Online Appendix
Supply, Demand, Institutions, and Firms

Daniel Haanwinckel
UCLA and NBER

May 29, 2023

A Proofs

Section 4: Task-based production function

Proof of Lemma 1: Allocation is assortative and labor constraints bind

I proceed by proving two lemmas that, together, imply the desired result. I use the term *candidate solution* to refer to tuples of output and schedules $\left\{q,\{m_h\}_{h=1}^{H}\right\}$ that satisfy all constraints in the assignment problem.

Lemma 4. *If there exists a candidate solution $\left\{q,\{m_h(\cdot)\}_{h=1}^{H}\right\}$ such that one can find two tasks $x_1 < x_2$ and two worker types $h_1 < h_2$ with $m_{h_1}(x_2) > 0$ and $m_{h_2}(x_1) > 0$, then there exists an alternative candidate solution $\left\{q',\{m'_h(\cdot)\}_{h=1}^{H}\right\}$ that achieves the same output ($q = q'$) but has a slack of labor of type h_1 ($l_{h_1} > \int_0^\infty m'_{h_1}(x)dx$).*

Proof. Let $\Delta = x_2 - x_1$ and pick $\tau \in (0, \min\{m_{h1}(x_2), m_{h2}(x_1)e_{h_2}(x_1+\Delta)/e_{h_1}(x_1+\Delta)\})$. Because $m_h(\cdot)$ is right continuous and the efficiency functions $e_h(\cdot)$ are strictly positive and continuous, I can find $\delta > 0$ such that $m_{h_1}(x) > \tau\ \forall x \in [x_2, x_2+\delta)$ and $m_{h2}(x_1)e_{h_2}(x_1+\Delta)/e_{h_1}(x_1+\Delta) > \tau\ \forall x \in [x_1, x_1+\delta)$.

Now construct $\left\{q',\{m'_h(\cdot)\}_{h=1}^{H}\right\}$ identical to $\left\{q,\{m_h(\cdot)\}_{h=1}^{H}\right\}$, except for:

$$m'_{h_1}(x) = m_{h_1}(x) - \tau, \qquad x \in [x_2, x_2+\delta)$$

$$m'_{h_2}(x) = m_{h_2}(x) + \tau\frac{e_{h_1}(x)}{e_{h_2}(x)}, \qquad x \in [x_2, x_2+\delta)$$

$$m'_{h_2}(x) = m_{h_2}(x) - \tau \frac{e_{h_1}(x+\Delta)}{e_{h_2}(x+\Delta)}, \qquad x \in [x_1, x_1+\delta)$$

$$m'_{h_1}(x) = m_{h_1}(x) + \tau \frac{e_{h_1}(x+\Delta)}{e_{h_2}(x+\Delta)} \frac{e_{h_2}(x)}{e_{h_1}(x)}, \qquad x \in [x_1, x_1+\delta)$$

I need to prove that $\left\{q', \{m'_h(\cdot)\}_{h=1}^{H}\right\}$ satisfies all constraints in the assignment problem and has a slack of labor h_1, and that $m'_h(\cdot) \in RC$. Starting with the latter, note that $m'_h(\cdot)$ is always identical to $m_h(\cdot)$ except in intervals of the form $[a,b)$. In those intervals, $m'_h(\cdot)$ is a continuous transformation of $m_h(\cdot)$. So, because $m_h(\cdot)$ is right continuous, so is $m'_h(\cdot)$. In addition, $m'_h(x) > 0 \ \forall x \in \mathbb{R}_{>0}$ by the condition imposed when defining δ. So $m'_h(\cdot) \in RC$.

Next, the blueprint constraints are satisfied under the new candidate solution because second and fourth rows increase task production of particular complexities in a way that exactly offsets decreased production due to the first and third rows, respectively. Total labor use of type h_2 is identical under both allocations, because the additional assignment in the second row is offset by reduced assignment in the third row. Finally, decreased use of labor type h_1 follows from log-supermodularity of the efficiency functions, which guarantees that the term multiplying τ in the fourth row is strictly less than one. So labor added in that row is strictly less than labor saved in the first row. □

Lemma 5. *Any candidate solution with slack of labor is not optimal.*

Proof. Consider two cases:

If there is slack of labor of the highest type, $h = H$: By the feasibility condition in the definition of blueprints, $u_H = \int_0^\infty b(x)/e_H(x)dx$ is finite. Denote the slack of labor of type H in the original candidate solution by $S_H = l_H - \int_0^\infty m_H(x)dx$. Now consider an alternative candidate solution with $q' = q + S_H/u_H$, $m'_H(x) = m_H(x) + (S_H/u_H)b(x)/e_H(x)$, and $m'_h(\cdot) = m_h(\cdot) \ \forall h < H$. That candidate solution satisfies all constraints and achieves a strictly higher level of output. Thus, the original candidate solution is not optimal.

Otherwise: Then there is a positive slack $S_h = l_h - \int_0^\infty m_h(x)dx$ for some $h < H$, and no slack of type H. I will show that it is possible to construct an alternative allocation with the same output and positive slack of labor type H. Using that alternative allocation, one can invoke the first part of this proof to construct a third allocation with higher output.

Remember that the domain of f imposes $l_H > 0$. Because there is no slack of labor H, there must be some $\underline{x}$ with $m_H(\underline{x}) > 0$. Pick an arbitrarily small $\tau > 0$. By right continuity of m_H, there is a small enough $\delta > 0$ such that $m_H(x) > \tau\ \forall x \in [\underline{x}, \underline{x} + \delta)$. Let $\tilde{u}_h = \int_{\underline{x}}^{x+\delta} e_H(x)/e_h(x)dx < \infty$ and define $g = \min\{\tau, S_h/\tilde{u}_h\}$.

Now consider an alternative candidate solution identical to the original one, except that $m'_H(x) = m_H(x) - g$ in the interval $[\underline{x}, \underline{x} + \delta)$ and $m'_h(x) = m_h(x) + g e_H(x)/e_h(x)$ in the same interval. The new candidate solution satisfies all constraints, has right continuous and non-negative assignment functions, and has slack of labor of type H. □

Proof of Lemma 1, except non-arbitrage condition. From Lemma 5, we know that any optimal solution must not have any slack. The same Lemma implies that any candidate solution satisfying the conditions in Lemma 4 is also not optimal. So any optimal solution must be such that for any two tasks $x_1 < x_2$ and two types $h_1 < h_2$, $m_{h_2}(x_1) > 0 \Rightarrow m_{h_1}(x_2) = 0$ and $m_{h_1}(x_2) > 0 \Rightarrow m_{h_2}(x_1) = 0$. This property can be re-stated as: for any pair of types $h_1 < h_2$, there exists at least one number ${}_{h_1}\bar{x}_{h_2}$ such that $m_{h_2}(x) = 0\ \forall x < {}_{h_1}\bar{x}_{h_2}$ and $m_{h_1}(x) = 0\ \forall x > {}_{h_1}\bar{x}_{h_2}$. By combining all such requirements together, there must be $H-1$ numbers $\bar{x}_1, \ldots, \bar{x}_{H-1}$ such that, for any type h, $m_h(x) = 0\ \forall x \notin [\bar{x}_{h-1}, \bar{x}_h]$ (where $\bar{x}_0 = 0$ and $\bar{x}_H = \infty$ are introduced to simplify notation).

Because there is no overlap in types that get assigned to any task (except possibly at the thresholds), the blueprint constraint implies that $m_h(x) = b(x)/e_h(x)\ \forall x \in (\bar{x}_{h-1}, \bar{x}_h)$. Right continuity of assignment functions means that the thresholds must be assigned to the type on the right.

It remains to be shown that the thresholds are unique and non-decreasing. To see that, recall that $b(x) > 0$ and $e_h(x) > 0\ \forall h$. Now start from type $h = 1$ and note that the integral $\int_0^{\bar{x}_1} m_1(x)dx = \int_0^{\bar{x}_1} b(x)/e_1(x)dx$ is strictly increasing in $\bar{x}_1$. Thus, there is only one possible $\bar{x}_1 \geq 0$ consistent with full labor use of type 1. One can then proceed by induction, showing that for any type $h > 1$, the thresholds $\bar{x}_h$ is greater than $\bar{x}_{h-1}$ and unique, for the same reason as in the base case.

Proof of the non-arbitrage condition (Equation 2) is provided in the next section of this Appendix. □

Proposition 1, curvature of the production function: formulas for elasticities and proofs (including Equation 2)

Elasticities: I denote by $c = c(\boldsymbol{w}, q)$ the cost function, use subscripts to denote derivatives regarding input quantities or prices, and omit arguments in functions to simplify the expressions. Then, for any pair of worker types h, h' with $h < h'$:

$$\frac{c c_{h,h'}}{c_h c_{h'}} = \begin{cases} \dfrac{\rho_h}{s_h s_{h'}} & \text{if } h' = h+1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{(Allen partial elasticity of substitution)}$$

$$\frac{f f_{h,h'}}{f_h f_{h'}} = \sum_{\mathfrak{h}=1}^{H-1} \xi_{h,h',\mathfrak{h}} \frac{1}{\rho_{\mathfrak{h}}} \qquad \text{(Hicks partial elasticity of complementarity)}$$

$$\text{where} \quad \rho_h = b_g(\bar{x}_h) \frac{f_h}{e_h(\bar{x}_h)} \left[\frac{d}{d\,\bar{x}_h} \ln\left(\frac{e_{h+1}(\bar{x}_h)}{e_h(\bar{x}_h)} \right) \right]^{-1}$$

$$\xi_{h,h',\mathfrak{h}} = \left(\mathbf{1}\{h \geq \mathfrak{h}+1\} - \textstyle\sum_{k=\mathfrak{h}+1}^{H} s_k \right) \left(\mathbf{1}\{\mathfrak{h} \geq h'\} - \textstyle\sum_{k=1}^{\mathfrak{h}} s_k \right)$$

$$\text{and} \quad s_h = \frac{f_h l_h}{f} = \frac{c_h l_h}{c}$$

Proofs: Constant returns to scale and concavity follow easily from the definition of the production function. Let's start with concavity. Suppose that there are two input vectors $\boldsymbol{l}^1$ and $\boldsymbol{l}^2$, achieving output levels q^1 and q^2 using optimal assignment functions m_h^1 and m_h^2, respectively. Now take $\alpha \in [0,1]$. Given inputs $\bar{\boldsymbol{l}} = \alpha \boldsymbol{l}^1 + (1-\alpha)\boldsymbol{l}^2$, one can use assignment functions defined by $\bar{m}_h(x) = \alpha m_h^1(x) + (1-\alpha) m_h^2(x)\ \forall x, h$ to achieve output level $\bar{q} = \alpha q^1 + (1-\alpha)q^2$, while satisfying blueprint and labor constraints. So $f(\bar{\boldsymbol{l}}, b) \geq \bar{q}$. For constant returns, note that, given $\alpha > 1$, output αq^1 is attainable with inputs $\alpha \boldsymbol{l}^1$ by using assignment functions $\alpha m_h^1(x)$. Together with concavity, that implies constant returns to scale.

Lemma 1 implies that, given inputs $(\boldsymbol{l}, b_g(\cdot))$, the optimal thresholds and the optimal production level satisfy the set of H labor constraints with equality. I will now prove results that justify using the implicit function theorem on that system of equations. That will prove twice differentiability and provide a path to obtain elasticities of complementarity and substitution.

Definition 4. *The excess labor demand function* $\boldsymbol{z} : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{>0} \to \mathbb{R}^H$ *is given by:*

$$z_h(q, \bar{x}_1, \ldots, \bar{x}_{H-1}; \boldsymbol{l}) = q \int_{\bar{x}_{h-1}}^{\bar{x}_h} \frac{b_g(x)}{e_h(x)} dx - l_h$$

Lemma 6. *The excess labor demand function is* C^2.

Proof. We need to show that, for all components $z_h(\cdot)$, the second partial derivatives exist and are continuous. This is immediate for the first derivatives regarding q and $\boldsymbol{l}$, as well as for their second own and cross derivatives (which are all zero).

The first derivative regarding threshold $\bar{x}_{h'}$ is:

$$\frac{\partial z_h(\cdot)}{\partial \bar{x}_{h'}} = q\left[\mathbf{1}\left\{h'=h\right\}\frac{b_g(\bar{x}_h)}{e_h(\bar{x}_h)} - \mathbf{1}\left\{h'=h-1\right\}\frac{b_g(\bar{x}_h)}{e_{h+1}(\bar{x}_h)}\right]$$

Because blueprints and efficiency functions are continuously differentiable and strictly positive, this expression is continuously differentiable in $\bar{x}_h$. The cross-elasticities regarding q and $\boldsymbol{l}$ also exist and are continuous. □

Lemma 7. *The Jacobian of the excess labor demand function regarding $(q,\bar{x}_1,\ldots,\bar{x}_{H-1})$, when evaluated at a point where $\boldsymbol{z}(\cdot) = \mathbf{0}_{H\times 1}$, has non-zero determinant.*

Proof. The Jacobian, when evaluated at the solution to the assignment problem, is:

$$J = \begin{bmatrix} \frac{l_1}{q} & q\frac{b_g(\bar{x}_1)}{e_1(\bar{x}_1)} & 0 & 0 & \cdots & 0 & 0 \\ \frac{l_2}{q} & -q\frac{b_g(\bar{x}_1)}{e_2(\bar{x}_1)} & q\frac{b_g(\bar{x}_2)}{e_2(\bar{x}_2)} & 0 & \cdots & 0 & 0 \\ \frac{l_3}{q} & 0 & -q\frac{b_g(\bar{x}_2)}{e_3(\bar{x}_2)} & q\frac{b_g(\bar{x}_3)}{e_3(\bar{x}_3)} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{l_{H-1}}{q} & 0 & 0 & 0 & \cdots & -q\frac{b_g(\bar{x}_{H-2})}{e_{H-1}(\bar{x}_{H-2})} & q\frac{b_g(\bar{x}_{H-1})}{e_{H-1}(\bar{x}_{H-1})} \\ \frac{l_H}{q} & 0 & 0 & 0 & \cdots & 0 & -q\frac{b_g(\bar{x}_{H-1})}{e_H(\bar{x}_{H-1})} \end{bmatrix}$$

The determinant is:

$$|J| = (-1)^{H+1} q^{H-2}\left[\prod_{h=1}^{H-1}\frac{b_g(\bar{x}_h)}{e_{h+1}(\bar{x}_h)}\right]\sum_{h=1}^{H}\left(l_h\prod_{i=2}^{h}\frac{e_i(\bar{x}_{i-1})}{e_{i-1}(\bar{x}_{i-1})}\right)$$

which is never zero, since $q > 0$ (from feasibility of blueprints and $l_H > 0$) and $b(x), e_h(x) > 0\ \forall x, h$.

□

Lemmas 6 and 7 mean that the implicit function theorem can be used at the solution to the assignment problem to obtain derivatives of the solutions to the system of equations imposed by the labor constraints. These solutions are $q(\boldsymbol{l}) = f(\boldsymbol{l}, b_g(\cdot))$ and $\bar{x}_h(\boldsymbol{l})$. Because $\boldsymbol{z}$ is C^2, so are the production function and the thresholds as functions of inputs.

Obtaining the ratios of first derivatives in Lemma 1 and the elasticities of complementarity and substitution in Proposition 1 is a matter of tedious but straightforward algebra, starting from the implicit function theorem. For the non-arbitrage condition in Lemma 1, a simpler approach is to define the allocation problem in terms of choosing output and thresholds, and then use a Lagrangian to embed the labor constraints into the objective function. Then, the result of Lemma 2, along with the constant returns relationship $q = \sum_h l_h f_h$, emerge as first order conditions, after noting that the Lagrange multipliers are marginal productivities.

When working towards second derivatives, it is necessary to use the derivatives of thresholds regarding inputs. For reference, here is the result:

$$\frac{d\bar{x}_h}{dl_{h'}} = \frac{e_h(\bar{x}_h)}{qb_g(\bar{x}_h)} \frac{f_{h'}}{f_h} \left[\mathbf{1}\left\{h \geq h'\right\} - \sum_{i=1}^{h} s_i \right]$$

One can verify $\frac{d\bar{x}_h}{dl_{h'}} > 0 \Leftrightarrow h \geq h'$. Adding labor "pushes" thresholds to the right or to the left depending on whether the labor which is being added is to the left or to the right of the threshold in question.

