. Changes in  $a$  alone have an opposite (if more modest) effect: the number of workers performing mid-level tasks grows relative to the extremes. This reflects productivity growth in low tasks by low talents and in high tasks by high talents inducing reductions in shadow task prices at the top and the bottom and movements of some lower and higher talents into mid-level tasks. However, when changes in the  $b$  and  $a$  parameters are combined, it is the former that dominates. Although, the  $b$  parameter change induces quite large

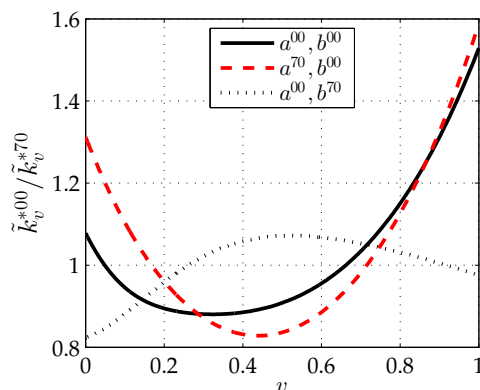


Figure IV.1. : Relative changes in  $\tilde{k}_v^*$  from the 1970s to the 2000s. Allowing  $b$  to change,  $a$  to change and both  $a$  and  $b$  to change.

changes in the numbers of workers performing particular tasks, this is achieved

with only modest occupational reassignments of given workers. As shown in Figure IV.4, low-mid level talents reduce their task assignment, but by no more than 2%, high-mid level talents increase their task assignment, but by no more than 3%.

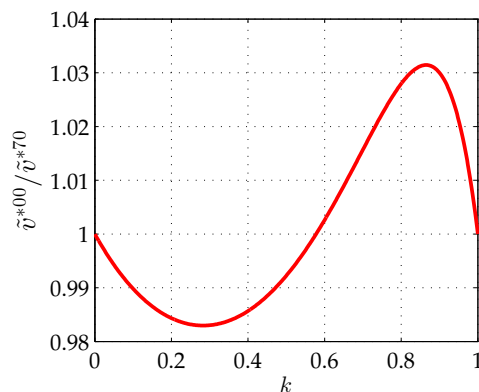


Figure IV.2. : Relative changes in task assignment  $\tilde{v}^*$  induced by the shift in  $b$  from 1970 to 2000.

#### WAGE CHANGES

An implication of the modest change in task assignment induced by the shift in the  $b$  function is that equilibrium wage growth over talents is also only modestly altered by this shift. As shown in Figure IV.3,  $b$  changes alone induce very slight compression in wage differentials across low-to-mid talents and very slight expansion across mid-to-high talents. Changes in the  $a$  function also depress wage growth across talents at the bottom and raise it at the top, but the effect is much more pronounced.

#### MARGINAL TAX CHANGES

The shift in the  $b$  function alone has limited impact on the relative wage-effort elasticities and on the wage compression term. Combined with its small impact on wage growth over talents, it has a correspondingly modest effect on marginal taxes, see Figure IV.4a. In contrast, the shift in the  $a$  function has a much more significant impact on wage growth and on the relative wage-effort elasticities. It has a much more significant effect on optimal marginal taxes and accounts for most of the adjustment between the 1970s and the 2000s.

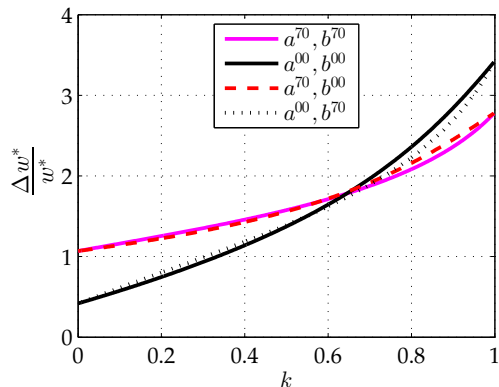


Figure IV.3. : Equilibrium wage growth  $\Delta w^*/w^*$  for the parameter combinations  $(a_{70}, b_{70})$ ,  $(a_{70}, b_{00})$ ,  $(a_{00}, b_{70})$  and  $(a_{00}, b_{00})$ .

#### V. OPTIMAL POLICY UNDER A RAWLSIAN OBJECTIVE

In this subsection, we recompute optimal taxes under a Rawlsian societal objective that attaches positive weight only to the utility of the lowest talent. The results are given in Table V.1. Relative to the benchmark case, the government's enhanced concern for redistribution translates into higher marginal income tax rates at nearly all income levels, larger subsidies to low and middle income earners and larger average taxes on those in the top quartile. However, the impact of technical change remains unaltered. Marginal taxes fall on low to middle income quantiles (but not at the very lowest), rise on high income quantiles (but not the very highest). Average taxes rise sharply at the bottom of the income distribution and fall at mid to higher incomes. The largest beneficiary are those at the 75 income percentile who see the largest reduction in average taxes (from 34% to 28.2%).