Proof of Corollary 1: Distance-dependent complementarity

This is proven by inspecting the sign of the weights $\xi_{h,h',\mathfrak{h}}$ above. When $h = h'$, these terms are negative for all i. Changing h' by one, either up or down, changes one of the $\xi_{h,h',\mathfrak{h}}$ from negative to positive while keeping the others unchanged. So there must be an increase in the elasticity of complementarity since all of the $\rho_{\mathfrak{h}}$ are positive. Every additional increment or decrement of h' away from h involves a similar change of sign in one of the $\xi_{h,h',\mathfrak{h}}$, leading to the same increase in complementarity.

Proof of Lemma 2: Differences in skill intensity, monopsony, and task assignment

We can write the problem of the firm under monopsony as:

$$\pi_j = \max_{\boldsymbol{l}_j} p_g f\left(\boldsymbol{l}_j, b_g\right) - \sum_{h=1}^{H} \omega_h \frac{l_{h,j}^{1+\frac{1}{\beta}}}{L_h^{\frac{1}{\beta}}}$$

Which has first order conditions:

$$p_g f_h \left(\boldsymbol{l}_j, b_g\right) = \frac{\beta+1}{\beta} \omega_h \left(\frac{l_{h,j}}{L_h}\right)^{\frac{1}{\beta}}$$

Taking ratios for $(h+1)/h$, using Equation 2, and introducing the firm-specific task threshold notation:

$$\frac{e_{h+1}\left(\bar{x}_{h,j}\right)}{e_h\left(\bar{x}_{h,j}\right)} = \frac{\omega_{h+1}}{\omega_h} \left(\frac{l_{h+1,j}}{l_{h,j}}\right)^{\frac{1}{\beta}} \left(\frac{L_{h+1,j}}{L_{h,j}}\right)^{-\frac{1}{\beta}} \quad h \in \{1, \ldots, H-1\} \tag{11}$$

The desired result follows from the comparative advantage assumption, making the task threshold $\bar{x}_{h,j}$ increasing in $l_{h+1,j}/l_{h,j}$ if all firms face the same supply parameters.

Proof of Proposition 2: Complementarity patterns may differ between firms

For firms producing $g = 1$, the production function is $f(\boldsymbol{l}, b_1) = \sum_{h=1}^{H} l_h e_h(0)$, since each unit measure of tasks $x = 0$ corresponds to one unit of output. Using the first order condition of problem of the firm under monopsony (from the previous proof), we find:

$$p_g e_h(0) = \frac{\beta+1}{\beta} \omega_h \left(\frac{l_{h,j}}{L_h}\right)^{\frac{1}{\beta}} \quad \forall h$$

From here, it is clear that there is no change in employment for any $h \neq 1$. For $h = 1$, because the left-hand side is invariant in this partial equilibrium exercise, $l_{1,j}$ changes proportionately to L_1, such that the ratio $l_{1,j}/L_1$ remains invariant—and thus, the posted wage $w_{h,j}$ does not change either.

For firms producing $g = 2$, it is sufficient to show that all task thresholds move to the right following an increase in L_1. To see that, plug the labor supply expression into Equation 11 to find a monotonic link between posted wages and task thresholds:

$$\frac{e_{h+1}\left(\bar{x}_{h,j}\right)}{e_h\left(\bar{x}_{h,j}\right)} = \frac{w_{h+1,j}}{w_{h,j}}$$

Rewrite Equation 11 with task thresholds as the only endogenous variables (note that when the labor choices are divided, the choice of quantity cancels out):

$$\frac{e_{h+1}\left(\bar{x}_{h,j}\right)}{e_h\left(\bar{x}_{h,j}\right)} = \frac{\omega_{h+1}}{\omega_h}\left(\frac{\int_{\bar{x}_{h,j}}^{\bar{x}_{h+1,j}} \frac{b_g(x)}{e_{h+1}(x)}dx}{\int_{\bar{x}_{h-1,j}}^{\bar{x}_{h,j}} \frac{b_g(x)}{e_h(x)}dx}\right)^{\frac{1}{\beta}} \left(\frac{L_{h+1,j}}{L_{h,j}}\right)^{-\frac{1}{\beta}} \qquad h \in \{1,2\}$$

If we take logs and implicitly differentiate with respect to $\log L_1$, we find:

$$\frac{d\bar{x}_{1,j}}{d\log L_1} = \frac{1 + \frac{d\bar{x}_{2,j}}{d\log L_1}\frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}}{\beta\left[\frac{e_1\left(\bar{x}_{1,j}\right)}{e_2\left(\bar{x}_{1,j}\right)}\right]\frac{d}{d\bar{x}_{1,j}}\left[\frac{e_2\left(\bar{x}_{1,j}\right)}{e_1\left(\bar{x}_{1,j}\right)}\right] + \frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})} + \frac{b_g(\bar{x}_{1,j})}{l_1 e_1(\bar{x}_{1,j})}}$$

$$\frac{d\bar{x}_{2,j}}{d\log L_1} = \frac{\frac{d\bar{x}_{1,j}}{d\log L_1}\frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})}}{\beta\left[\frac{e_2\left(\bar{x}_{2,j}\right)}{e_3\left(\bar{x}_{2,j}\right)}\right]\frac{d}{d\bar{x}_{2,j}}\left[\frac{e_3\left(\bar{x}_{2,j}\right)}{e_2\left(\bar{x}_{2,j}\right)}\right] + \frac{b_g(\bar{x}_{2,j})}{l_3 e_3(\bar{x}_{2,j})} + \frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}}$$

The comparative advantage assumption implies that the derivatives of efficiency ratios are positive. Thus, all individual terms in those expressions are positive, the second equation implies that both thresholds move in the same direction. Tedious but straightforward algebra shows that they move to the right if and only if:

$$\beta\left[\frac{e_1\left(\bar{x}_{1,j}\right)}{e_2\left(\bar{x}_{1,j}\right)}\right]\frac{d}{d\bar{x}_{1,j}}\left[\frac{e_2\left(\bar{x}_{1,j}\right)}{e_1\left(\bar{x}_{1,j}\right)}\right] + \frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})} + \frac{b_g(\bar{x}_{1,j})}{l_1 e_1(\bar{x}_{1,j})} > \frac{\frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})}\frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}}{\beta\left[\frac{e_2\left(\bar{x}_{2,j}\right)}{e_3\left(\bar{x}_{2,j}\right)}\right]\frac{d}{d\bar{x}_{2,j}}\left[\frac{e_3\left(\bar{x}_{2,j}\right)}{e_2\left(\bar{x}_{2,j}\right)}\right] + \frac{b_g(\bar{x}_{2,j})}{l_3 e_3(\bar{x}_{2,j})} + \frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}}$$

This expression is always true. To see why, note that the right-hand size is bounded above by one of the terms on the left-hand side:

$$\frac{\frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})}\frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}}{\beta\left[\frac{e_2\left(\bar{x}_{2,j}\right)}{e_3\left(\bar{x}_{2,j}\right)}\right]\frac{d}{d\bar{x}_{2,j}}\left[\frac{e_3\left(\bar{x}_{2,j}\right)}{e_2\left(\bar{x}_{2,j}\right)}\right] + \frac{b_g(\bar{x}_{2,j})}{l_3 e_3(\bar{x}_{2,j})} + \frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}} < \frac{\frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})}\frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}}{\frac{b_g(\bar{x}_{2,j})}{l_2 e_2(\bar{x}_{2,j})}} = \frac{b_g(\bar{x}_{1,j})}{l_2 e_2(\bar{x}_{1,j})}$$

Section 5: Markets and wages

Proofs in this section are written for a more general version of the model with heterogeneous non-wage amenities at the firm level, denoted by a_j and with good-specific averages $\bar{a}_g$. That general version is described in Appendix B.2 below.

Proof of Lemma 3: Firm problem and representative firms

I start by establishing that the solution must have positive employment of all types. The marginal product of an efficiency unit of labor of the highest type is bounded below by $1/\int_0^\infty b_g(x)/e_H(x)dx = \underline{f}_H$, which is strictly positive due to the feasibility condition imposed on blueprints. Consider the strategy of posting a fixed payment $y_{Hj}(\varepsilon) = \bar{y} \geq \underline{y}$ to all workers with $\varepsilon > \underline{\varepsilon}_{Hj}$. Profit from workers of type H associated with that strategy are bounded below by $\int_{\underline{\varepsilon}_{Hj}}^\infty N_H a_j \bar{y}^\beta/\omega_H(\varepsilon)^\beta r_H(\varepsilon)(p_g \underline{f}_H \varepsilon - \bar{y})d\varepsilon$. That expression is assured to be positive for high enough $\underline{\varepsilon}_{Hj}$ (note that $\omega_h(\varepsilon)$ is always finite in an equilibrium). Thus, positive employment of skilled workers following that strategy is more profitable than not employing any of those workers.

A positive amount of l_H ensures that all other types are employed as well. Consider a particular type $h < H$ and whether it is optimal to set $l_h = 0$, fixing employment of all other types. Because $l_H > 0$, $\bar{x}_{H-1}$ is finite, and thus threshold $\bar{x}_h$ (the highest task performed by h) is guaranteed to be finite as well. Then, from Equation 2, the marginal product of type h is bound below by $\underline{f}_H e_h(\bar{x}_{H-1})/e_H(\bar{x}_{H-1})$. A similar reasoning as above establishes that employing small quantities of labor h is more profitable than setting $l_h = 0$.

The rest of the proof follows from the logic described in the text. The threshold $\underline{\varepsilon}_{hj}$ is chosen so that the worker with the least amount of efficiency units pays for himself, bringing in revenue equal to the minimum wage. Below that, labor payments — which are bound by the minimum wage — will necessarily exceed marginal revenue from those workers. For every $\varepsilon > \underline{\varepsilon}_{hj}$, the firm chooses $y_{hj}(\varepsilon)$ by equating marginal revenue from workers of that (h,ε) combination with their marginal cost. For high enough ε, that leads to the constant markdown rule, implying that earnings are proportional to marginal product of labor — and thus linear in ε. Workers close to the cutoff are still profitable, but for them, the minimum wage constraint binds.

To see why these solutions do not depend on amenities, such that there is a representative firm for each good g, first note that a_j is a multiplicative term in both $C_h\left(y_{hj}, \underline{\varepsilon}_{hj}, a_j\right)$ and $l_h\left(y_{hj}, \underline{\varepsilon}_{hj}, a_j\right)$. Now remember that the task-based production function has constant returns to scale. Thus, the profit function can be rewritten as $\pi(a_j) = a_j\pi(1)$. Amenities scale up employment and production while keeping average labor costs constant.

Proof of Proposition 3: Wage differentials across firms

I start by proving a useful Lemma that shows how proportional terms dividing task requirements can be interpreted as physical productivity shifters.

Lemma 8. *If $b_g(x) = b(x)/z_g$ for a blueprint $b(\cdot)$ and scalar $z_g > 0$, then $f(\boldsymbol{l}, b_g(\cdot)) = z_g f(\boldsymbol{l}, b(\cdot))$.*

Proof. Plug $b_g(x) = b(x)/z_g$ into the assignment problem defining the task-based production function. Change the choice variable to $q' = q/z_g$. The z_g terms in the task constraint cancel each other and the maximand changes to $z_g q'$. The result follows from noting that $\max_{\{\cdot\}} z_g q' = z_g \max_{\{\cdot\}} q'$ and that the resulting value function is $f(\boldsymbol{l}, b(\cdot))$ by definition. □

Now I proceed to the proof of each statement of Proposition 3 separately.

Proof of part 1: From Lemma 8, $f_h(\boldsymbol{l}, b_g(\cdot)) = z_g f_h(\boldsymbol{l}, b(\cdot))$. Also note $\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, \bar{a}_g) = \bar{a}_g \boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, 1)$ and $\boldsymbol{C}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, \bar{a}_g) = \bar{a}_g \boldsymbol{C}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, 1)$, and remember that the task-based production function has constant returns to scale (and so marginal productivities are homogeneous of degree zero). Now let $\tilde{F} = F_g/\bar{a}_g$ and rewrite the first order conditions of the firm (7), (8) and the zero profits condition (10) imposing the conditions from this proposition:

$$
\begin{aligned}
p_g z_g f_h(\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, 1), b(\cdot)) \exp(\underline{\varepsilon}_{hg}) &= \underline{y} && \forall h, g \\
p_g z_g f_h(\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, 1), b(\cdot)) \frac{\beta}{\beta+1} &= w_{hg} && \forall h, g \\
\bar{a}_g \left[p_g z_g f(\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, 1), b(\cdot)) - \sum_{h=1}^{H} C_h(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, 1) \right] &= \bar{a}_g \tilde{F} && \forall g
\end{aligned}
$$

To see that these equations imply a representative firm for the economy, plug in $\underline{\boldsymbol{\epsilon}}_g = \underline{\boldsymbol{\epsilon}}$, $\boldsymbol{w}_g = \boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_H\}$, and $p_g = p/z_g$ for common $\underline{\boldsymbol{\epsilon}}$, $\boldsymbol{\lambda}$, and p. All dependency on g is eliminated, showing that the solution of the problem of the firm is the same for all firms in the economy and that prices are inversely proportional to physical productivity shifters z_g (such that marginal revenue product of labor is equalized across firms). □

Proof of part 2: Without a minimum wage, there is no motive for a cutoff rule: $\underline{\varepsilon}_{hg} = 0$. In addition, the labor supply curve becomes isoelastic with identical elasticities for all worker types:

$$l_h\left(w_{hg},\cdot,\bar{a}_g\right)=\bar{a}_g\left(\frac{w_{hg}}{\omega_h}\right)^{\beta}$$
$$C_h\left(w_{hg},\cdot,\bar{a}_g\right)=w_{hg}l_h\left(w_{hg},\cdot,\bar{a}_g\right)$$
$$\text{where } \omega_h=\left(\sum_g J_g\bar{a}_g w_{hg}^{\beta}\right)^{\frac{1}{\beta}}$$

Rewrite the first order conditions on wages as in the proof of part 1 above:

$$p_g z_g f_h\left(\boldsymbol{l}\left(\boldsymbol{w}_g,\cdot,1\right),b(\cdot)\right)\frac{\beta}{\beta+1}=w_{hg} \qquad \forall h,g$$

Also, rewrite the zero profit condition as:

$$\begin{aligned}F_g&=p_g z_g f\left(\boldsymbol{l}\left(\boldsymbol{w}_g,\cdot,\bar{a}_g\right),b(\cdot)\right)-\sum_{h=1}^{H}C_h\left(\boldsymbol{w}_g,\cdot,\bar{a}_g\right)\\&=p_g z_g\sum_{h=1}^{H}l_h\left(w_{hg},\cdot,\bar{a}_g\right)f_h\left(\boldsymbol{l}\left(\boldsymbol{w}_g,\cdot,1\right),b(\cdot)\right)-\sum_{h=1}^{H}w_{hg}l_h\left(w_{hg},\cdot,\bar{a}_g\right)\end{aligned}$$

I claim that $\boldsymbol{w}_g=(F_g/\bar{a}_g)^{1/(\beta+1)}\boldsymbol{\lambda}$ for some vector $\boldsymbol{\lambda}=\{\lambda_1\ldots,\lambda_H\}$. From the labor supply equation, that implies $l_{hg}=F_g^{\beta/(\beta+1)}\bar{a}_g^{1/(\beta+1)}\ell_h$, where $\ell_h=\omega_h^{-\beta/(\beta+1)}$. Plugging these expressions in the rewritten zero profit condition yields $\sum_h \ell_h\lambda_h=1\ \forall g$, showing that the claim does not contradict optimal entry behavior; instead, optimal entry merely imposes a normalization on the $\boldsymbol{\lambda}$ vector.

The corresponding prices that lead to zero profits are:

$$\begin{aligned}\Rightarrow p_g&=\frac{(\beta+1)F_g}{z_g f\left(\boldsymbol{l}\left(\boldsymbol{w}_g,\cdot,\bar{a}_g\right),b(\cdot)\right)}\\&=\frac{\beta+1}{z_g f\left(\boldsymbol{\ell},b(\cdot)\right)}\left(\frac{F_g}{\bar{a}_g}\right)^{\frac{1}{\beta+1}}\end{aligned}$$

Finally, plugging these results into the first order conditions yields:

$$f_h\left(\boldsymbol{\ell},b\right)\beta=\lambda_h \qquad \forall h,g$$

Which again has no dependency on g, showing that the claimed solution solves the problem for all firms. □

Proof of part 3: Under the conditions from this part, labor supply curves are isoelastic, as shown in the proof of part 2 above. It is easily shown, using that isoelastic expression for $l_h(\cdot)$, that:

$$\left(\frac{w_{h'g'}}{w_{hg'}}\right) \Big/ \left(\frac{w_{h'g}}{w_{hg}}\right) = \left[\left(\frac{l_{h'g'}}{l_{hg'}}\right) \Big/ \left(\frac{l_{h'g}}{l_{hg}}\right)\right]^{\frac{1}{\beta}}$$

Under the condition imposed on labor input ratios, the right hand side is positive. The proof follows from noting that the desired ratio of earnings is equal to the ratio of wages in the left hand side.

□

Proof of Proposition 4: Supply shocks

For notational simplicity, in this proof we set p_1 as the numeraire, so $p_2/p_1 = p_2$. The proof proceeds in two parts. First, we will obtain an expression for the skill wage premium as a function of p_2 and model parameters, so that the main result can be derived. Next, we obtain the expression that pins down p_2 to prove that it is decreasing in L_2/L_1.

From the constant mark-down rule and the fact that blueprints are degenerate:

$$w_{h,1} = \frac{\beta}{\beta+1} e_h(x_1) \qquad\qquad w_{h,2} = \frac{\beta}{\beta+1} e_h(x_2) p_2$$

To obtain the shares $s_{h,g}$ as functions of p_2, start with optimal firm creation, which implies that profits per firm must be proportional to entry costs; coupled with the fact that with no minimum wage, profits are proportional to revenues:

$$\frac{q_1}{F_1} = \frac{q_2 p_2}{F_2}$$

Next, optimal consumption implies:

$$\frac{Q_2}{Q_1} = \frac{q_2 J_2}{q_1 J_1} = \left(\frac{\gamma_2}{\gamma_1}\frac{1}{p_2}\right)^{\sigma}$$

Combining both expressions:

$$\frac{J_2}{J_1} = \left(\frac{\gamma_2}{\gamma_1}\right)^{\sigma} \frac{F_1}{F_2} p_2^{1-\sigma}$$

Now we are ready to derive expressions for employment shares:

$$\begin{aligned}
s_{h,1} =& \frac{J_1 w_{h,1}^{\beta}}{J_1 w_{h,1}^{\beta} + J_2 w_{w,2}^{\beta}} \\
=& \left[1 + \frac{J_2}{J_1}\left(\frac{w_{h,2}}{w_{h,1}}\right)^{\beta}\right]^{-1} \\
=& \left[1 + \left(\frac{\gamma_2}{\gamma_1}\right)^{\sigma} \frac{F_1}{F_2} p_2^{1-\sigma} \left(\frac{e_h(x_2) p_2}{e_h(x_1)}\right)^{\beta}\right]^{-1} \\
=& \left[1 + \left(\frac{\gamma_2}{\gamma_1}\right)^{\sigma} \frac{F_1}{F_2} \left(\frac{e_h(x_2)}{e_h(x_1)}\right)^{\beta} p_2^{\beta+1-\sigma}\right]^{-1}
\end{aligned}$$

and $s_{h,2} = 1 - s_{h,1}$.

Neither the employment shares nor wages depend on L_h directly. So, the effects of supply shocks on the mean log wage gap are fully mediated by p_2. This result is specific to the case with degenerate blueprints. It simplifies the analytical solution of the model and helps isolate the role of general equilibrium effects through prices and firm entry.

Then, to obtain the first price of the proposition, one just needs to combine the expressions above to write the mean log wage gap and differentiate it with respect to $\log p_2$. This is simple once one notes that the elasticity of $s_{h,2}$ with respect to p_2 is $(\beta + 1 - \sigma)s_{h,1}$.

Finally, we need to prove that p_2 is decreasing in L_2/L_1. To do that, we will use an expression linking aggregate production to aggregate consumption (in ratios), which only depends on p_2 and model parameters:

$$\left(\frac{\gamma_2}{\gamma_1}\frac{1}{p_2}\right)^{\sigma} = \frac{L_1 s_{1,2} e_1(x_2) + L_2 s_{2,2} e_2(x_2)}{L_1 s_{1,1} e_1(x_1) + L_2 s_{2,1} e_2(x_1)}$$

where, once again, the assumption of degenerate blueprints helps with tractability.

After careful manipulations, this expression can be rewritten as:

$$\frac{L_2}{L_1} = \frac{\frac{e_1(x_1)}{F_1} - \frac{e_1(x_2)}{F_2}\left[\frac{e_1(x_2)}{e_1(x_1)}\right]^{\beta} p_2^{1+\beta}}{\frac{e_2(x_2)}{F_2} - \frac{e_2(x_1)}{F_1}\left[\frac{e_2(x_1)}{e_2(x_2)}\right]^{\beta} p_2^{-1-\beta}} \left[\frac{e_2(x_1)}{e_2(x_2)}\right]^{\beta} p_2^{-1-\beta} \frac{\frac{\gamma_1^{\sigma}}{F_1} + \frac{\gamma_2^{\sigma}}{F_2}\left[\frac{e_2(x_2)}{e_2(x_1)}\right]^{\beta} p_2^{1+\beta-\sigma}}{\frac{\gamma_1^{\sigma}}{F_1} + \frac{\gamma_2^{\sigma}}{F_2}\left[\frac{e_1(x_2)}{e_1(x_1)}\right]^{\beta} p_2^{1+\beta-\sigma}}$$

To show that p_2 is decreasing in L_2/L_1, we only need to show that the right-hand side of this expression is decreasing in p_2. This is easy to see for all terms except the last fraction. If $\sigma \leq 1+\beta$, one only needs to multiply the standalone $p_2^{-1-\beta}$ and the last numerator to obtain a fraction that is obviously decreasing in p_2. If instead $\sigma > 1+\beta$, then one needs to use the comparative advantage assumption to see that the perm multiplying $p_2^{1+\beta-\sigma}$ in the numerator is larger than the same term in the denominator of that expression. This, coupled with the fact that $1+\beta-\sigma < 0$, is enough to establish that the fraction is decreasing in p_2, given that the first term is the same in both the numerator and the denominator.

Proof of Proposition 5: Changes in firm costs affect the returns to skill

Before proving the Proposition, I derive a Lemma that states that blueprints that are more intensive in complex tasks lead to higher gaps in marginal productivity, holding constant the quantity of labor. This Lemma is conceptually similar to the monotone comparative statics in Costinot and Vogel (2010).

Lemma 9. *Let b and b' denote blueprints such that their ratio $b'(x)/b(x)$ is strictly increasing. Then:*

$$\frac{f_{h+1}(\boldsymbol{l},b')}{f_h(\boldsymbol{l},b')} > \frac{f_{h+1}(\boldsymbol{l},b)}{f_h(\boldsymbol{l},b)} \quad h = 1,\ldots,H-1$$

Proof. Fix $\boldsymbol{l}$, let $q = f(\boldsymbol{l},b)$ and $q' = f(\boldsymbol{l},b')$. Now construct $b''(x) = b'(x)q'/q$. From Lemma 8, it follows that $f(\boldsymbol{l},b'') = q$ and $f_h(\boldsymbol{l},b'') = f_h(\boldsymbol{l},b')$ $\forall h$. I will show that the statement holds for b and b'', and since b'' and b' lead to the same marginal products, the desired result holds.

Because b and b'' lead to the same output given the same vector of inputs, but $b''(x)/b(x)$ is increasing, there must be a task x^* such $b''(x) < b(x)$ $\forall x < x^*$ and $b''(x) > b(x)$ $\forall x > x^*$. To see why they must cross at least once at x^*, suppose otherwise (one blueprint is strictly more than other for all x): there will be a contradiction since task demands are strictly higher for one of the blueprints, but they still lead to the same production q given the same vector of inputs. From this crossing point, differences before and after emerge from the monotonic ratio property.

Now note from the non-arbitrage condition (2) in Lemma 1, along with log-supermodularity of $e_h(x)$, that the statement to be proved is equivalent to

$$\bar{x}'_h \geq \bar{x}_h \quad h \in \{1,...,H-1\}$$

where $\bar{x}'_h$ denotes thresholds under the alternative blueprint b''.

I proceed by using compensated labor demand integrals to show that thresholds differ as stated above. Denote by h^* the type such that $x^* \in [\bar{x}_{h^*-1}, \bar{x}_{h^*})$. The proof will be done in two parts: starting from $\bar{x}'_1$ and ascending by induction up to $\bar{x}_{h^*-1}$, and next starting from $\bar{x}_{h-1}$ and descending by induction down to $\bar{x}_{h^*}$. Note that if $h^* = 1$ or $h^* = H$, only one part is required.

Base case $\bar{x}_1$: The equation for $h = 1$ is $\int_0^{\bar{x}_1} \frac{b(x)}{e_1(x)} dx = \frac{l_1}{q}$ under the original blueprint, and $\int_0^{\bar{x}'_1} \frac{b''(x)}{e_1(x)} dx = \frac{l_1}{q}$ under the new one. Equating the right hand side of both expressions and rearranging yields:

$$\int_{\bar{x}_1}^{\bar{x}'_1} \frac{b''(x)}{e_1(x)} dx = \int_0^{\bar{x}_1} \frac{b(x) - b''(x)}{e_1(x)} dx$$

Since $b(x) \geq b''(x)$ for $x < x^*$, the right-hand side is positive, and then the equality will only hold if $\bar{x}'_1 \geq \bar{x}_1$.

Ascending induction rule: Suppose $\bar{x}'_{h-1} \geq \bar{x}_{h-1}$ and $h < h^*$. I will prove that $\bar{x}'_h \geq \bar{x}_h$. To do so, use the fact that $\frac{l_h}{q}$ is the same under both the old and new blueprints to equate the labor demand integrals, as was done in the base case. This yields the following equivalent expressions:

$$\begin{aligned}
\int_{\bar{x}_h}^{\bar{x}'_h} \frac{b''(x)}{e_h(x)} dx &= \int_{\bar{x}_{h-1}}^{\bar{x}'_{h-1}} \frac{b(x)}{e_h(x)} dx + \int_{\bar{x}'_{h-1}}^{\bar{x}_h} \frac{b(x) - b''(x)}{e_h(x)} dx \\
&= \int_{\bar{x}_{h-1}}^{\bar{x}_h} \frac{b(x)}{e_h(x)} dx + \int_{\bar{x}_h}^{\bar{x}'_{h-1}} \frac{b''(x)}{e_h(x)} dx
\end{aligned}$$

It is enough to show that the expression is positive, ensuring that $\bar{x}'_h \geq \bar{x}_h$. Consider two cases. *If* $\bar{x}'_{h-1} \leq \bar{x}_h$, then use the first expression. The induction assumption guarantees positivity of the first term, and the integrand of the second term is positive because $\bar{x}_h < z^*$. *If instead*

$\bar{x}'_{h-1} > \bar{x}_h$, the second expression is more convenient. There, all integrands are positive and the integration upper bounds are greater than the lower bounds.

Base case $\bar{x}_{H-1}$ and descending induction rule: Those are symmetric to the cases above. □

In a competitive economy, thresholds are the same for all firms. Given total endowments of labor efficiency units $\boldsymbol{L}$ and aggregate demand for tasks $B(x) = Q_1 b_1(x) + Q_2 b_2(x)$ (where Q_g denotes aggregate demand for good g before the shock), wages w_h must be proportional to marginal productivities $f_h(\boldsymbol{L}, B(\cdot))$, because the labor constraints that determine thresholds and marginal productivities in the task-based production function are the labor clearing conditions for this economy.

Aggregate demand for tasks following the shock is $B'(x) = Q'_1 b_1(x) + Q'_2 b_2(x)$. As noted above, wages after the shock are proportional to $f_h(\boldsymbol{L}, B'(\cdot))$. But $B(x, Q'_1, Q'_2)/B(x, Q_1, Q_2)$ is increasing in x if $Q'_2/Q'_1 > Q_2/Q_1$. And an increase in relative taste for good 2, holding all else equal, necessarily implies an increase in aggregate consumption of good 2 relative to good 1. Thus, Lemma 9 implies that wage gaps increase as stated in the Proposition.

Section 6: Wage inequality and sorting in Brazil

Proof of Proposition 6: Identification, estimation, and inference

The goal of this proof is to show that Assumptions 1 through 6, coupled with the smoothness of the economic model (which makes the $a(\cdot)$ function differentiable), imply that the econometric model satisfies standard identification conditions for a parametric nonlinear least squares panel regression. The panel dimension is the region, as there are several different endogenous outcomes by region. Discussion of the identification assumptions in the context of Brazil is left to Appendix D.4.

The non-standard part of the proposed identification strategy is the inversion of region-specific parameters using a subset of the endogenous variables. Assumptions 3 and 4 imply that this condition is satisfied. See Appendix D.4 for a discussion of why invertibility is feasible in the theoretical model. Then, the model to be estimated is the one described in Assumption 5:

$$\boldsymbol{Y}_r = \tilde{a}\left([\boldsymbol{Z}'_r, PB(\boldsymbol{y}_r)']', \boldsymbol{\theta}^G\right) + \boldsymbol{u}_r$$

which is a nonlinear simultaneous equation model where the set of "exogenous" covariates is expanded to include the endogenous outcomes selected by the $PB(\cdot)$ function. The fact that those variables are listed both on the left- and right-hand sides is irrelevant, since for those equations, the error is always zero. Thus, they bear no consequence for the least squares procedure. Alternatively, one could define an equivalent model omitting those equations.

For exogeneity of this model, I need $E\left[\boldsymbol{u}_r|\boldsymbol{Z}_r, PB(\boldsymbol{Y}_r)\right] = 0$. From assumptions 1 and 3, $E\left[\boldsymbol{u}_r|\boldsymbol{Z}_r, \hat{\boldsymbol{\theta}}^R\left(PB(\boldsymbol{Y}_r)|\boldsymbol{Z}_r, \boldsymbol{\theta}_0^G\right)\right] = 0$. Since $\hat{\boldsymbol{\theta}}^R(\cdot)$ is a measurable injective function in the first argument, conditioning on Z_r and $PB(\boldsymbol{Y}_r)$ is the same as conditioning on Z_r and $\hat{\boldsymbol{\theta}}^R\left(PB(\boldsymbol{Y}_r)|\boldsymbol{Z}_r, \boldsymbol{\theta}_0^G\right)$, proving the desired result.

This result, along with assumptions 2, 5, and 6, are standard assumptions for a nonlinear least squares panel model with exogenous covariates, no unobserved heterogeneity, and errors that may have an arbitrary variance-covariance matrix within regions.

B Appendix to the theory

B.1 Definition of the task-based production function

Here, I make two notes about the task-based production function. The first is that the assignment model is very general. The function $m_h(x)$ allows firms to use multiple worker types for the same task, the same worker in disjoint sets of tasks, and discontinuities in assignment rules.

The second note is on the restriction $f : \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{>0} \times \{b_1(\cdot), \dots, b_G(\cdot)\} \to \mathbb{R}_{\geq 0}$: that is, there must be a positive input of the highest labor type. This assumption simplifies proofs and ensures well-behaved derivatives, because the feasibility requirement of blueprints requires a positive quantity of the highest skilled labor type.

That assumption is not restrictive for the applications in this paper. That's because with isoelastic demand curves for very skilled workers, they become arbitrarily cheap when their quantity is close to zero.

In a more general formulation, blueprints might require at least one worker of a minimum worker type $\underline{h}$ — if none is available, lower types have zero marginal productivity. This property might be useful for models of endogenous growth and innovation.

B.2 Firm sizes and non-wage amenities

The basic framework shows that firms producing the same good are identical in all aspects, including firm size. In addition, the model imposes strong links between firm size differences and wage premiums. In this Appendix, I show that those restrictions can be relaxed by allowing for dispersion in firm-specific non-wage amenities—without invalidating any of the theoretical results of the paper.