Table V.1—: Optimal Tax Rates on Real Labor Income: Rawlsian Case.

|                  | Decade | Percentiles of Income |                  |                  |                  |                  |                  |
|------------------|--------|-----------------------|------------------|------------------|------------------|------------------|------------------|
|                  |        | 10 <sup>th</sup>      | 25 <sup>th</sup> | 50 <sup>th</sup> | 75 <sup>th</sup> | 90 <sup>th</sup> | 99 <sup>th</sup> |
| <b>Averages</b>  | 70s    | -144.1                | -79.4            | - 6.1            | 34.0             | 41.3             | 34.5             |
|                  | 00s    | -54.1                 | -39.4            | -6.3             | 28.2             | 39.6             | 32.4             |
| <b>Marginals</b> | 70s    | 80.6                  | 78.9             | 74.4             | 62.2             | 38.0             | -0.4             |
|                  | 00s    | 70.3                  | 71.1             | 70.3             | 63.0             | 41.4             | -3.9             |

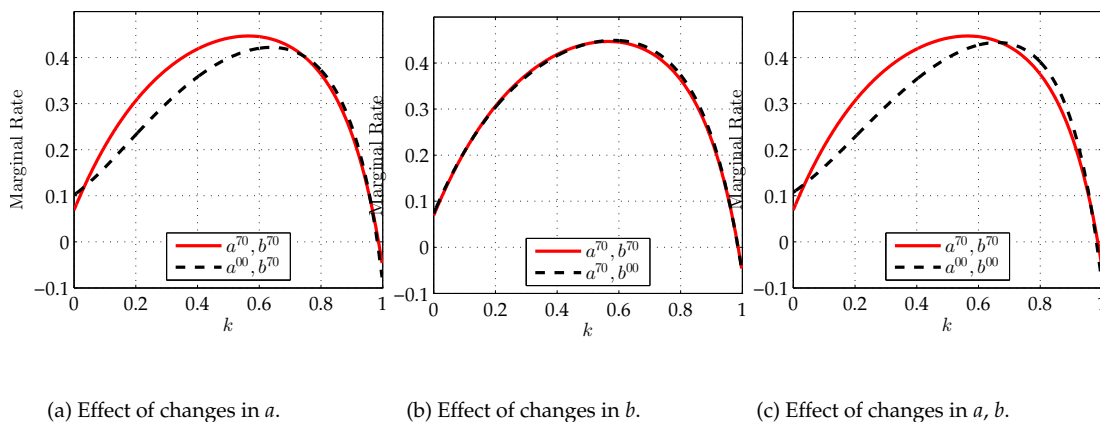


Figure IV.4 : Marginal tax effects.

## VI. A MODEL WITH INTRA-OCCUPATIONAL WAGE DISPERSION AND A HEAVY-TAILED TALENT DISTRIBUTION

In the main text, we consider a baseline model in which worker talent and task complexity are compressed to single dimensional variables. This formulation simplifies the theoretical analysis, facilitates empirical identification and connects models of technical change, assignment and taxation in a very direct way. However, the model's equilibrium does not permit intra-task wage variation. In this appendix we explore the extent to which our baseline results are qualified by its omission. In particular, to the extent that such wage variation is underpinned by variation in talent that is unrelated to tasks, the link between wages and tasks is weakened. A natural conjecture is that the responsiveness of policy to (task-level) technical change is similarly weakened. As described in Section VI. , we enhance the contribution of talent unrelated to tasks and seek a lower bound for the responsiveness of policy to technical change.

In the remainder of the appendix, we proceed as follows. First, we detail a general framework that accommodates high dimensional talent and high dimensional tasks. We show how this framework can be specialized to yield the model of Rothschild and Scheuer (2013) and our (baseline) model in the main text and, in so doing, relate the two. We then develop a specialization that is intermediate between these cases. In this specialization, there are two aspects of worker talent: one captures comparative advantage in complex tasks, the other the ability to do all things well. The latter creates intra-task wage variation. Similar to our baseline model, comparative advantage types partition the ordered space of occupations amongst themselves. We attribute all wage variation within these partitions to variations in absolute advantage uncorrelated with task. Such

variation diffuses the effect of task-based technological change across the wage distribution and the impact of changes in taxes directed at a particular income across the set of tasks, task shadow prices and wages. These effects dampen the impact of technical change on policy design. In taking the model to the data we assume a coarse set of comparative advantage types and attribute all residual wage variation within the comparative advantage partitions (about 75% of total wage dispersion in our CPS sample) to variations in absolute advantage. We attribute none to measurement error in incomes, hours or tasks and absorb some measured inter-occupational wage variation into the residual by keeping the set of comparative advantage types and number of partitions small. These assumptions enlarge the dampening effect of absolute advantage variation. We find that the impact of technical change on policy is smaller, but the direction is unchanged. The qualitative conclusion that marginal taxes should be reduced on low to middle incomes, but raised on higher ones (with opposite adjustments in the extreme tails) remains intact. We interpret these results as lower bounds for the responsiveness of policy to technical change.

#### A. A GENERAL FRAMEWORK

We first develop a general framework in which, as in Section II, (consumption and effort) allocations are defined as functions of workers' types and the domain of the production function is the space of effort allocations. Subsequently, we introduce assignment. Relative to the main text the generality lies in our treatment of the talent space.

Assume that the workers are partitioned across talents according to a probability space  $(\Theta, \mathcal{F}, P)$ , where, to begin with,  $\Theta$  is an arbitrary set. Let  $\mathcal{A} = \mathcal{C} \times \mathcal{E}$  denote a set of allocations with each allocation a pair of measurable functions  $c \in \mathcal{C}$ ,  $c : \Theta \rightarrow \mathbb{R}_+$ , and  $e \in \mathcal{E}$ ,  $e : \Theta \rightarrow \mathbb{R}_+$ , describing the consumption and effort of differently talented workers. The set  $\mathcal{E}$  is further restricted to be a Banach space. Let  $F : \mathcal{E} \rightarrow \mathbb{R}_+$  denote a production function defined directly on the space of effort allocation functions, with  $F$  concave and (Fréchet) differentiable. The government's problem is:

$$(VI.1) \quad \sup_{\mathcal{A}} \int U(c(\theta), e(\theta)) P(d\theta)$$

subject to  $\forall \theta, \theta'$ ,

$$(VI.2) \quad U(c(\theta), e(\theta)) \geq U\left(c(\theta'), \frac{\omega(\theta', e)e(\theta')}{\omega(\theta, e)}\right)$$

and

$$(VI.3) \quad \int \{\omega(\theta, e)e(\theta) - c(\theta)\} P(d\theta) \geq G,$$

where (VI.2) and (VI.3) are, respectively, the incentive-compatibility and resource constraints and wages  $\omega$  are given by the Fréchet derivative of  $F$ .<sup>8</sup>  $G \in \mathbb{R}_+$  is government spending.

This framework can be interpreted as the reduced form of an economy with assignment. Let  $(V, \mathcal{V})$  denote a measurable space of tasks and  $\mathcal{M}$  the set of finite measures on  $V$ . Interpret such measures as allocations of effective labor across tasks. Let  $H : \mathcal{M} \rightarrow \mathbb{R}_+$  be a production function (now defined on the space of effective labor allocations) and  $\mu : \Theta \times V \rightarrow \mathbb{R}_+$  a (measurable) productivity kernel giving the productivity of each talent in each task. As in Section III, assignment is efficient in the competitive equilibria and planner's problems that we consider. Thus, as there, an indirect production function over effort allocations is recoverable from an assignment problem:

$$(VI.4) \quad F(e) = \sup_{m, \Lambda} \left\{ H(m) : \begin{array}{l} \forall B \in \mathcal{V}, \int_B m(dv) = \int_B \int_{\Theta} \mu(\theta, v) e(\theta) \Lambda(d\theta, dv) \\ \forall B \in \mathcal{F}, \int_B P(d\theta) = \int_V \int_B \Lambda(d\theta, dv) \end{array} \right\},$$

where  $m$  is a distribution of effective labor across tasks and  $\Lambda$  is a distribution of workers across tasks and talents. The constraints in (VI.4) ensure that these distributions are consistent with one another, the effort allocation and the underlying distribution of talent. It follows that, as in the main text, the planner's problem with assignment can be decomposed into an outer step (VI.1) in which  $(c, e)$  are chosen and an inner step (VI.4) in which  $(m, \Lambda)$  are chosen (to determine  $F$  at  $e$ ).

#### B. REFORMULATION OF THE GENERAL FRAMEWORK

Enlarging the dimension of talent to allow for multiple attributes that interact differently across tasks greatly complicates the pattern of binding incentive constraints. Rothschild and Scheuer (2013) observe that if  $\Theta$  is uncountable, then (almost all) talents earning the same wage receive the same consumption and effort. Consequently, allocations can be re-expressed as functions of (one dimensional) wages and in this form only local incentive constraints bind. This is an important simplifying insight that we utilize below. However, it is not costless as it requires the introduction of rather complicated constraints that relate the (endogenous) wage distribution to allocations (as functions of wages). In the remainder of this section, we make one simplification: in the inner assignment problem, we restrict attention to effective labor allocations described by (density) functions  $l : V \rightarrow \mathbb{R}_+$  rather than measures. Thus, we exclude atoms of effective labor in tasks. As before, let  $H$  denote the production function, but defined now on the domain of such densities. In addition, assume that  $H$  is (Fréchet) differentiable with derivative  $\partial H$ . The shadow price of task  $v$  output is  $\partial H(l; v)$

<sup>8</sup>We continue to exclude externalities which is a focus of Rothschild and Scheuer (2014) and Lockwood, Nathanson and Weyl (2014).