The fundamentals of the model need to be modified as follows. When the entrepreneur creates a firm, it gets a random draw of amenities $a_j > 0$ from a good-specific distribution that has mean $\bar{a}_g$. Normalize $a_j = 1$ for home production. Worker preferences are now given by:

$$U_i(c, j) = c \cdot a_j^{\frac{1}{\beta}} \cdot \left[\exp\left(\eta_{ij}\right)\right]^{\frac{1}{\lambda}}$$

The idiosyncratic vector η_{ij} is randomly drawn from the same distribution as before. The probability of a worker (h, ε) choosing a particular option j is given by:

$$\Pr\left(0 = \underset{j' \in \{0,1,\dots,J\}}{\arg\max}\ V_{ih}(\varepsilon, j')\right) = \frac{(\varepsilon z_{0,h})^{\lambda}}{(\varepsilon z_{0,h})^{\lambda} + \omega_{\varepsilon,h}^{\lambda}}$$

$$\Pr\left(j = \underset{j' \in \{0,1,\dots,J\}}{\arg\max}\ V_{ih}(\varepsilon, j')\right) = \frac{\omega_{\varepsilon,h}^{\lambda}}{(\varepsilon z_{0,h})^{\lambda} + \omega_{\varepsilon,h}^{\lambda}} a_j \left(\frac{\mathbf{1}\{\varepsilon \geq \underline{\varepsilon}_{hj}\} y_{hj}(\varepsilon)}{\omega_{\varepsilon,h}}\right)^{\beta} \qquad \text{for } j \geq 1$$

$$\text{where } \omega_{\varepsilon,h} = \left(\sum_{j=1}^{J} \mathbf{1}\{\varepsilon \geq \underline{\varepsilon}_{hj}\} a_j y_{hj}(\varepsilon)^{\beta}\right)^{\frac{1}{\beta}}$$

This expression makes is clear that a_j terms becomes a proportional shifter in the firm-level labor supply curve. Given the same posted wage, a firm with a_j twice as large as another will attract twice as many workers, and thus use twice as many efficiency units of labor in production. Lemma 3 can then be extended:

Complement to Lemma 3. *Among firms producing the same good, differences in output and employment are proportional to differences in amenities a_j.*

Finally, Proposition 3 can be rewritten in the following way:

Proposition 3a.

1. *If* $b_g(x) = b(x)/z_g$ *for scalars* $z_1, \ldots, z_G$ *and the ratio* $F_g/\bar{a}_g$ *is the same for all firm-produced goods, then there are no firm-level wage premiums:*

$$\log y_{hg}(\varepsilon) = \max\left\{\upsilon_h + \log\varepsilon, \log \underline{y}\right\}$$

where $\upsilon_1, \ldots, \upsilon_H$ *are scalar functions of parameters.*

2. *If there is no minimum wage and* $b_g(x) = b(x)/z_g$*, wages are log additive:*

$$\log y_{hg}(\varepsilon) = \upsilon_h + \log\varepsilon + \frac{1}{1+\beta}\log\left(\frac{F_g}{\bar{a}_g}\right)$$

3. *If there is no minimum wage and there are firm types* g, g' *and worker types* h' h *such that* $\ell_{h'g'}/\ell_{hg'} > \ell_{h'g}/\ell_{hg}$ *(that is, good* g' *is relatively more intensive in* h'*), then:*

$$\frac{y_{h'g'}(\varepsilon)}{y_{hg'}(\varepsilon)} > \frac{y_{h'g}(\varepsilon)}{y_{hg}(\varepsilon)}$$

What makes a firm "high-wage" in this generalized model is not simply a high entry cost, but a high entry cost relative to average amenities provided by the firm. That is because the model implies a compensating variation for vertical differences in amenities. If firms producing a given good—say, mineral ores—are on average much worse workplaces, they must pay more to achieve the same firm size on average.

With vertical differences in amenities, the model can rationalize any distribution of firm sizes in the economy. Conversely, if firm sizes are not of primary concern, then the model can be simplified by omitting amenities. This is the approach I use in the main paper.

B.3 Tinbergen's race

The following proposition considers a case in which the supply of skill, demand for task complexity, and minimum wages rise in tandem:

Proposition 7 (Race between technology, education, and minimum wages)**.** *Start with a baseline economy characterized by parameters* $\left(\left\{e_h, N_h, z_{0,h}\right\}_{h=1}^{H}, \left\{b_g, F_g, \bar{a}_g\right\}_{g=1}^{G}, z, T, \beta, \lambda, \sigma, \underline{y}\right)$, *where* T *is the stock of entry input (which is normalized to one in the main text). Consider a new set of parameters denoted with prime symbols. Assume* e_h *are decreasing functions to simplify interpretation (more complex tasks are harder to produce). Let* Δ_0, Δ_1 *and* Δ_2

denote arbitrary positive numbers and consider the following conditions:

1. $N'_h = \Delta_0 N_h \ \forall h$ *and* $T' = \Delta_0 T$*: The relative supply of factors remains constant.*
2. $e'_h(x) = e_h\left(\frac{x}{1+\Delta_1}\right) \ \forall h$*: Workers become better at all tasks and the degree of comparative advantage becomes smaller for the current set of tasks (e.g. both high school graduates and college graduates improve at using text editing software, but the improvement is larger for high school graduates).*
3. $b'_g(x) = \frac{1}{1+\Delta_1} b_g\left(\frac{x}{1+\Delta_1}\right) \ \forall g$*: Production requires tasks of increased complexity.*
4. $z' = (1+\Delta_2)z$*,* $z'_{0,h} = (1+\Delta_2)z_{0,h} \ \forall h$*, and* $\underline{y}' = (1+\Delta_2)\underline{y}$*: productivity and minimum wage rise in the same proportion.*

If these conditions are satisfied, the equilibrium under the new parameter set is identical to the initial equilibrium, except that prices for goods are uniformly lower: $p'_g = p_g/(1+\Delta_2)$ *and* $P' = P/(1+\Delta_2)$*.*[30]

Proof. The proof is simple once one notes that the difference between the two economies is a linear change of variables in the task space $x' = (1+\Delta_1)x$, coupled with a reduction in task demand by a factor of $(1+\Delta_2)$. Let $\bar{x}^g_h$ denote task thresholds for firm g in the original equilibrium. Thresholds $(1+\Delta_1)\bar{x}^g_h$ lead to exactly the same unit labor demands, except for a proportional reduction:

$$\int_{(1+\Delta_1)\bar{x}^g_{h-1}}^{(1+\Delta_1)\bar{x}^g_h} \frac{b'_g(x')}{e'_h(x')} dx' = \int_{(1+\Delta_1)\bar{x}^g_{h-1}}^{(1+\Delta_1)\bar{x}^g_h} \frac{1}{(1+\Delta_1)(1+\Delta_2)} \frac{b_g(x'/(1+\Delta_1))}{e_h(x'/(1+\Delta_1))} dx' = \frac{1}{1+\Delta_2} \int_{\bar{x}^g_{h-1}}^{\bar{x}^g_h} \frac{b_g(x)}{e_h(x)} dx$$

So if firms use exactly the same labor inputs, they will produce $(1+\Delta_2)$ times more goods. But because $p'_g = p_g/(1+\Delta_2)$, total and marginal revenues are the same. Since all other equilibrium variables are the same, all equilibrium conditions are still satisfied. □

Proposition 7 delineates balanced technological progress in this economy. Production becomes more efficient by using tasks that are more complex. At the same time, the skill of workers increases, changing the set of tasks where skill differences are relevant. If minimum wages remain as important, then there is a uniform increase in living standards. Wage differences between worker groups and across firms for workers in the same group remain stable.

[30] Using the exponential-gamma parametrization, changes in comparative advantage functions and blueprints are equivalent to $\alpha'_h = \alpha_h/(1+\Delta_1)$, $\theta'_g = (1+\Delta_1)\theta_g$, $k'_g = k_g$, and $z'_g = (1+\Delta_2)z_g$.

B.4 Discussion of missing minimum wage channels

In this appendix, I briefly discuss three minimum wage channels that are not present in this paper. The first is interactions of minimum wage with labor market concentration. By using a "monopsonistic competition" assumption and assuming that the β parameter is common across regions and skill levels, my model rules out the possibility that labor market power varies significantly across regions, as suggested by the empirical work of Azar et al. (2019). My assumptions also rule out the possibility that, by reallocating labor from smaller to larger firms, the minimum wage increases the labor market power of the latter—a channel that is present in the theoretical model of Berger, Herkenhoff and Mongey (2022*b*).

The reason why my framework abstracts from these channels is simplicity. Adding concentration requires not only a more complicated model but also significant effort in precisely defining specific labor markets (such that concentration measures are meaningful). I believe that abstracting from those dimensions does not have first-order implications for my analysis for two reasons. First, low-wage workers in Brazil typically have low levels of schooling. Those workers may not have very specialized skills, and so their potential labor markets may be large and thus less likely to be concentrated. Second, despite not including that feature, the estimated model has a very good cross-sectional fit with respect to formal employment rates for unskilled workers and the size of the minimum wage spike. So, to the extent that regional differences in market power may exist, they may be relatively small.

The second channel that is not explicitly included is capital-labor substitution. The task-based production function could directly account for different forms of capital replacing workers at particular tasks, in the style of Acemoglu and Autor (2011). The reason why this omission is arguably not very consequential is because the firm creation side of the model may account for it. Specifically, the entry input entrepreneurs use to create firms may be interpret as including capital investment. And the association of larger entry costs with a blueprint that is more intensive in complex tasks is a representation of capital-skill complementarity.

One may be concerned that entry inputs are not a good representation of capital because they are a one-time investment. A firm may respond to the minimum wage by scaling up with no need to purchase more capital. The reason why this is probably not a significant constraint is that I only use the model for long-run analyses, and what is most relevant for the calculation of the target moments is the share of workers of each type employed by all firms producing the same good.

The final channel not included in the paper are endogenous increases in worker efficiency in response to the minimum wage. Such "efficiency wage" effects may arise either because of reciprocity/fairness concerns, or because workers would choose to put in more effort at some utility cost to avoid being disemployed following a minimum wage hike. The second effect is the most important for the analysis of employment and wage effects.

The omission of these worker effort effects is not likely to be consequential because, to the extent that workers do that, it should be reflected in a larger minimum wage spike. That is because workers would put the necessary effort to be above the recruitment bar, but they do not need to put in so much effort that it overcomes the wage mark-down. The model matches the data well with a fairly small mark-down—if anything, the spike is over-predicted, not under-predicted. If we estimated an augmented model where a quantitatively important number of workers bunch at the minimum wage due to endogenous effort, than we would need mark-downs to be even smaller to match the size observed spikes. The augmented model would have an additional force against disemployment. But it would also have smaller mark-downs, which lead to stronger disemployment effects. After accounting for both of those changes, comparative statics regarding wages and employment would likely be similar.

C Numerical implementation

C.1 Task-based production function

The basic logic of obtaining compensated labor demands in this model is to use the non-arbitrage equation 2 from Lemma 1 to obtain thresholds as functions of marginal productivity gaps. Then, compensated labor demands can be obtained through numerical integration of Equation 3.

The exponential-Gamma parametrization is helpful because it provides a simple closed form solution for thresholds and the labor demand integrals. Consider the slightly more general version of the parameterization shown in the main text (allowing for heterogeneous k_g by good and productivity shifters z_g):

$$e_h(x) = \exp(\alpha_h x) \qquad \alpha_1 < \alpha_2 < \cdots < \alpha_{H-1} < \alpha_H$$

$$b_g(x) = \frac{x^{k_g-1}}{z^g \Gamma(k_g) \theta_g^{k_g}} \exp\left(-\frac{x}{\theta_g}\right) \qquad (z_g, \theta_g, k_g) \in \mathbb{R}^3_{>0}$$

Then, the compensated labor demand integral can be written as a function of thresholds in two ways: either in terms of incomplete gamma functions or as a power series.

$$\bar{x}_h\left(\frac{f_{h+1}}{f_h}\right) = \frac{\log f_{h+1}/f_h}{\alpha_{h+1}-\alpha_h} \tag{12}$$

$$\begin{aligned}
\ell_{hg}\left(\bar{x}_{h-1},\bar{x}_h\right) &= \int_{\bar{x}_{h-1}}^{\bar{x}_h} \frac{b_g(x)}{e_h(x)}dx \\
&= \begin{cases} \frac{1}{z_g\Gamma(k_g)}\left(\frac{1}{\Upsilon_{hg}\theta_g}\right)^{k_g}\left[\gamma\left(\Upsilon_{hg}\bar{x}_h,k_g\right)-\gamma\left(\Upsilon_{hg}\bar{x}_{h-1},k_g\right)\right] & \text{if } \Upsilon_{hg}\neq 0 \\ \frac{1}{z_g k_g\Gamma(k_g)}\left[(\bar{x}_h/\theta_g)^{k_g}-(\bar{x}_{h-1}/\theta_g)^{k_g}\right] & \text{otherwise} \end{cases} \qquad (13) \\
&= \begin{cases} \sum_{m=0}^{\infty}\frac{\bar{x}_h^{k_g}\exp\left(-\Upsilon_{hg}\bar{x}_h\right)\left(\Upsilon_{hg}\bar{x}_h\right)^m-\bar{x}_{h-1}^{k_g}\exp\left(-\Upsilon_{hg}\bar{x}_{h-1}\right)\left(\Upsilon_{hg}\bar{x}_{h-1}\right)^m}{z_g\theta_g^{k_g}\Gamma(k_g+m+1)} & \text{if } \Upsilon_{hg}\neq 0 \\ \frac{1}{z_g k_g\Gamma(k_g)}\left[(\bar{x}_h/\theta_g)^{k_g}-(\bar{x}_{h-1}/\theta_g)^{k_g}\right] & \text{otherwise} \end{cases} \qquad (14)
\end{aligned}$$

where $\Upsilon_{hg}=\alpha_h+\frac{1}{\theta_g}$, $\gamma(\cdot,\cdot)$ is the lower incomplete Gamma function, and $\Gamma(\cdot)$ is the Gamma function.

Expression 13 is simple to code and fast to run in software packages such as Matlab, where optimized implementations of the incomplete Gamma function are available.[31] When $\Upsilon_{hg}<0$, that expression requires calculating complex numbers as intermediate steps. This is not a problem in Matlab.

If using complex numbers is not convenient or reduces computational efficiency, then the power series representation in 14 should be used. In my Julia implementation, I only use real (floating point) numbers. I use formulation 13 when $\Upsilon_{hg}\geq 0$, and 14 when $\Upsilon_{hg}<0$. Another option, not used in this paper, is to change the normalization of α_h such that they are all non-negative.

Calculating the production function and its derivatives — that is, solving for output and marginal productivities given labor inputs — is not needed in the equilibrium computation nor in estimation. However, it might be useful for other purposes. Those numbers are obtained from a system of H equations implied by requiring that labor demand equals

[31] Note that Matlab's *gammainc* yields a normalized incomplete Gamma function, so dividing by $\Gamma(k_g)$ is not necessary.

labor available to the firm. The choice variables can be either $(q, \bar{x}_1, \dots, \bar{x}_{H-1})$ or $f_1, \dots, f_H$. Moving from thresholds and output to marginal productivities, or vice-versa, is a matter of applying the constant returns relation $\sum_h f_h = q$.

C.2 Equilibrium

Solving for equilibrium can seem challenging at first glance. Using a convenient set of choice variables reduces the problem to solving a square system of $(H+1) \times G$ equations. First, I use the "price" of the entry input (that is, the Lagrange multiplier for the entrepreneur) instead of the price of the final good as the numeraire. Then, I use the following procedure to map guesses of firm-specific task thresholds, firm-level output, and prices for each good into a vector of $(H+1) \times G$ "residuals" which must be zero in an equilibrium:

1. Start with values for mean output $\bar{q}_g$ and task thresholds $\bar{\boldsymbol{x}}_g = \{\bar{x}_{1g}, \dots, \bar{x}_{Hg}\}$ for the representative firms of each type, along with prices for goods p_g.

2. Use the compensated labor demand integral for the task-based production function to find average labor demands $\bar{l}_{hg}$ (Equation 3 in the text, or Equation 13 in Appendix C if using the exponential-Gamma parametrization).

3. Find marginal products of labor f_{hg} via the non-arbitrage conditions (2) and the constant returns to scale relationship $\sum_h f_{hg} \bar{l}_{hg} = \bar{q}_g$.

4. Employ the first order conditions of the firm (7) and (8) to find wages w_{hg} and rejection cutoffs $\underline{\varepsilon}_{hg}$, respectively.

5. Calculate relative consumption $Q_g/Q_1 = (p_g/p_1)^{-\sigma}$ and relative firm entry $J_g/J_1 = (Q_g/Q_1)/(\bar{q}_g/\bar{q}_1)$.

6. Pin down entry of firm type 1 (and thus all others) with entrepreneurial talent clearing: $J_1 = T/(\sum_g F_g J_g/J_1)$.

7. Calculate the real minimum wage as the sum of the minimum wage parameter and the price index implied by the guess of prices for goods.

8. For each $h \in \{1, \dots, H\}$, integrate over ε to find labor supply and labor costs for each firm:

    (a) Choose minimum and maximum values $\varepsilon_{h,lowest}$ and $\varepsilon_{h,highest}$ for numerical integration, based on quantiles of the r_h distribution. In my application I use 0.001

and 0.999 as quantiles.