and the wage of a worker of talent  $\theta$  in task  $v$  is:  $\partial H(l, v)\mu(\theta, v)$ . Workers choose their tasks to maximize their wages:

$$(VI.5) \quad v^*(\theta; l) \in \arg \max_v \partial H(l; v)\mu(\theta, v).$$

Let  $w^*(\theta; l) = \partial H(l; v^*(\theta; l))\mu(\theta, v^*(\theta, l))$  denote the (maximized) wage of a worker of talent  $\theta$  given the labor allocation  $l$ . By (VI.5),  $l$  implies a distribution of workers over wages and tasks:

$$(VI.6) \quad R(l)(w, v) := \int_{\Theta} 1(w^*(\theta; l) \leq w, v^*(\theta; l) \leq v)P(d\theta).$$

In this setting, define an allocation to be a triple  $(l, \tilde{c}, \tilde{e})$  with  $\tilde{c} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and  $\tilde{e} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  mapping *wages* rather than talents to consumption and effort choices. Note that from (VI.5), given  $l$ , any worker receiving wage  $w$  in task  $v$  has productivity  $\frac{w}{\partial H(l; v)}$  and an effective labor of  $\frac{w}{\partial H(l; v)}\tilde{e}(w)$ . Consistency of  $l$  with  $\tilde{e}$  thus requires:

$$(VI.7) \quad \forall v : \quad l(v) = \int_0^\infty \frac{w\tilde{e}(w)}{\partial H(l; v)}R(l)(dw, v).$$

In addition, the wage distribution  $Q$  must equal the wage marginal of  $R(l)$ :

$$(VI.8) \quad \forall w : \quad Q(w) = \int_0^w \int_V R(l)(dw', dv').$$

Hence, the government's problem can be re-expressed as:

$$(VI.9) \quad \sup_{\tilde{c}, \tilde{e}, l, Q} \int_0^\infty U(\tilde{c}(w), \tilde{e}(w))Q(dw)$$

subject to (VI.7) and (VI.8),  $\forall w, w'$ ,

$$(VI.10) \quad U(\tilde{c}(w), \tilde{e}(w)) \geq U\left(\tilde{c}(w'), \frac{w'\tilde{e}(w')}{w}\right),$$

and

$$(VI.11) \quad \int_0^\infty \{w\tilde{e}(w) - \tilde{c}(w)\}Q(dw) \geq G.$$

The main difficulties in (VI.9) are the constraints (VI.7) and (VI.8) relating  $Q$  to  $l$  and  $\tilde{e}$ , absent which everything would reduce to a standard Mirrlees problem.

## ROTHSCHILD AND SCHEUER'S SPECIALIZATION

Rothschild and Scheuer (2013) make progress by assuming that: (i)  $V = \{1, 2\}$ , (ii)  $\Theta = \mathbb{R}^2$  with  $\theta = (\theta_1, \theta_2)$  and  $\mu(\theta, v) = \theta_v$  and (iii)  $P$  has a density  $p$ . These restrictions make (VI.7) and (VI.8) manageable.  $H$  is now a function of only two variables  $(l_1, l_2)$  and  $R$  is given by:

$$(E.6'a) \quad R(l)(w, 1) = \int_{\underline{\theta}}^{\frac{w}{\partial H(l;1)}} \int_{\underline{\theta}}^{\theta_1 \frac{\partial H(l;1)}{\partial H(l;2)}} p(\theta_1, \theta_2) d\theta_2 d\theta_1$$

$$(E.6'b) \quad R(l)(w, 2) = \int_{\underline{\theta}}^{\theta_2 \frac{\partial H(l;2)}{\partial H(l;1)}} \int_{\underline{\theta}}^{\frac{w}{\partial H(l;2)}} p(\theta_1, \theta_2) d\theta_2 d\theta_1.$$

The low dimensionality of  $V$  (and, hence,  $l$ ) suggests an inner-outer approach to solving (VI.9) quite distinct from that described in the main text. The inner component maximizes (VI.9) over  $(\tilde{c}, \tilde{e})$  subject to (VI.7), (VI.10) and (VI.11) with  $l$  fixed and  $Q$  set to  $Q(\cdot) = \sum_{v=1,2} R(l)(\cdot, v)$ . This problem is a standard Mirrlees problem augmented by (VI.7). The outer component maximizes the resulting value function over  $l$ . Note that here assignment of effective labor across tasks  $l$  is solved for in the *outer step*, in contrast to the formulation in the main text where this is done in the inner step.

## AN ALTERNATIVE SPECIALIZATION

We now present a version of the general formulation given above in which workers are distributed across two talent attributes. The first interacts with tasks and affects comparative advantage, the second influences the ability to do all things and absolute advantage. Let  $V := [0, 1]$  and  $\Theta := \{1, \dots, K\} \times \mathbb{R}_+$ . Denote elements of  $\Theta$  by  $\theta := (k, \psi)$  and assume that  $\mu$  has the form  $\mu(\theta, v) := \psi a_k(v)$  where the function  $a$  is log super-modular in  $(k, v)$ . Workers of a given  $k$  type have the same profile of relative wages and the same preference ordering over tasks; variations in  $\psi$  cause workers to be more or less good at all things and underpin intra-task wage dispersion. Let  $\{\pi_k\}_{k=1}^K$  denote the distribution of workers across  $k$  and  $\{f_k\}_{k=1}^K$  the densities of workers over  $\psi$  conditional on  $k$ . The latter are assumed to satisfy  $E_k[\psi] = 1$ . Also let  $H(l) = [\int_0^1 b(v)l(v)^{\frac{\epsilon-1}{\epsilon}} dv]^{\frac{\epsilon}{\epsilon-1}}$ . As in the main text,  $k$ -types (i.e. workers with a common  $k$ , but potentially different values of  $\psi$ ) sort themselves across tasks with those of a given type  $k$  distributing themselves over a sub-interval  $[\tilde{v}_{k-1}, \tilde{v}_k]$  so as to ensure a common value for  $w_k := \partial H(l, v) a_k(v)$ . The wage received by a  $(k, \psi)$ -type is  $w = \psi w_k$  and, since  $E_k[\psi] = 1$ ,  $w_k$  is the average wage per unit of effort received by  $k$ -types. In this setting, it is useful to define the  $\psi$ -weighted average effort ("effec-

tive labor supply") of the  $k$ -th type:

$$(VI.13) \quad \ell_k = \int_0^\infty \psi e(k, \psi) f_k(\psi) d\psi = \int_0^\infty \frac{w}{w_k} \tilde{e}(w) f_k\left(\frac{w}{w_k}\right) \frac{dw}{w_k},$$

where the second term expresses effort as a function of  $(k, \psi)$  and the third re-expresses it as a function of the wage and uses  $w = \psi w_k$  to change variables from  $\psi$  to  $w$ . The vector of effective labor supplies  $\ell = \{\ell_k\}$  (rather than the full allocation of effective labor over tasks  $l$ ) is sufficient to determine the average wages  $\{w_k\}$  and, hence, the complete wage distribution. Similar to the main text it can be shown that given  $\ell = \{\ell_k\}$ , final output is:

$$(VI.14) \quad Y(\ell) := \max_{\{\tilde{v}_k\}} \left[ \sum_{k=1}^K B_k(\tilde{v}_{k-1}, \tilde{v}_k) \ell_k^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}},$$

with  $B_k(\tilde{v}_{k-1}, \tilde{v}_k) = \left( \int_{\tilde{v}_{k-1}}^{\tilde{v}_k} b(v)^\varepsilon a_k(v)^{\varepsilon-1} dv \right)^{\frac{1}{\varepsilon}}$ . In addition, the wage terms  $\{w_k\}$  are given by:

$$(VI.15) \quad w_k(\ell) = \left( \frac{Y(\ell)}{\ell_k} \right)^{\frac{1}{\varepsilon}} \tilde{B}_k(\ell),$$

where  $\tilde{B}_k(\ell)$  is the value of  $B_k(\tilde{v}_{k-1}, \tilde{v}_k)$  at the optimized task thresholds from (VI.14) and the notation makes the dependence of  $w_k$  on  $\ell$  explicit. Combining (VI.13) and (VI.15) gives an analogue of (VI.7):

$$(E.7'') \quad \ell_k = \int_0^\infty \frac{w}{w_k(\ell)} \tilde{e}(w) f_k\left(\frac{w}{w_k(\ell)}\right) \frac{dw}{w_k(\ell)}.$$

Variations in absolute advantage  $\psi$  create variations in the wages paid to members of a given  $k$ -type group. Thus, they create wage variation within each partition of the occupational space  $[\tilde{v}_k, \tilde{v}_{k+1}]$ . The overall wage distribution  $Q$  satisfies an analogue of (VI.8):

$$(E.8'') \quad Q(w) = \int_0^w \sum_{k=1}^K f_k\left(\frac{w'}{w_k(\ell)}\right) \frac{\pi_k}{w_k(\ell)} dw'.$$

The government's problem can then be expressed as:

$$(VI.16) \quad \sup_{\tilde{c}, \tilde{e}, \ell, Q} \int U(\tilde{c}(w), \tilde{e}(w)) Q(dw)$$

subject to (VI.10), (E.7'') and (E.8''), and the resource constraint:

$$\int \{w\tilde{e}(w) - \tilde{c}(w)\}Q(dw) \geq G.$$

Assuming that the first order approach is valid (i.e. that the workers' envelope condition is sufficient for incentive compatibility), the implied optimal marginal tax paid by a worker earning  $w$  is:

$$(VI.17) \quad \tau(w) = \frac{\frac{1-Q(w)}{wq(w)}\Psi(w)\mathcal{H}(w)}{1 + \frac{1-Q(w)}{wq(w)}\Psi(w)\mathcal{H}(w)} + \frac{\mathcal{N}(w)}{1 + \frac{1-Q(w)}{wq(w)}\Psi(w)\mathcal{H}(w)},$$