(b) Split the space $[\varepsilon_{h,lowest}, \varepsilon_{h,highest}]$ into (at most) $2G+1$ segments, based on two thresholds for each g: one based on the minimum employment requirement, and another based on the point where the minimum wage ceases to bind.

(c) For each of those segments:

i. Create an array of discrete values of ε, uniformly spaced between the endpoints of the segment (inclusive).

ii. For each point, calculate $\omega_{h,\varepsilon}$, then the shares of workers choosing each individual firm, the corresponding units of labor going to each firm, and labor cost. Each point should have "mass" corresponding to the density at the point, times the distance between halfway to the previous point until halfway to the next point. For the boundaries, the distance is from the point to the next or previous halfway point.

9. Calculate the error in the system of equations, which has two components:

(a) For each h, g, the deviation between labor demand $\bar{l}_{hg}$ found in Step 2 and the labor supply from Step 8. I normalize those residuals such that they are measured in terms of shares of the total workforce.

(b) The relative deviation between profits and the entry cost parameter F_g (given that the "price" of the entry input is normalized to one).

I make two important notes about the trapezoidal integration in Step 8. One could be tempted to just use a constant grid of ε values. But that significantly reduces the accuracy of numerical differentiation of the system of equations. That is: we want the errors calculated through that procedure to change continuously with respect to the initial guesses. Using the endogenous grid based on the precisely calculated thresholds in ε space is crucial for that.

Second, the procedure could be more simply described as trapezoidal integration, without having to think about the "mass" of each individual discrete point of ε. But the analogy of each point having a weight makes clear that the trapezoidal integration is, effectively, creating a discretized "data set" that can be used to simulate moments from the model. Thus, the same procedure doubles down as a simulation tool, in addition to serving to find equilibrium. See the next subsection for details.

That system of equations can be solved using standard numerical procedures, with the restrictions that $\bar{q}_g > 0$, $p_g > 0$, and $0 \leq \bar{x}_{1g} \leq \bar{x}_{2g} \leq \cdots \leq \bar{x}_{Hg}$ $\forall g$. These restrictions can be imposed through transformations of the choice variables: log prices, log quantities, log of the lowest task thresholds $\bar{x}_{1g}$, and log of differences between consecutive thresholds $\bar{x}_{hg} - \bar{x}_{h-1,g}$ for $h = 2, \ldots, H-1$.

The procedure may be sensitive to starting points for some parameters. I solve this issue in two ways. First, I create a separate routine to provide a reasonable guess for the starting point. In essence, the procedure makes sure that initial task thresholds are such that, for all g, employment shares of each type is at least $0.1/H$. This is to make sure that derivatives regarding task thresholds are not zero in the starting point. For the prices and quantities, I just try a small grid and choose the combination with the lowest maximum for the loss vector.

The second way to address the issue is to try a potentially large number of starting points, and also different optimization algorithms. My code tries a maximum of 50 attempts. If a point is found that has maximum residual of 10^{-10} or less, the equilibrium-finding procedure stops. If no solution that precise is found, it takes the one with the smallest maximum residual among all 50 attempts. If the maximum residual is 10^{-4} or less, it is considered a success. Otherwise, the procedure fails.

C.3 Simulating measures of wage inequality

As explained in the previous section, the procedure used to calculate the equilibrium "errors" doubles down as a simulation tool. I include an option in that function to save a data set with all discrete combinations of (h, ε, g) with the corresponding weights (i.e., shares of workforce) and log earnings.

In the quantitative exercise, I need to calculate some moments at the educational level. It is straightforward to create a version of the same data set with a variable for observable educational group. To do so, one needs to "expand" the data so that each observation in the old data corresponds to three observations in the new. The weight of the old observation is split among the new three based on the probabilities $P(\hat{h}|h)$. From the new data set, is is straightforward to calculate metrics such as between-group wage gaps and within-group variances.

The only moments that require more thinking are the variance decomposition components.

To reason about AKM decompositions in the theory, I need a two-period version of the model, from which panel data could be simulated if needed. I assume that, with some probability $R > 0$, workers re-draw their full vector of idiosyncratic preferences $\boldsymbol{\eta}_i$ from period one to period two. I also assume that only part of the efficiency units of labor of a worker is transferable: $\log \varepsilon_{t=2} = A \log \varepsilon_{t=1} + (1 - A^2)^{0.5} \log \varepsilon'$, where ε' is a new i.i.d. draw from the same distribution of efficiency units (given h). After the re-draws, the labor market clears in the same way as in period 1.

Because the cross-sectional distribution of $(h, \varepsilon, \boldsymbol{\eta})$ remains the same as before, firm choices and the equilibrium allocation remain the same, except for the identity of workers employed by each firm. That model of job-to-job transitions implies that, whenever a given worker type (h, ε) is employed in equilibrium by the two firm types, there is a positive probability that some of those workers moved from a firm of type $g = 1$ to another of type $g = 2$ (and vice-versa).

Furthermore, I assume that firms are large, in the sense that there are many movers and firm fixed effects in the AKM regression are precisely estimated. Together with Lemma 3, that assumption implies that all firms producing the same good will have the same estimated fixed effect.

Given these assumptions, the results of an AKM decomposition of log wages using simulated panel data are identical to running a two-way fixed effects model based on simulated data from one period, using a "worker id" indicator for each combination of (h, ε) and a "firm id" indicator for each good. Each observation is a (h, ε, g) cell. The regression is weighted by the share of the employed population in the corresponding cell. Finally, the estimated worker fixed effects are shrinked by the factor A, since they correspond only to the portable portion of productivity. The persistence parameter A is calibrated such that the R^2 of the simulated AKM regression is 0.9, about the same as the empirical regressions.[32]

This approach ignores granularity issues in the simulation of AKM moments. That is conceptually consistent with the way the corresponding moments are estimated from the data, since the KSS estimator is not subject to limited mobility bias.

[32] The persistence parameter is allowed to change between 1998 and 2012 and between regions.

Table D1: Sample sizes for the 151 selected microregions

	1998			2012		
	Min.	Mean	Max.	Min.	Mean	Max.
Panel A: Base year						
Adult population (thousands)	69	396	7,037	82	512	8,240
Formal workers in RAIS (thousands)	16	121	3,117	26	216	4,954
Establishments in RAIS	743	9,216	190,784	2,352	15,887	288,929
Panel B: Three year panel around base year						
Unique workers in connected set (thousands)	7	93	2,500	18	178	4,181
Unique establishments in connected set	132	2,527	62,416	598	6,637	135,819

Notes: Panel A shows sample sizes for each microregion in 1998 and 2012. Adult population is the count of all individuals between 18 and 54 (inclusive), using Census data. RAIS is the matched employer-employee data set. Panel B shows the numbers of workers and establishments used in the estimation of two-way fixed effects models, using data from 1997 through 1999 ("1998") and 2011 through 2013 ("2012").

D Appendix to the quantitative exercises

D.1 Sample sizes

Sample sizes for the descriptive statistics and quantitative exercises are displayed in Table D1.

D.2 Variance decomposition using Kline, Saggio and Sølvsten (2018)

The estimation of variance components follows the methodology proposed in Kline, Saggio and Sølvsten (2018), henceforth KSS. For each period (1998 and 2012), I use a three-year panel centered around the base year. The sample used for estimation is the largest leave-one-out connected set. This concept differs from the usual connected set in matched employer-employee datasets because it requires that firms need to be connected by at least two movers, such that removing any worker from the sample does not disconnect this set. Table D1 presents the size of that largest connected set in each period.

I implement the variance decomposition using the Julia code provided by KSS.[33] There are some implementation choices required in this estimation, stated below:

- Dealing with controls (year fixed effects): "Partialled out" prior to estimation.
- Maximum number of interactions: 300

[33] Currently available at `https://github.com/HighDimensionalEconLab/VarianceComponentsHDFE.jl`.

- Sample selection: includes both movers and stayers. The leave-out procedure leaves a whole match out, not simply a worker-time observation.
- Number of simulations for JLA algorithm: 200

D.3 Validation of the task-based production function: robustness

Table D2 shows additional versions of the validation exercises from Table 3. Panel A repeats the results from that table for quick referencing. Panels B and C show sample restrictions where regions where the minimum wage binds more strongly are eliminated. That exercise tests whether the log-wage complementarities shown in Column (5) are mechanical consequences of minimum wages. That could be a concern since minimum wages censor the bottom of the wage distribution, and thus reduce the possibility of cross-firm wage differentials for unskilled workers.

The coefficient of interest falls by 28% from Panel A to Panel B, but remains statistically significant. The further sample restriction from Panel B to Panel C has essentially no effect on the estimated coefficient, which remains statistically distinguishable from zero. Thus, I conclude that minimum wages are not the primary cause for the log wage complementarities.

In Panel D, I explore an alternative measure of skill, constructed in the following way. First, I split workers into 12 age groups (each group includes three years of age, except the last, which includes workers 51 through 54). Next, I use data from 1997 only to run a regression of log wages on schooling fixed effects, age fixed effects, and firm fixed effects. Thus, it accounts for nonlinearities in returns to schooling, the role of age, and nets out some of the effects of firms on log wages. The measure is normalized to range from zero to 15, so that the magnitude of the coefficient can be more easily comparable to the ones from the other panels. The firm-level averages and leave-out averages are recalculated using the Mincerian measure.

I find that the results are very similar for all outcomes. In unreported results, I also find that results hold if the skill measure is just dummies for the three educational groups, as used in the remainder of the quantitative exercises. I conclude that the results are not sensitive to the particular metric of worker skill I use.

Table D2: Validation of the task-based production function: robustness.

	Non-routine cognitive task content				Log wage
	(1)	(2)	(3)	(4)	(5)
Panel A: baseline estimates					
Coefficient	0.07921	0.06304	0.00663	0.00343	0.00162
Standard error	(0.00049)	(0.00159)	(0.00077)	(0.00086)	(0.00045)
r2	0.26216	0.40172	0.84463	0.85033	0.95789
N	93,606	11,551,108	2,673,660	2,673,659	14,996,848
Panel B: 101 microregions where spike ≤5% of formal emp.					
Coefficient	0.08138	0.06166	0.00827	0.00531	0.00117
Standard error	(0.00053)	(0.00175)	(0.00073)	(0.00084)	(0.00039)
r2	0.26849	0.40415	0.84489	0.85056	0.9572
N	82,711	10,333,034	2,415,618	2,415,617	13,142,099
Panel C: 44 microregions where spike ≤2% of formal emp.					
Coefficient	0.08331	0.06116	0.00941	0.00678	0.00113
Standard error	(0.00061)	(0.00214)	(0.00085)	(0.00098)	(0.00048)
r2	0.2762	0.40159	0.84052	0.84619	0.95668
N	60,230	7,567,905	1,774,798	1,774,796	9,510,389
Panel D: Mincerian measure of skill					
Coefficient	0.07373	0.05314	0.00519	0.00297	0.00159
Standard error	(0.00043)	(0.00182)	(0.00074)	(0.00086)	(0.00042)
r2	0.27312	0.40156	0.84461	0.85033	0.95789
N	93,606	11,551,108	2,673,660	2,673,659	14,996,848

Notes: See notes from Table 3.

D.4 Discussion about identification

D.4.1 Avoiding incidental parameter bias

A central challenge in the empirical model is allowing for region-specific heterogeneity in labor demand parameters, formal employment shifters, and overall productivity levels (which are strong determinants of how binding the minimum wage is in each region). It would not be realistic, for example, to assume that regional labor demand is orthogonal to education, or that education is orthogonal to productivity. Thus, when specifying the unobserved supply, demand, and productivity parameters, the structural model needs to account for the possibility of such correlations.

One approach would be to add flexible fixed effects to model to capture such unobserved heterogeneity. But that solution would be incomplete, since there may be heterogeneous trends in addition to heterogeneous levels. For example, rural regions could on average be less educated initially, face stronger educational growth, and receive stronger shocks to TFP and relative demand for unskilled labor due to the commodities boom.

A worse problem with the fixed effects approach would be incidental parameter bias, since the model is nonlinear. There exist methods to deal with incidental parameter bias in such panel models (e.g., Hahn and Kuersteiner, 2002; Hahn and Newey, 2004). However, they rely on large T asymptotics. Since I am estimating a long-run model, those methods are not appropriate.

This is the motivation for specifying the regression-style models for the biased demand parameters, and using a subset of the endogenous outcomes to invert the flexible region-specific parameters. Three region-specific outside option parameters are recovered from formal employment rates in 1998, capturing heterogeneity in outside options at the microregion-education group level. The formal employment rate for high school workers in 2012 recovers the common region-specific shock to outside options for all groups. That could reflect, for instance, location-specific changes in the enforcement of labor regulations, which affects informality rates (Almeida and Carneiro, 2012).[34] Local TFP in each period is inferred from the minimum wage bindingness level. In effect, those endogenous outcomes are used as covariates, somewhat analogously to how empirical strategies such as Lee (1999) use measures of minimum wage bindingness as independent variables in regressions. An important difference is that the inversion procedure explicitly takes into account that observed bindingness depends on several other characteristics at the local level in addition to TFP, such as the educational distribution and labor demand characteristics.

Inversion requires that there should be no error in formal employment rates for 1998, the employment rate of high school workers in 2012, and the minimum wage bindingness variable (Assumption 3). That is because the model is nonlinear: even if there is mean-zero error, it could still introduce bias to the model, which would not go away with an increase in the number of regions.

As mentioned in the main text, the residuals $\boldsymbol{u}_r$ include misspecification in functional forms, omitted variables, and sampling error. Functional form issues are not an issue, since the model can always match observed formal employment rates and levels of minimum wage bindingness by shifting the flexible productivity and outside option parameters. As for omit-

[34] I choose high school workers as the reference group because it corresponds to a large share of the workforce in both periods, thus providing more precise estimates of the formal employment rate.

ted variables, Assumption 3 can be viewed as a normalization: the "z" parameters to be inverted should be interpreted as encompassing all factors that drive formal employment and bindingness other than the wage index.

Sampling error could be an issue, but it is made less relevant by the sample restrictions I use. The most imprecise measure is the formal employment rate of college workers in 1998, as they are by far the smallest worker group and the sample is smaller (and less educated) in 1998. But since the sample is selected to have regions with at least 1,000 formal workers with college education (and thus more than 1,000 adults with college education), the sampling error is minimal. The largest estimated standard error is 0.013, for a point estimate of 0.654. That region has a small population, such that its weight in estimation is not large. The mean standard error, using the region-specific estimation weights, is 0.005. That is, standard errors are about 1% of the point estimates, and 2% in the region with the most imprecise estimate. Thus, they are unlikely to cause significant bias.

D.4.2 Identifying variation and instrumental variables analogy

The estimator can be interpreted as a nonlinear instrumental variables model. The population share instruments have a primary effect ("first stage") on the endogenous total supply of skilled labor to the formal sector. Time is used as an instrument for common changes in the three time-varying demand-side parameters: blueprint complexity of advanced firms, entry cost ratios between firms, and relative taste for advanced goods. That is: conditional on observed changes in minimum wage bindingness and labor supply, the only time-varying factors are the three demand shocks. That approach is analogous to that of papers such as Katz and Murphy (1992), where a time trend is interpreted a change in unobserved shocks conditional on labor supply.

The interaction of time with initial sectoral shares in agriculture and manufacturing is inspired by papers that use shift-share instruments to gauge the effects of trade shocks between regions. That is clear by noting that the equations for the three time-varying demand parameters can be written as time changes within microregion, and each of the initial sectoral shares can have an independent effect on those changes that is different from their impacts on initial levels.

The simultaneous equation least squares estimator can then be interpreted as stacking the first stages and reduced forms, which is one way to estimate an IV model (in the classic IV model, one would estimate them as a set of seemingly unrelated regressions). A potential concern is that the residuals of first stages will be correlated with those of the reduced forms. This is an important reason why the model needs to allow for within-region correlated errors, even between different time periods. It is not the only reason, though. As another example, an unobserved factor that affects the wage for high school workers would mechanically affect the two between-group wage gaps.

I also rely on some exogenous variation in the bindingness level of the minimum wage. It comes from the assumption that region-time-specific TFP is mean independent of the residuals conditional on all instruments and outside option parameters. The estimator uses that variation to infer how minimum wage bindingness maps into the size of the spike and the share of the employed workforce close to the minimum wage. That information, in turn, identifies the firm-level labor supply elasticity β and the skewness parameter of the distribution of efficiency units, χ.