where  $q(w)$  is the wage density,  $\Psi(w)$  is the normalized multiplier on the incentive constraint,  $\mathcal{H}(w) = \frac{1+\mathcal{E}^u}{\mathcal{E}^c}$ , with  $\mathcal{E}^u$  the uncompensated and  $\mathcal{E}^c$  the compensated labor supply elasticities. The first right hand side component of (VI.17) is the conventional Mirrlees tax term; the second is the wage compression term. A tax induced effort perturbation at  $w$  impacts the effective labor of each  $k$ -type population (since each includes some workers receiving wage  $w$ ). This, in turn, after task migration of  $k$  types, affects the  $w_k$  terms and, hence, the entire wage distribution. These effects are seen by decomposing the wage compression term numerator  $\mathcal{N}$  as:

$$\mathcal{N}(w) = N \cdot L(w),$$

where  $L(w)$  is a column vector giving the impact of a perturbation in  $\tilde{e}(w)$  on the vector of effective labor supplies  $\ell = \{\ell_k\}$  and  $N$  is a row vector giving the shadow value of a perturbation in  $\ell$  on the distribution of wages. Specifically, the  $k$ -th element of  $N$  is:

$$N_k := \frac{1}{\ell_k} \int_0^\infty \lambda(w') \sum_{\kappa=1}^K \left( \frac{w_\kappa(\ell)}{q_\kappa(w'; w_\kappa(\ell))} \frac{\partial q_\kappa(w'; w_\kappa(\ell))}{\partial w_\kappa} \right) \varepsilon_{w_\kappa, \ell_\kappa} q_\kappa(w'; w_\kappa(\ell)) dw',$$

where  $\varepsilon_{w_\kappa, \ell_\kappa} := \frac{\ell_k}{w_\kappa(\ell)} \frac{\partial w_\kappa(\ell)}{\partial \ell_k}$  is the elasticity of the  $\kappa$ -th comparative advantage type's average wage with respect to the  $k$ -th type's effective labor (and an analogue of the relative wage-effort elasticities from the main text),  $\frac{w_\kappa(\ell)}{q_\kappa(w'; w_\kappa(\ell))} \frac{\partial q_\kappa(w'; w_\kappa(\ell))}{\partial w_\kappa}$  is the elasticity of the  $\kappa$ -th type's conditional wage density at  $w'$  with respect to  $w_\kappa$  and  $\lambda(w') = V(w') + \chi\{w'\tilde{e}(w') - \tilde{c}(w')\}$  is the societal value of the allocation given to workers earning  $w'$  (i.e. the utility plus the shadow value of the resource surplus these workers generate). As in the main text, technical change impacts the wage functions  $w_\kappa$  and the wage elasticities  $\frac{\ell_k}{w_\kappa(\ell)} \frac{\partial w_\kappa(\ell)}{\partial \ell_k}$  directly and via the reallocation of  $k$ -types across tasks. The affect of these changes on the wage distribution and tax policy's ability to shape this distribution is now diffused through the densities  $q_k$ . They modify the hazard term  $\frac{1-Q(w)}{q(w)w}$  (the ana-

logue of  $\frac{\Delta w_{k+1}^*}{w_{k+1}^*} \frac{1-\Pi_k}{\pi_k}$  in the main text) and, hence, the Mirrlees term in (VI.17). They also change the numerator of the wage compression term which incorporates the elasticities  $\frac{\ell_k}{w_k(\ell)} \frac{\partial w_k(\ell)}{\partial \ell_k}$ .

#### BRINGING THE FORMULATION TO DATA

Via the introduction of the additional talent attribute  $\psi$ , the above formulation permits intra-task wage variation and an unbounded wage distribution. In these dimensions it advances the baseline model in the main text. However, numerically solving (VI.16) becomes challenging as  $K$  becomes large. Thus, in contrast to the simpler baseline model in which the set of  $k$  values is uncountable,  $K$  is restricted to equal four in the calculations below.<sup>9</sup> The above formulation inherits many of its parameters from the baseline model. In our calculations below, we retain earlier values for these parameters. In particular, utility parameters are reused, the final goods production function is assumed to be Cobb-Douglas and our earlier estimates of  $b$  and  $a$  retained. Values for the conditional densities  $\{f_k\}$ , the new parameters in (VI.16), are required. To derive estimates of these, and following the approach in the main text, we use a worker's occupation to infer his or her  $k$ -type. Specifically, we order occupations by average wage, use this ordering and a worker's occupation to rank workers and then recombine workers into  $K = 4$  ranked and equally sized  $k$  groups. Thus, a worker is of type  $k = 1$  if she belongs to the first  $\frac{1}{K}$  of workers by (ordered) occupation, she is of type  $k = 2$  if she belongs to the next  $\frac{1}{K}$  workers by (ordered) occupation and so on. To capture  $\psi$  dispersion, we fit Burr Type XII distributions to each demeaned  $k$ -group. The density of a Burr Type XII distribution converges asymptotically to a Pareto density in the right tail, but admits a non-Paretian form that better accounts for wage data over the remainder of its domain.<sup>10</sup> The Burr XII is estimated by maximum likelihood; the estimation procedure accounts for top coded observations. For the  $k = 4$  group, our estimated Pareto tail parameter is 3.07. The fitted Burr XII distributions account for about 75% of the overall wage dispersion in our sample.

#### RESULTS

Table VI.1 gives the resulting optimal marginal tax rates. Relative to the benchmark case considered in the main text, this exercise generates similar values for marginal tax rates on middle incomes, but, as expected, quite different values for marginal rates on incomes in the tails of the distribution. The latter are shaped by the tails of the Burr XII distributions.

<sup>9</sup>These calculations use the non-linear optimizer SNOPT. The code is available on request.

<sup>10</sup>The Burr Type XII distribution has density  $g(x) = \frac{cdx^{c-1}}{(1+x^c)^{d+1}}$  and hazard  $\frac{1-G(x)}{g(x)x} = \frac{1+x^c}{cdx^c}$ , where  $c$  and  $d$  are parameters. It was proposed as flexible description of income distributions by Singh and Maddala (1976). McDonald (1984) concludes that the Burr XII outperforms many other heavy-tailed distributions in describing the distribution of income.

Table VI.1—: Optimal Tax Rates on Real Labor Income.

|                  | Decade | Percentiles of Income |                  |                  |                  |
|------------------|--------|-----------------------|------------------|------------------|------------------|
|                  |        | 10 <sup>th</sup>      | 50 <sup>th</sup> | 90 <sup>th</sup> | 99 <sup>th</sup> |
| <b>Marginals</b> | 70s    | 41.8                  | 48.2             | 42.9             | 44.5             |
|                  | 00s    | 43.2                  | 45.7             | 43.7             | 43.4             |

The response of optimal marginal tax rates to technological change, while dampened, is qualitatively similar to before: marginal rates rise over the lowest two income deciles, fall over low-mid income deciles (between the second and seventh decile), rise over higher income deciles (seventh to ninth) and fall at the very top, see Figure VI.1 which gives changes in rates. The dampening of the response, marginal tax rates change by at most 2.5 points at a given income compared to a maximum of 8.5 points in the baseline case, stems from the introduction of variation in absolute advantage that diffuses the impact of both technical change and taxes on wages. However, by assuming only  $K = 4$  comparative advantage groups and by attributing all residual measured wage variation to absolute advantage, we make this dampening force very large. We interpret these numerical results as lower bounds on responsiveness of optimal policy to technical change.

After decomposition, the Mirrlees and Wage Compression terms display similar movements to those in the main text: changes in the Mirrleesian term prevail over most of the income quantile domain, but are dampened by offsetting movements in the wage compression term. For example, at the fourth income decile, the Mirrlees term falls by more than 4 points, while the wage compression term rises by about 2 points. The logic behind changes to the Mirrlees and compression terms is essentially that in the main text. Expressed in terms of changes to the wage distribution rather than the structure of binding incentive constraints, technical change has compressed wage differentials at the bottom, but increased them at the top. This has thinned out the wage distribution in the tails, but fattened it in the middle. Since, other things equal, it is desirable to impose higher marginal taxes in areas of the distribution where the wage density is low and few will be distorted, the Mirrlees component falls on the low-middle incomes, but rises on high incomes (and on very low incomes in the extreme lower tail). The wage compression effect is enhanced by technical change as reduced substitutability of workers across tasks gives the government, through tax policy, more leverage over task shadow prices and wages. This creates a wage compression motive for raising taxes at the bottom and lowering them at the top. Overall change in the tax code depends on the balance of these forces, with the wage compression force only predominating in the extreme tails.

Overall, while quantitative responses are more muted than in the benchmark

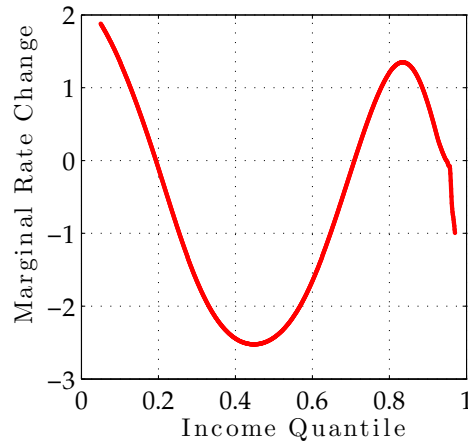


Figure VI.1 : Change in Marginal Taxes.

case, the broad policy prescription of modest marginal tax reductions over a broad band of lower-middle income quantiles combined with an increase over higher quantiles emerges as a robust finding.