One advantage of my approach is that it "corrects" for differences in the shape of the wage distribution that could be driven by different supply and demand characteristics across regions. Those might be confounders both because they may correlate with TFP and because they have independent effects on wages, and thus affect empirical measures of bindingness such as the size of the minimum wage spike or how the minimum wage compares to the mean or median of the log wage distribution. In addition, I do not need to specify a reference point at which the minimum wage is assumed to have no effects, as in Lee (1999) or Autor, Manning and Smith (2016). That is useful for capturing possible general equilibrium effects which could affect the upper tail of the distribution. As a potential downside, I have to specify a fully parametric model, which may not be accurate. When evaluating the fit of the model, I will argue that the model is flexible enough to accurately portray the shape of the wage distribution, particularly at the left tail.

The variation in labor supply, labor demand, and minimum wage bindingness induced by the instruments is then used to identify the remaining general parameters of the model:

Worker types: The comparative advantage of high school workers $\mu_{\hat{h}=2}$ is identified from the initial mean log wage gap between high school workers and those with less than high school. To identify the dispersion in comparative and absolute advantage within educational groups, I need to combine two kinds of information for each of them. The first is the overall level of wage dispersion, measured through the initial variance of log wages within group. The second piece of information is revealed by how the changes in the variance of log wages correlate with changes in skill premiums at the microregion.[35]

Outside options: The four region-specific parameters are inferred from observed formal employment rates, as described above. The two shocks to outside options at the education level (for less than high school and for college workers) are identified by matching the average employment rates for those groups. Finally, the preference parameter λ, which regulates the macro elasticity of labor supply, is identified by the correlation between employment rates and the predicted inclusive value of formal employment, which is a function of wages and the number of firms of each type in the economy.

Blueprint shape and elasticity of substitution between goods: Those two parameters have important implications for sorting and the aggregate substitution patterns between worker types. The first, k, determines the extent to which the skill-intensive firms are specialized. The second, σ, determines how good-specific output, and thus firm entry and aggregate employment by firm type, responds to shocks that affect relative costs, such as changes in skill premiums induced by supply or demand shocks. That has strong implications for how mean log wage gaps between groups respond to those shocks, as well as the contribution of firm premiums to within-group inequality. Thus, the two parameters are jointly recovered from cross-sectional correlations between supply and demand shocks, sorting, skill premiums between groups, and variances of log wages within groups.

D.4.3 Identifying variation in the Brazilian context

The variation used to identify the impact of supply comes from the dramatic rise educational achievement in Brazil. The country has historically low levels of schooling (see Chapter 5 in Engerman and Sokoloff, 2012, for a discussion of the historical development of schooling institutions in the Americas). In 1989, average years of schooling were 5.1 in Brazil, compared to 6.1 in Mexico, 7.11 in Venezuela, or 8.4 in Chile (calculated using statistics compiled in SEDLAC, 2022). But with the return to democracy in 1985, following more than 20 years of military dictatorship, a series of reforms helped set a new trajectory for schooling achievement in the country.

These developments started at the end of the military dictatorship. A constitutional amend-

[35] If there is significant dispersion in comparative advantage in a group, then the variance of log wages within that group should increase with skill-premiums. Alternatively, if all of the productivity dispersion is in absolute advantage, then log wages within a group move in tandem. Because the estimation procedure is joint, that logic is valid after netting out the contribution of other factors such as minimum wages, which may have strong independent effects on within-group variances of log wages.

ment passed in 1983 ("Emenda Calmon") imposed minimum expenditure requirements on education: at least 13% of federal resources and 25% of state and minicipality-level resources. The dictatorship argued that the amendment was not binding without another law regulating it. Congress acted, and the new law was passed in 1985. Later, the new Constitution of 1988 enshrined that law, with the federal expenditure requirement increasing to 18%. The new Constitution also gave municipalities more autonomy in how to organize their educational systems.

More systematic efforts to expand schooling followed in the 1990's and 2000's. In 1996, a new law ("Lei de Diretrizes e Bases da Educação Nacional") established guidelines and attributed formal responsabilities to federal, state, and municipal agents in promoting the universalization of schooling. In 1995, the federal government created an effective system to collect school quality data at the national level ("Saeb"). Another system for evaluating secondary education followed in 1998 ("Enem"). In 2001, the federal government implemented a national cash transfer program conditional on school enrollment ("Bolsa-Escola", later incorporated into the "Bolsa Família" program). And starting in 2005, the "ProUni" program subsidizes low-income students who wished to attend private colleges and universities (public universities are tuition-free in Brazil, but few low-income students are able to pass the entry exams). This list of reforms and policies, which is not exhaustive, shows that that the rise in schooling achievement in Brazil was not an accident, nor should be viewed as "automatic" consequence of economic growth.[36]

The model allows for trends in labor demand that correlate with schooling achievement measured in 1998, as well as with initial employment shares in agriculture and manufacturing and overall wage levels (relative to the minimum wage). Thus, the variation in disentangles the effect of supply from that of demand comes from regions where the growth in schooling achievement was faster or slower than expected, compared to other locations that were similar in 1998. I argue that this variation is plausibly exogenous. Reverse causality is unlikely because it takes years or decades for household or local government decisions to be reflected into shares of the adult population belonging to each educational group.

Why does schooling rise faster in some regions, compared to others? It could be due to differences in policies implemented before 1998, or due to the fact that some national policies could affect regions differently. As an example of the former, the Brazilian Federal

[36]Indeed, economic growth was much more significant in the 1960's and 1970's than the 1980's and early 1990's.

District (where the capital, Brasília, and a few other cities are located) implemented a local cash transfer program in 1995, six years before the national program. As for the latter, the minimum expenditure requirements from "Emenda Calmon" and the 1988 Constitution were more binding in some states than in others, such that some were more strongly affected by that policy.

D.4.4 Threats to identification

At this point, it is worth emphasizing some threats that could hinder identification in other models, but are not problematic for my estimator:

- Labor demand shocks cause endogenous responses in labor market participation, leading to simultaneity bias in supply: not a problem because supply of labor to the formal sector is a modeled endogenous outcome.
- On average, regions that are initially more "backward"—lower education and TFP, for example—experience both more rapid growth in education and more biased labor demand shocks (regional convergence): not a problem because demand shocks may correlate with initial education and sectoral shares.
- Outside options for educated workers might be worse in places with higher demand for skilled labor, or places where the supply of educated workers grows faster, or regions experiencing more technical change: not a problem because region-education-specific outside option parameters are not assumed to be independent of demand, supply, or TFP (though they must be orthogonal to the unmodeled residuals).
- Outside options are becoming worse for low-educated workers relative to college workers, because of unmodeled factors leading to a decline in the number of informal jobs in the economy: not a problem because of the flexible education-time-specific outside option parameters.
- Outside options for all workers are becoming worse in regions that are developing faster, again due to a stronger decline in informal jobs in those regions: not a problem because of the flexible region-time-specific outside option parameters, which need to be orthogonal to the residuals but may be arbitrarily correlated with local supply and demand factors.

Still, there may be threats to identification. One particular concern is an imperfect mapping between education groups and worker productivity in the model. For example, average school quality may be higher in large urban areas, compared to more rural microregions. That would introduce non-random measurement error, a possible source of bias.

I argue that the model is robust to some forms of correlated misspecification of both absolute or comparative advantage, if they affect workers of all educational groups in the same microregion. For absolute advantage, the result follows from noting that the productivity shifters z_{rt} are flexible, and thus would absorb proportional differences in productivity for all workers. For comparative advantage, the model is robust to region and time differences in the α_h parameters that correlate with labor demand shifters, as long as the α_h vary in the same proportion for all h. To see why, look at Proposition 7, shown in Appendix B.3. It shows how a the effects of such proportional shocks to the α_h can be "compensated" by corresponding proportional changes in task complexity θ, leaving the wage distribution unchanged.

One could think of other forms of misspecification that would be more serious. For example, the quality of newly created colleges might be lower than that of preexisting ones, such that in places where college expansion is stronger, the average human capital of college graduates might be lower compared to workers without college. In that case, the estimated effects of increased supply of skill on the labor market may be underestimated (possibly introducing bias in the estimated effects of demand shocks as well). Investigating that potential source of bias is beyond the scope of this paper.

D.5 Estimation

D.5.1 Numerical implementation of the loss function

The estimation procedure is implemented using the Julia programming language (Bezanson et al., 2017). There are two major challenges in the implementation of the loss function. The first is the need to account for the inversion procedure described in the main text. The second is the need to minimize the chance that no equilibrium can be find. The issue is that, with 302 region-time combinations, it is possible that parameter guesses are such that it is hard to find all of the equilibria. This is a problem for estimation, because if even one equilibrium is not found, the loss function cannot be calculated. While one can impose ad hoc shortcuts such as assuming the loss function is large in such cases, those shortcuts can lead the optimization procedure astray, making it fail to converge or converge to points that could be local instead of global minimums.

I start with creating two alternative formulations of the equilibrium-finding procedure that incorporate the inversion procedure. The first one is used for equilibria corresponding to the 1998 time period. In those, I include four choice variables, corresponding to the parameters to be inverted: $\hat{z}^{RH}_{r,1}$, $\hat{z}^{RH}_{r,3}$, $\hat{z}^{RT}_{r,1998}$, and $z_{r,1998}$. Then, I add four "residuals" corresponding to the formal employment rates for the three educational groups and the minimum wage bindingness.

The second version is used for the 2012 period. It only has two additional variables, $\hat{z}^{RT}_{r,2012}$ and $z_{r,2012}$, and two additional residuals, the formal employment rate for high school workers and minimum wage bindingness.

The evaluation of the loss function will then try to solve equilibria for each region separately (using parallel processing if multiple cores are available). First, it will attempt to solve for the 1998 equilibria using the alternative equilibrium-finding procedure above (trying up to 50 starting points, as described in Appendix C). If it fails, it will try to match at least minimum wage bindingness and employment for high school workers (that is, using the procedure for 2012). If even that fails, it will try to solve for an equilibrium with no inversion.

In case an equilibrium without the full inversion is found, the procedure will try to use that as a starting point to achieve complete inversion. Specifically, if only an equilibrium with no inversion at all is found, that equilibrium is used as a starting point to find an equilibrium using the 2012 inversion. Then, if an equilibrium with 2012 inversion is found, then that is used as a starting point for the desired 1998 inversion.

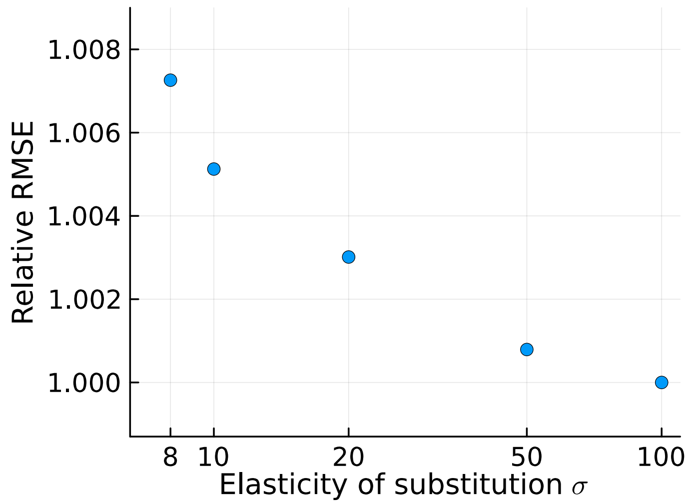
Next, the procedure tries to solve for the actual 2012 equilibrium. There, it will use some of the outside options parameters found for 1998. Again, if the equilibrium with inversion cannot be found, the procedure will attempt to find an equilibrium without inversion. That equilibrium will then be used as a starting point to find the equilibrium with inversion.

The estimator then proceeds to the Jacobian. There, it will use all of the equilibria found in the first evaluation as starting points, leading to large computational gains.

The estimation loss function allows for incomplete inversion. This is addressed by including all endogenous outcomes, including the ones used in the inversion, in the sum of squared deviations to be minimized. The endogenous outcomes that need to be zero by the inversion procedure receive a high equation weight.

That sequence of steps is somewhat complicated, but highly effective. In practice, the procedure will report using equilibria without full inversion only for points very far from the global minimum.

Figure D1: Relative mean squared error with fixed σ



D.5.2 Estimator and starting points

I use the Levenberg-Marquardt optimization algorithm. All parameters are transformed to eliminate the need for constrained optimization. I begin with a set of parameters that produced somewhat realistic moments, with elasticities $\beta = 4$, $\lambda = 0.5$, and $\sigma = 2$. Then, I started the optimization procedure using that starting point and nine others in parallel. The other starting points had random Uniform[-0.5,0.5] shifts (in terms of transformed parameters) compared to the base one.

The best result from this first step was then used in a second draw of starting points. There, the random shifts in transformed were smaller (between -0.1 and 0.1). The best point from that second draw is the optimal point shown in the paper. Most of the other points were very close in terms of estimated parameters and values of the loss function. The complete process took about four weeks using 180 CPU cores in a modern compute cluster.

I also experimented with other heuristics to generate starting points, different optimization algorithms, and weighting schemes. My conclusion is that the procedure is not very sensitive to most implementation choices. However, abandoning equation weights leads to much worse quality of fit for some moments. That is because there is significant differences in the variance of residuals in different equations.

D.5.3 Elasticity σ at the boundary of the parametric space

As explained in the main text, two parameters are found to be at the boundary of the parametric space. One of them implies that, for workers with less than secondary schooling, all of the within-group variation comes from dispersion in efficiency units of labor ε, not labor types h. The second is that goods appear to be perfect substitutes in production. Specifically, the estimation procedure stopped at a point with $\sigma = 100.0074$. At that point, marginal changes in σ had almost no effect on the loss function. Because that parameter is central to comparative statics, I spent some time studying that result.

I started the analysis by checking whether that the large σ was an outlier. I found that, even though the initial points in the first draw had values around 2 for that elasticity, the estimation procedure moved in the direction of a much higher σ for almost all of them.

Next, I ran a series of additional estimation exercises where the σ was constrained to four different values: 8, 10, 20, and 50. For $\sigma = 10$ and above, the starting point for all other parameters was the optimal point. For $\sigma = 8$, I used the optimal point and six additional random points (using uniform shifts between -0.1 and 0.1).

Figure D1 shows the relative root mean squared error for those additional exercises (for $\sigma = 8$, it picks the best result). That figure shows a smoothly declining pattern. The slope is considerably larger for lower values, suggesting that quality of fit starts falling fast when the elasticity becomes small.

As the final step in the analysis, I looked into the quality of fit separately by moment. My goal was to understand what aspect of the data lead the estimator to a large value for σ. I find that the average predicted values for all moments remain the same. However, the R2 for the variance of log wages for college workers goes from 0.05 to -0.05 as σ falls from 100 to 8. That observation is consistent with the discussion in Appendix D.4, where I discuss what kinds of variation help pin down each parameter. I conclude that substantial responses in reallocation are needed to better explain the cross-sectional differences in the variance of log wages for college workers.

D.5.4 Estimates of demand parameters

Table D3 shows estimates of the $\delta_i^{d,t}$ demand-side parameters. The coefficients are reported for demeaned variables within each period, such that the constants capture the year-specific averages of the parameter transformations. Those averages point to an overall demand shock

Table D3: Estimates of demand parameters

	$\log \theta_{2,r,t}$		$\log\left(\frac{F_{2,r,t}}{F_{1,r,t}}\right)$		$\log\left(\frac{\gamma_{2,r,t}}{1-\gamma_{2,r,t}}\right)$	
	1998	2012	1998	2012	1998	2012
Constant	0.77	1.43	9.36	6.69	1.99	1.71
	(0.13)	(0.18)	(0.51)	(0.29)	(0.03)	(0.04)
Initial share high school	0.15	2.26	-0.70	-0.16	0.45	0.10
	(0.40)	(1.88)	(3.87)	(4.08)	(0.64)	(0.67)
Initial share college	3.42	-0.45	2.19	0.48	1.62	-1.80
	(0.97)	(2.00)	(8.43)	(5.38)	(0.27)	(0.79)
Initial share agriculture	0.53	0.12	-1.80	-4.60	-0.14	-0.71
	(0.32)	(0.50)	(1.84)	(1.64)	(0.24)	(0.25)
Initial share manufacturing	-0.34	-1.87	-6.80	-7.51	-1.46	-1.89
	(0.41)	(0.38)	(2.07)	(1.74)	(0.29)	(0.32)
Current log min. wage minus mean log wage	0.43	0.60	-0.21	-1.96	0.35	-0.10
	(0.13)	(0.22)	(0.61)	(1.38)	(0.09)	(0.19)

Notes: Estimates of the $\delta_i^{d,t}$ demand-side parameters. All of the variables are demeaned within time period, and thus the constants measure mean parameter values for each year. Standard errors, shown in parentheses, are cluster-robust at the region level, calculated using the sample analogue of the asymptotic formula from Proposition 6.

that combines three elements. First, task complexity requirements at the skill-intensive firms are increasing. Second, the relative entry cost ratio falls, such that it becomes relatively easier (from the point of view of entry inputs) to create skill-intensive firms. And third, there is a reduction in the relative taste for the skill-intensive good (corresponding to an exogenous average increase in the price for the low-skill good, since $\sigma \to \infty$ in the estimated model).

The interpretation of the other coefficients is not straightforward clear, since they correspond to partial correlations. However, it is worth pointing out that several of them have economically meaningful magnitudes and are statistically significant. That points to the importance of allowing for those correlations in the empirical model.

D.5.5 Benchmark regression models for quality of fit

I use two benchmark models to gauge the quality of fit within sample.

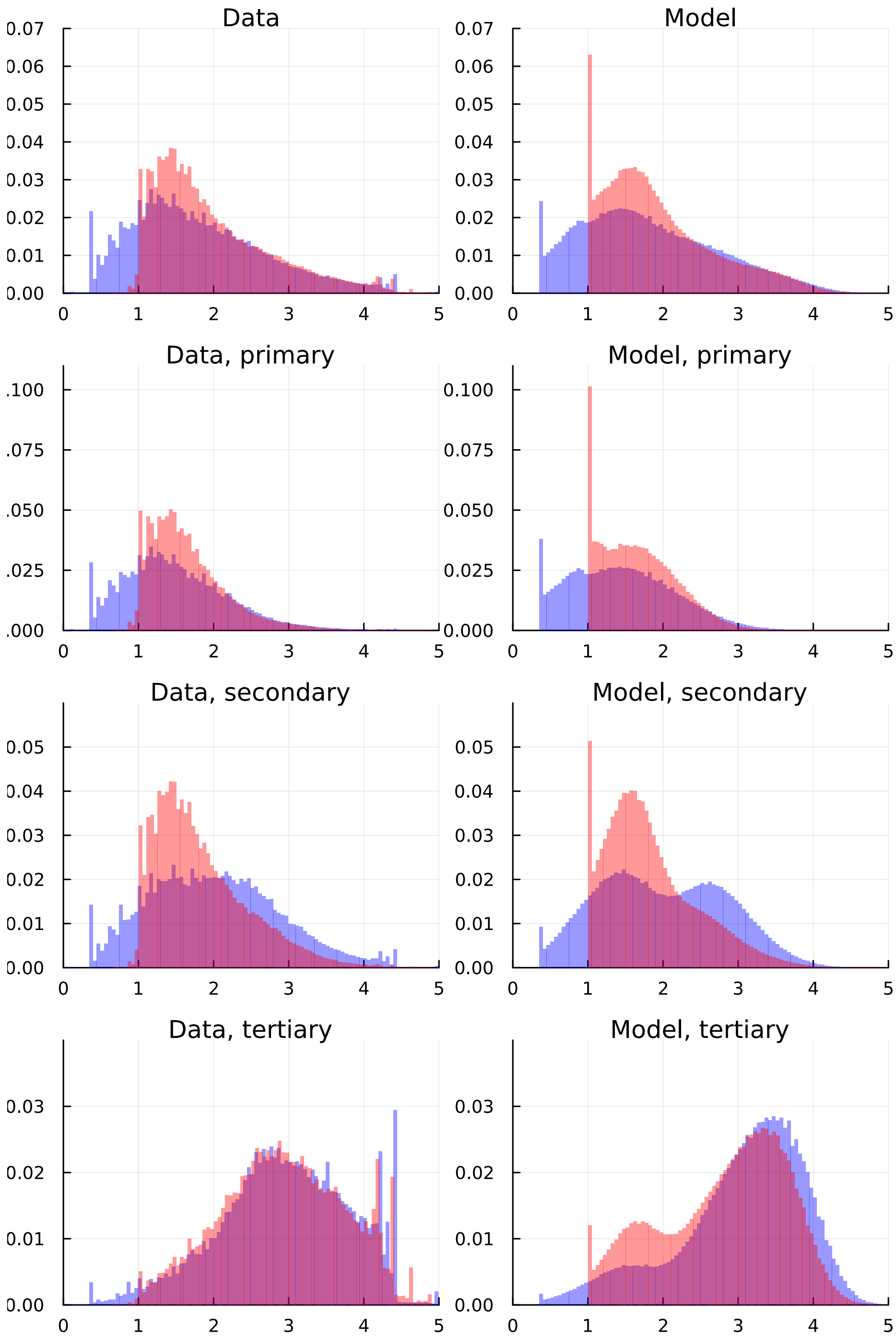
Simple OLS: I run separate regressions for each moment. For all outcomes except the formal employment rates, the regressions include both time periods (302 observations in each). The regressors are time effects, share of adults with high school, share of adults with college, and the difference between the minimum wage and the mean log wage. I run two additional regressions, one for formal employment rates of adults with less than secondary, and the same outcome for adults with college education. Each uses data only for 2012 (151 observations each). The regressors are a constant, the lagged employment rate (i.e., for the same group in 1998), and the current formal employment rate for high school workers. That makes the employment rate regression comparable to the structural model, as it features region-education and region-time effects estimated by matching lagged participation values and the employment rates for high school workers. The model has a total of 51 parameters ($9 \times 5 + 2 \times 3$). This is the exact number of estimated parameters in the structural model.

Large OLS: That model is an augmented version of the Simple OLS with more regressors and allowing for nonlinearities in the effect of the effective minimum wage. For outcomes other than employment rates, the regressors are time effects, current share of adults with high school, initial share of adults with high school (that is, for the same region in 1998), current share of adults with college, initial share of adults with college, initial share of workforce in agriculture, initial share of workforce in manufacturing, effective minimum wage, and effective minimum wage squared. For the formal employment regressions, the regressors are those Simple OLS model along with all others mentioned above. That yields a total of 112 parameters ($9 \times 10 + 2 \times 11$).

D.5.6 Additional measures of fit

In this section, I show additional measures of the quality of fit. I start with a comparison of the national histogram of log wages to that predicted by the model. The top panels in Figure D2 shows that the model closely approximates the real histogram, highlighting the quality of fit in both the inequality and relative formal employment across worker groups and regions. The other panels shows separate histograms for each educational group. Again, the model fits the data very well. The worst fit is for college workers. That is consistent with the lower quality of fit shown in Table 5 for the returns to college and the variance of log wages for college-educated workers. This lower quality of fit comes from the fact that thos moments have more residual variance in the data, and thus receive lower weight in the estimation procedure.

Figure D2: Distribution of log wages, data and model



Notes: This figure shows histograms of log wages using 0.05-sized bins, for the whole adult population and separately by educational group (Less than secondary, Secondary, and Tertiary). The histograms represent real and simulated data for all 151 microregions in the sample.

Table D4: Cross-sectional quality of fit (R2) within time periods

	Model		Simple OLS		Large OLS	
	1998	2012	1998	2012	1998	2012
Moments	(1)	(2)	(3)	(4)	(5)	(6)
Wage inequality measures						
Secondary / less than secondary	0.008	0.236	0.024	0.293	0.184	0.377
Tertiary / secondary	0.142	0.181	0.121	0.289	0.264	0.636
Within less than secondary	0.33	0.126	0.467	0.507	0.659	0.59
Within secondary	0.123	0.625	-0.085	0.341	0.301	0.668
Within tertiary	0.128	-0.292	0.289	0.05	0.371	0.302
Two-way fixed effects decomposition						
Variance establishment effects	0.328	0.329	0.259	0.29	0.328	0.415
Covariance worker, estab. effects	0.21	0.699	0.311	0.536	0.425	0.715
Formal employment rates						
Less than secondary	1.0	0.905	1.0	0.915	1.0	0.959
Secondary	1.0	1.0	1.0	1.0	1.0	1.0
Tertiary	1.0	0.076	1.0	0.471	1.0	0.619
Minimum wage bindingness						
Log min. wage - mean log wage	1.0	1.0	1.0	1.0	1.0	1.0
Share < log min. wage + 0.05	0.802	0.61	0.616	0.519	0.836	0.737
Share < log min. wage + 0.30	0.855	0.856	0.683	0.626	0.854	0.884

Notes: This table displays the within-year quality of fit of the model, as measured by the R2 metric. The R2 can be negative if the model fits the data more poorly than a constant equal to the weighted mean of the target moment. The table also shows the quality of fit of the two benchmark OLS models described in Appendix D.5.5.

Next, I investigate whether the model is able to explain the cross-sectional variation within years. Table D4 shows that, for almost all target moments, the R2 metrics are positive. The only exception is the variance of log wages for college workers, which is the moment with the worst fit in the aggregate. Table D4 also shows the corresponding measures of fit for the two benchmark OLS models described in Appendix D.5.5. Similar to the discussion of the overall quality of fit, the Simple OLS model is comparable to the structural model. The Large OLS model fits the data better in most dimensions, but again, the differences are not large with respect to the minimum wage bindingness measures, two-way fixed effects moments, and employment rate for workers with less than secondary.

The following exercise verifies the quality of fit regarding the spike and the share close to the minimum wage, separately by education. Those measures are not targeted by the estimation procedure, and thus serve as a test of whether the distributional assumptions on

Table D5: Minimum wage spike and share close to the minimum wage by education

	Data		Model		R2
	1998	2012	1998	2012	Model
Moments	(1)	(2)	(3)	(4)	(5)
Less than sec., up to 5 log points	0.041	0.077	0.042	0.108	0.572
Secondary, up to 5 log points	0.022	0.05	0.014	0.063	0.685
Tertiary, up to 5 log points	0.005	0.009	0.003	0.017	0.01
Less than sec., up to 30 log points	0.117	0.287	0.133	0.288	0.816
Secondary, up to 30 log points	0.054	0.22	0.055	0.211	0.89
Tertiary, up to 30 log points	0.01	0.032	0.013	0.06	-0.031

Notes: This table displays national averages by year and the R2 quality-of-fit measure for additional moments that are not targeted in the estimation procedure: the size of the spike and share close to the minimum wage by educational group.

worker productivity seem warranted. In addition, if β varies strongly by skill, instead of being common as assumed in the model, then the data and the model would likely disagree regarding the relative size of the spike for different educational groups.

Table D5 shows that this is not the case. The overall pattern of a good fit for the spike in 1998, and an over-estimate in 2012, holds for all worker types. The fit of share close to the minimum wage is excellent for workers with secondary or less. For college workers, the R2 metric is close to zero, but the shares are very low to begin with. Thus, the lack of excellent quality of fit there is likely not very consequential for counterfactual analysis.

Finally, I investigate whether the good quality of fit is being driven by the largest regions, which are more strongly weighted in the estimation procedure. In Table D6, I shows that this is not the case. That table follows the same structure of Table 5 shown in the text. The only difference is that region weights are not used to calculate the averages and R2 metrics. To be clear, this is not a separate estimation exercise: the same parameter estimates are being used to calculate the simulated moments in each region-time, both for the structural model and the benchmark OLS models. Quality of fit decreases a bit for all models, but the overall conclusions from the main text still hold.

Table D6: Quality of fit with equal weights for all regions

Moments	Data 1998 (1)	Data 2012 (2)	Model 1998 (3)	Model 2012 (4)	R2 Model (5)	Benckmark R2 Simple (6)	Benckmark R2 Large (7)
Wage inequality measures							
Secondary / less than secondary	0.478	0.131	0.494	0.116	0.717	0.723	0.764
Tertiary / secondary	0.978	0.953	1.022	0.911	0.031	-0.094	0.068
Within less than secondary	0.362	0.212	0.34	0.209	0.507	0.603	0.716
Within secondary	0.681	0.307	0.647	0.311	0.816	0.724	0.827
Within tertiary	0.755	0.612	0.712	0.628	0.141	0.362	0.42
Total variance of log wages	*0.633*	*0.442*	*0.672*	*0.455*	*0.617*		
Two-way fixed effects decomposition							
Variance establishment effects	0.101	0.049	0.107	0.048	0.457	0.383	0.462
Covariance worker, estab. effects	0.036	0.034	0.046	0.04	0.12	0.11	0.256
Variance worker effects	*0.37*	*0.317*	*0.406*	*0.283*	*0.25*		
Correlation worker, estab. effects	*0.193*	*0.256*	*0.215*	*0.333*	*-0.127*		
Formal employment rates							
Less than secondary	0.256	0.336	0.256	0.333	0.934	0.942	0.968
Secondary	0.425	0.509	0.424	0.509	1.0	1.0	1.0
Tertiary	0.534	0.632	0.533	0.636	0.836	0.917	0.936
Minimum wage bindingness							
Log min. wage - mean log wage	-1.237	-0.831	-1.237	-0.831	1.0	1.0	1.0
Share < log min. wage + 0.05	0.046	0.062	0.042	0.092	0.541	0.487	0.688
Share < log min. wage + 0.30	0.121	0.235	0.136	0.259	0.842	0.672	0.857

Notes: This table is identical to Table 5, except that all of the averages and R2 measures are calculated without using region weights.

D.6 Counterfactuals

D.6.1 Additional decomposition outcomes

Table D7 performs decomposition exercises identical to those in Table 6, but for different outcomes.

D.6.2 Demand shocks

As explained in the main text, I group several time-varying changes under the "demand" umbrella. There are two points to warrant further discussion. The first is why outside options were included as a demand shock. The second is on the interpretability of the effects of each component in isolation.

Table D7: Effects of supply, demand, and minimum wage on other outcomes

Outcome	Base value (1)	All changes (2)	Individual effects S (3)	D (4)	M (5)	Interactions S+D (6)	S+M (7)	D+M (8)	Triple (9)
Panel A: Inequality between and within groups									
Between groups: 2/1	0.49	-0.34	-0.09	-0.27	-0.06	0.11	0.01	-0.04	0.00
Between groups: 3/2	1.00	-0.06	-0.05	-0.04	-0.07	0.05	0.01	0.05	-0.01
Within group: 1	0.39	-0.16	0.03	-0.04	-0.15	-0.04	0.01	0.03	-0.01
Within group: 2	0.65	-0.31	-0.04	-0.28	-0.11	0.04	0.01	0.07	-0.01
Within group: 3	0.69	-0.05	-0.06	0.02	-0.08	0.03	0.01	0.04	-0.01
Panel B: Two-way fixed effects decomposition									
Variance of log wages	0.72	-0.22	0.04	-0.20	-0.14	0.02	0.01	0.05	0.00
Var. worker effects	0.42	-0.12	0.01	-0.07	-0.05	-0.01	0.01	-0.02	0.01
Var. estab. effects	0.12	-0.06	-0.01	-0.08	-0.01	0.02	-0.01	0.02	0.00
2×Cov. worker, estab	0.12	-0.02	0.03	-0.03	-0.07	0.01	0.01	0.04	-0.01
Var. residuals	0.07	-0.02	0.00	-0.02	-0.01	0.00	0.00	0.00	0.00
Panel C: Formal employment rates									
All workers	0.32	0.11	0.04	0.11	-0.03	-0.00	0.01	-0.01	0.00
Group 1	0.27	0.07	0.00	0.12	-0.03	-0.00	0.01	-0.02	-0.00
Group 2	0.44	0.07	-0.01	0.11	-0.03	0.00	0.01	0.00	-0.00
Group 3	0.54	0.09	-0.01	0.10	-0.00	0.01	-0.00	-0.00	0.00

Notes: This table is similar to Table 6, except that it shows a different set of outcomes.

The main reason for grouping outside options with demand shocks is conceptual, related to the interpretation of what is the final good. The model specifies two technologies to produce the final good: either home production or combining the two goods produced by firms. Shocks to θ_g, γ_g, and F_g are changing the second technology. It is plausible that such changes could also change the relative "quality" of the final good produced by using the second technology. Including the estimated change in $z_{0,h}$ parameters as part of the demand shock bundle is an effective way to allow for that possibility in an agnostic way.

Changes in the technologies used by formal firms may not be the only reason why the $z_{0,h}$ parameters changed. Another example, previously mentioned in the paper, would be changes in the enforcement of labor regulations that make the formal sector more or less appealing to some workers. Whether such a shock is on the supply or demand side is a matter of interpretation—in this paper, I classify them as demand shocks.