## VII. INTRA-OCCUPATIONAL TASK VARIATION

The O\*NET database collects information on the content of occupations from two sources - occupational analysts and a direct survey of US workers and establishments. The latter permits some unbundling of occupations since those surveyed are asked to assess the knowledge (educational and training) requirements of the occupations with which they are associated. Variation in responses gives an indication of the variety of tasks that might be performed by different workers within the same occupation.<sup>11</sup> The O\*NET reports statistics summarizing the distribution of these responses. From the statistics contained within the latest release of the O\*NET we construct Leik's ordinal variation indices (see, Weisberg (1992));<sup>12</sup> we then average the indices associated with every occupation to create a single index for each. We interpret this index as a measure of disagreement amongst workers and establishments as to the knowledge content of an occupation (with a value of zero indicating complete agreement and a value of one maximal disagreement) and as a proxy for the variety of

<sup>11</sup>Variations in assessments of the skill and ability content of occupations would be preferable, but these assessments are obtained from analysts and are not requested in the establishment survey. At issue is whether the higher levels of training and education thought necessary by some respondents are indicative of higher productivities within a task or whether they are indicative of the performance of more complex tasks within a set of tasks defining the occupation. Our stance is that they are at least partly the latter.

<sup>12</sup>This index provides a measure of variation for ranked, categorical variables. The index is equal to zero for a degenerate distribution and assumes the maximum value of one for a polarized distribution with equal weight on the two extremal categories.

(complexity-ranked) tasks within an occupation. In Figure VII.1, we plot this (Lowess-smoothed) ‘disagreement’ index across all ranked occupations  $v$ . As

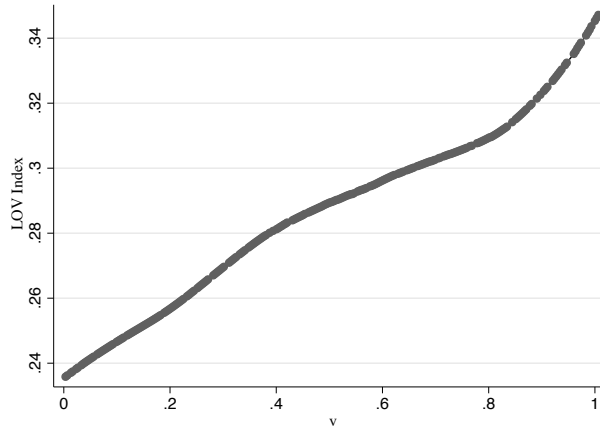


Figure VII.1. : Average Leik’s ordinal variation (LOV) index concerning the knowledge requirements of occupations.

the figure indicates, this index is greater than zero across all  $v$ 's. This suggests a degree of dispersion and disagreement in the survey answers. In addition, the value is increasing in  $v$  - the regression coefficient on non-smoothed data is statistically significant with value is .092 (.013). Overall, these results are consistent with intra-occupational, complexity-ranked task variety that is increasing in the occupation’s average wage. This, in turn, suggests that variation in average occupational wages is a lower bound for wage variation across (complexity-ranked) tasks, especially at the upper end of the wage distribution.

#### References

- Acemoglu, Daron, and David Autor.** 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” In *Handbook of Labor Economics*. Vol. 4B, , ed. David Card and Orley Ashenfelter, 1043–1171. Elsevier Press.
- Autor, David H, Frank Levy, and Richard J Murnane.** 2003. “The Skill Content of Recent Technological Change: An Empirical Exploration.” *Quarterly Journal of Economics*, 118(4): 1279–1333.
- Bonnans, J., and A. Shapiro.** 2000. *Perturbation Analysis of Optimization Problems*. Springer-Verlag: Springer Series in Operations Research.
- Heathcote, Jonathan, Fabrizio Perri, and Giovanni 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.



- Lockwood, Benjamin B, Charles G Nathanson, and E Glen Weyl.** 2014. "Taxation and the Allocation of Talent." Mimeo, Harvard University and University of Chicago.
- McDonald, James B.** 1984. "Some generalized functions for the size distribution of income." *Econometrica*, 52(3): 647–663.
- McFadden, Daniel, Peter Diamond, and Miguel Rodriguez.** 1978. "Measurement of the Elasticity of Factor Substitution and the Bias of Technical Change." In *Production Economics: A Dual Approach to Theory and Applications.* , ed. Melvyn Fuss and Daniel McFadden. Amsterdam: Elsevier.
- Rothschild, C., and F. Scheuer.** 2014. "A Theory of Income Taxation under Multi-dimensional Skill Heterogeneity." NBER WP 19822.
- Rothschild, Casey, and Florian Scheuer.** 2013. "Redistributive Taxation in the Roy model." *Quarterly Journal of Economics*, 128(2): 623–662.
- Singh, S., and G. Maddala.** 1976. "A function for size distribution of incomes." *Econometrica*, 44: 976–983.
- Weisberg, Herbert F.,** ed. 1992. *Central Tendency and Variability.* . 0 ed., SAGE Publications, Inc.