On the second point, it could be tempting to attach an economic interpretation to each component of the demand shock. Specifically, one could think of an increase in $\theta_{g=2,r,t}$ as skill-biased technical change (SBTC), and the reduction in the relative taste for the skill intensive good $\gamma_{g=2,r,t}/(1-\gamma_{g=2,r,t})$ as representing the commodities boom (which favored goods in the agricultural and mining sectors). To see why this interpretation is not warranted, consider SBTC. Given the formulation I use for the efficiency functions $e_h(x)$, an increase in $\theta_{2,r,t}$ leads to a relative increase in the cost for the skill-intensive good. But it would be reasonable to think that technological advancements such as personal computers, the internet, or programmable machines should reduce the cost of some goods that use skilled labor. Thus, SBTC may be better represented by a combination of primitives of the model, including not only θ_2 but also $\gamma_2/(1-\gamma_2)$ and F_2/F_1. A similar argument can be made for trade shocks, if, for example, higher demand for exports comes together with increases in quality requirements (Verhoogen, 2008).

Another way of framing this issue is that, to identify the independent effect of specific demand shocks such as SBTC or the commodities boom, we need additional exclusion restrictions. For example, one could impose the restriction that, in the empirical model of demand parameters, the interaction of the agricultural share with the time dummy corresponds to the effect of the commodities boom. I refrain from making such assumptions and focus instead on the role of demand shocks as a whole.

One may still be interesting to understand the mechanical effects of each shock in isolation. To that end, Table D8 decomposes the total demand shock.

D.6.3 Heterogeneity of minimum wage effects

The results from Table 8 are strongly heterogeneous along worker productivity categories, showing disemployment effects concentrated on those at the bottom of the productivity distribution. One possible counterpoint to those results is that, if they are true, then it should be fairly easy to detect such heterogeneity in reduced-form empirical designs. To the extent that those designs do not commonly find those negative effects, then that could constitute evidence against the model.

The problem with this argument is that it is difficult to condition on worker productivity in the data, which is almost always unobservable. Instead, the most common approach is to condition on a worker's wage before the introduction of the minimum wage. One potential pitfall of using this approach is that it may introduce bias from "regression to the mean." But even if that potential bias is addressed—as it is, for example, in Dustmann et al. (2021)—there is still the conceptual problem that wages are not equal to productivity in a model with firm wage premiums.

Table D8: Decomposition of demand shock

Outcome	All demand shocks (1)	Task demand (2)	Consumer taste (3)	Entry cost (4)	TFP and outside opt. (5)
Panel A: Inequality and sorting					
Mean log real wage	-0.06	-0.25	0.09	0.07	0.04
Variance of log wages	-0.20	-0.15	-0.13	0.06	0.02
Corr. worker, estab effects	0.08	-0.02	-0.12	0.21	0.02
Panel B: Inequality between and within groups					
Between groups: 2/1	-0.27	-0.24	-0.14	0.09	0.03
Between groups: 3/2	-0.04	0.09	-0.25	0.09	0.03
Within group: 1	-0.04	-0.08	0.04	0.00	0.01
Within group: 2	-0.28	-0.17	-0.17	0.05	0.02
Within group: 3	0.02	0.19	-0.06	-0.08	-0.02
Panel C: Two-way fixed effects decomposition					
Variance of log wages	-0.20	-0.15	-0.13	0.06	0.02
Var. worker effects	-0.07	-0.08	-0.04	0.04	0.01
Var. estab. effects	-0.08	-0.02	-0.03	-0.03	0.00
2×Cov. worker, estab	-0.03	-0.03	-0.05	0.04	0.01
Var. residuals	-0.02	-0.01	-0.01	0.01	0.00
Panel D: Formal employment rates					
All workers	0.11	-0.01	0.10	-0.10	0.12
Group 1	0.12	0.00	0.11	-0.10	0.11
Group 2	0.11	-0.04	0.10	-0.10	0.15
Group 3	0.10	-0.09	0.01	-0.03	0.20

Notes: Each column from (2) to (5) shows the marginal effect of changing each set of parameters described in the header. The decomposition is sequential. Column (3), for example, shows the effects of moving from models as of 1998, except that they have the θ_2 values of 2012; to other equilibria where the taste parameters γ_2 are also at their 2012 values.

I evaluate the consequences of this limitation in my empirical context with Table D9. It is identical to Table 8, except that workers are grouped by initial wage instead of productivity.[37] Consistent with the idea that wage groups are combinations of productivity groups, I find that the wage and employment effects extend into higher points of the distribution. Notably, if one ignores the equilibrium effects on returns to skill and entry, the elasticities of employment with respect to the mean wage become remarkably similar for the five bottom groups. That stability, however, is misleading under the lens of the model.

[37] This procedure requires assigning a wage to non-employed adults. For a given worker type (h, ε) in a given region, I split the non-formally employed across the two firm types according to the relative employment shares, and then assign the wage they would get at those firms. Then, all workers—employed or not—are ranked in increasing order according to that real or inputted wage, and the thresholds separating the groups are determined such that each of them corresponds to a similar amount of employment (as was done for Table 8).

Table D9: Wage and employment effects of the minimum wage by wage deciles

Wage. decile (1)	Pop. share (2)	Base wage (3)	Mean wage changes: Monops. (4)	Ret. sk. (5)	Gen. eq. (6)	Base emp. (7)	Emp. elasticities w.r.t.: Min. (8)	Mean (9)	·, monops. (10)
1	0.14	1.38	0.82	-0.08	-0.01	0.22	-0.26	-0.93	-0.69
2	0.12	1.86	0.75	-0.07	-0.01	0.27	-0.17	-0.89	-0.66
3	0.10	2.49	0.11	-0.07	0.02	0.30	-0.02	-1.95	-0.47
4	0.11	3.19	0.14	-0.08	0.02	0.31	-0.02	-1.97	-0.64
5	0.10	4.03	0.10	-0.08	0.02	0.32	-0.01	-2.37	-0.47
6	0.10	5.14	0.00	-0.08	0.03	0.34	-0.01		
7	0.09	6.27	0.00	-0.07	0.03	0.35	-0.00		
8	0.09	7.74	-0.00	-0.06	0.04	0.36	-0.00		
9	0.08	10.85	-0.00	-0.04	0.04	0.40	-0.00		
10	0.07	24.32	0.00	0.11	0.06	0.49	0.00		

Notes: Each row shows causal effects of an increase of 65 log points in the minimum wage in all regions for a subset of adults, grouped based on initial wage (see text for details). Wage effects are decomposed as described in Subsection 5.6: monopsony, returns to skill, and general equilibrium. Columns (8) and (9) report elasticities of employment with respect to the log real minimum wage or the mean wage for the group, respectively. Column (10) is similar to column (9) but only considers the monopsony channel.

D.6.4 Why do regressions find no employment effects of minimum wages in Brazil?

Finally, I address the issue of why previous reduced-form work studying the Brazilian case have not detected the negative employment effects. I focus on the descriptive results of Engbom and Moser (2022), as they study a similar period and the paper was recently published in a leading peer-reviewed journal. To be clear from the outset, this is not a criticism of that paper or of the authors. Indeed, they acknowledge the limitations of their reduced-form estimates, and most of their effort is spent in creating and estimating a structural model of the Brazilian economy. The point of this discussion is to argue that the identification of employment effects of minimum wages in the Brazilian context is challenging.

Engbom and Moser (2022) exploit variation in the "effective minimum wage," that is, the log of the national minimum wage minus the median log wage in each state-time combination, which they refer to as the Kaitz-50 index.[38] They run regressions of formal employment on the effective minimum wage, its square, and controls. This approach has a long tradition in labor economics, going back at least as far as Neumark and Wascher (1992, who used the minimum wage relative to the mean in the state-year instead of the median). In the specification they report in the paper, Engbom and Moser (2022) use state fixed effects and

[38] In other papers, the Kaitz index may be defined differently. In this discussion, I use their nomenclature.

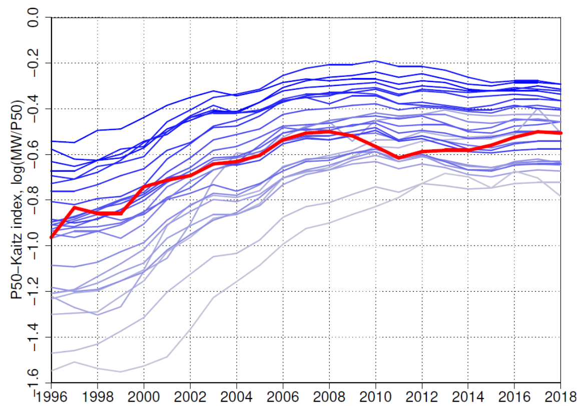


Figure D3: Variation in effective minimum wages at the state-time level

Notes: This is a copy of Appendix Figure B.10 in Engbom and Moser (2022). It shows the variation used to identify the effects of minimum wages on employment in Brazil.

state-specific time trends as controls.

One problem with this approach is that the median wage, used to construct the effective minimum wage, is an endogenous object. As emphasized in this paper, wages are determined at the local labor market level by a combination of region-specific supply and demand parameters. They correlate with each other, and also with local TFP. That introduces correlations between those factors and the Kaitz-50 index. On the supply side, I find that microregion-level changes in educational achievement are positively correlated with the change in the Kaitz-50 index, which is somewhat surprising. In addition, Table D3 shows that the current Kaitz index is a statistically significant predictor of demand-side parameters after controlling for initial characteristics at the microregion, with coefficients that vary between years. Those correlations may introduce omitted variable bias because all of those supply and demand shocks have large effects on employment rates even in the absence of minimum wage changes, as shown in Table D7. And because they correlate in differences, not only in levels, their effect is not absorbed by the state fixed effects.

To tackle those time-varying confounders, Engbom and Moser (2022) include region-specific time trends in regression models with many periods (the panel is at the yearly level, from 1996 through 2018). Intuitively, the assumption behind this approach is that the influence of these confounders on employment is well approximated by the linear trends, while the influence of the minimum wage is nonlinear. Another way of visualizing that assumption is: if one takes time differences two times for both employment rates and the Kaitz index, then the relationship between those transformed variables should reflect the impact of lo-

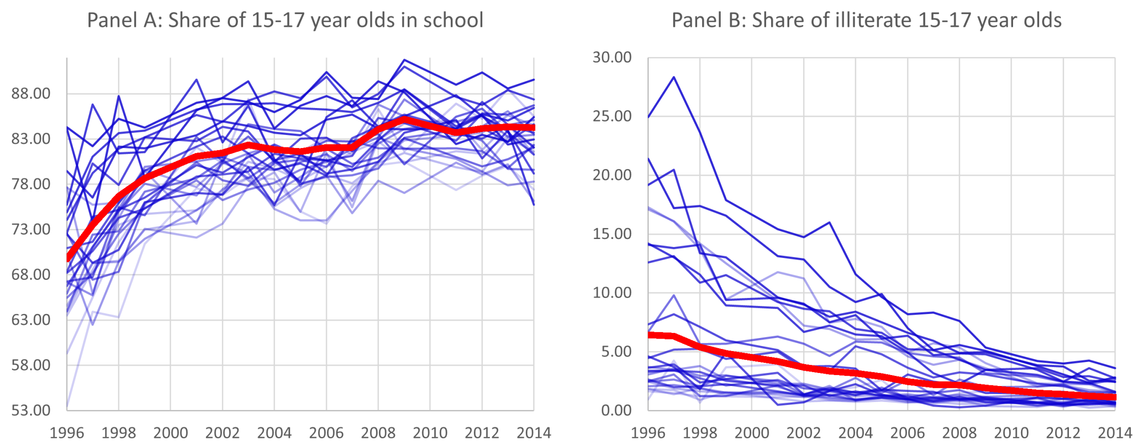


Figure D4: Evolution of educational outcomes by state

Source: PNAD survey. The series were obtained using the IpeaData online tool (available at http://www.ipeadata.gov.br).

cational changes in the latent distribution of wages (what could be described as TFP), not the direct effect of compositional changes between groups that have different intrinsic employment rates or of biased demand shocks that affect the latent productivity distribution and employment rates differently from a locational shift.

For the minimum wage, the non-linear part of the variation comes from faster minimum wage growth in the first half of the sample. This is evident from Figure D4, which is a copy of Appendix Figure B.10 from Engbom and Moser (2022)). The red thick line shows the national average for the Kaitz-50 index, while the blue lines show the Kaitz-50 index for each state.

Is the variation in minimum wages more nonlinear than the supply and demand shocks affecting the Brazilian economy? Below, I argue that this is not the case. Figure D4 shows two metrics related to the supply of young educated adults: the share of those between 15 and 17 who are in school, and the share of those between 15 and 24 who can read. Both graphs show steeper slopes early in the period, similarly to the minimum wage graph. This is an important issue, since formal employment rates vary dramatically by educational level. And among all adults, the young are more likely to be affected by the minimum wage.

A similar argument can be made for demand shocks. The variation in international commodity prices, shown in Figure D5, suggests that the influence of demand shocks may be much less smooth and monotonic than the impacts of minimum wages. In addition, Figure 2 in Costa, Garred and Pessoa (2016) shows that trends in Brazilian imports from, and exports to

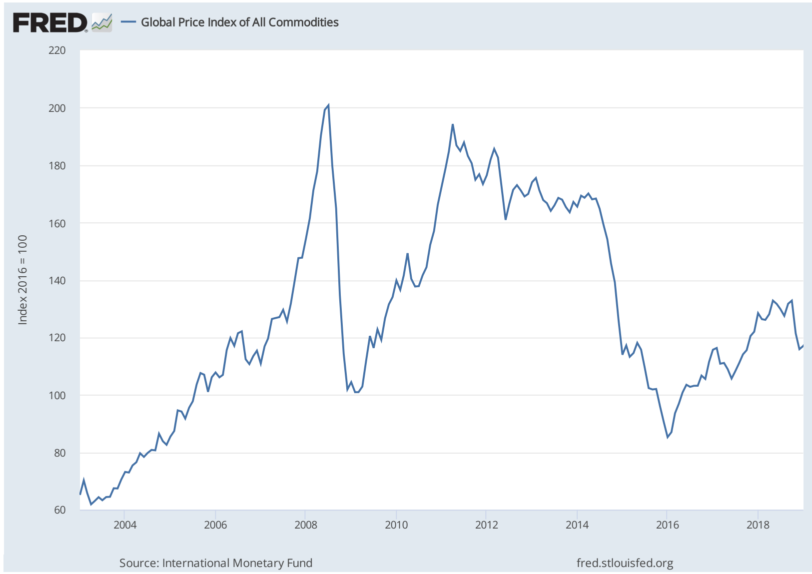


Figure D5: Global Price Index of All Commodities

China are also nonlinear. The export trends is nonmonotonic, and considerably further from the a line than trends in the Kaitx-50 index. Costa, Garred and Pessoa (2016) goes on to show that shocks to Chinese supply and demand have significant labor market effects at the microregion level.

One could think about alternative regression specifications, such as adding time fixed effects or higher-order trends at the state level. However, those approaches are not likely to solve the problem. That is because those terms absorb not only the confounders, but also the "good" variation introduced by the national minimum wage. The fundamental problem is the lack of a quasi-experiment that manipulates the minimum wage independently of other factors.

In addition to the possibility of omitted variable bias, the regressions may find no effects because they may measure short-run, instead of long-run, effects. To see why, note that the inclusion of state-specific trends means that the identifying variation is not coming from the long-run trend towards higher minimum wages. Instead, identification comes from deviations around these long-run trends: is employment particularly lower in years where the minimum wage is higher relative to the state-specific trend? If it takes time for the effects of minimum wages to materialize, then the regression will likely not detect them.

One can think of the structural approach used in this paper as a model designed to control for the influence of the supply and demand factors. The variation used to measure the effects of minimum wages is fundamentally the same: differences in bindingness of the minimum wage across regions, stemming from structural differences in education, total factor productivity, and local demand for skills. The effect of those local-level confounders is inferred from a series of additional outcomes at the local level, such as measured sorting. Thus, it provides a principled way to deal with those confounders.

Appendix Table D4 provides a test of whether the strong disemployment effects are rejected by the data. Specifically, if the employment effects predicted by the model were strongly at odds with what was observed at the microregion level, one would expect the R2 metric for the formal employment rate of workers with less than secondary in 2012 to be bad. Instead, it is 0.905.

The weakness of the structural approach is that it only measures effects of causal channels pre-specified by the econometrician. Given that my framework includes a uniquely wide array of causal pathways for the minimum wage, and given the threats that affect reduced-form designs in the Brazilian case, I believe that my estimates of minimum wage effects are the most reliable in this context. See Appendix B.4 for a discussion of minimum wage causal channels not included in my framework and why I believe adding them would not make a significant difference for my results.