Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution [*]

Daniel Haanwinckel[†]
Job Market Paper

December 31, 2018
Click here for the latest version

Abstract

This paper builds a tractable framework for analyzing the equilibrium effects of labor supply shocks, technical change, and minimum wages in an imperfectly competitive labor market environment with worker and firm heterogeneity. Goods are produced using task-based technologies exhibiting imperfect substitution between worker types. Firms specialize in the production of particular goods, which leads to differences in task requirements, entry costs, and workplace amenities. These differences generate firm heterogeneity in skill intensity, size, and wages. The model has three advantages relative to the canonical supply-demand-institutions framework typically used to study trends in wage inequality. First, task-based production with multiple worker types allows for plausibly rich formulations of the structure of technical change. Second, the model accounts for equilibrium effects of minimum wages, compressing the wage distribution and generating spillovers on quantiles where the minimum wage does not bind. Third, the model makes predictions regarding labor market sorting and cross-firm wage dispersion. I take a simple version of the model to Brazilian matched employer-employee data and show that it can fit several aspects of wage inequality: differences in mean log wages between educational groups, within education group variances, and two-way variance decompositions of log wages into worker and firm components. The model also matches reduced-form estimates of minimum wage spillovers. I use the estimated parameters to decompose observed changes in inequality and sorting into components attributable to increasing schooling achievement, technical change, and a rising minimum wage. Falling wage inequality in Brazil is primarily due to the minimum wage, while rising worker-firm assortativeness is found to be driven by technical change. The decomposition exercise also illustrates how responses to supply and demand shocks differ qualitatively from those predicted by models with a representative firm.

1   Introduction

A central task in labor economics is understanding the source of changes in the wage distribution. Three sets of explanatory factors have received the most attention in this literature: trends in the relative supply of skills, such as increasing college completion rates; shocks to relative demand for skills, such as skill-biased technical change; and changes in labor market institutions, such as minimum wages. The dominant approach in this literature employs competitive labor market models with a constant elasticity of substitution production function to quantify the relative importance of these factors (Bound and Johnson, 1992; Katz and Murphy, 1992; Card and Lemieux, 2001). This approach successfully captures trends in mean log wage gaps between broadly defined worker groups (e.g., college versus high school) and is still used by leading researchers (Autor, 2014). But it has limitations on three fronts. First, it cannot match other measures of wage inequality using its parsimonious formulation of demand shocks as skill-biased technical change (Card and DiNardo, 2002; Autor, Katz and Kearney, 2008). Second, the focus on between-group wage gaps also prevents that approach from fully accounting for the effects of changing minimum wages (DiNardo, Fortin and Lemieux, 1996; Lee, 1999). Third, it cannot be used to study between-firm wage dispersion for similar workers, a phenomenon that is now extensively documented (Manning, 2011; Hornstein, Krusell and Violante, 2011; Card et al., 2018) and that some economists suggest might have implications for the evolution of inequality (Card, Heining and Kline, 2013; Alvarez et al., 2018; Song et al., 2018).

Different strands of the wage inequality literature endeavor to address these limitations. Task-based models of comparative advantage are used to evaluate equilibrium effects of minimum wages (Teulings, 2000, 2003) and to model richer versions of demand-side shocks that provide a better fit to the data (Costinot and Vogel, 2010; Acemoglu and Autor, 2011; Lindenlaub, 2017). This approach still assumes competitive markets. Another strand tackles between-firm wage dispersion using models with search frictions (see Lentz and Mortensen (2010); Chade, Eeckhout and Smith (2017)) or monopsony power (see Ashenfelter, Farber and Ransom (2010)). Models in that strand do not account for the role of supply and demand in changing the marginal product of labor between firm types. There is also emerging reduced form literature finding that increased assortativeness between high wage workers and high wage firms, estimated with two-way fixed effects regressions (henceforth AKM regressions, after Abowd, Kramarz and Margolis (1999)), explains a substantial share of increased wage inequality in some countries (Card, Heining and Kline, 2013; Song et al., 2018). There is no agreement, however, on what causes changes in sorting — or, more fundamentally, on whether these AKM regressions are meaningful (Eeckhout and Kircher, 2011).

In this paper, I propose a tractable task-based framework that captures the equilibrium effects of technical change, labor supply shocks, and minimum wages on the wage distribution, while allowing for realistic worker-to-firm sorting patterns and firm-level wage premia. After studying the theoretical properties of the model, I show that it can be an effective quantitative tool for analyzing wages and sorting. The model can match several forms of wage dispersion: between worker groups, within groups, and across firms among similar workers. It can also match firm to worker sorting patterns and the causal effects of minimum wages measured using reduced form methods. The analytical and quantitative results reveal interactions between supply, demand, institutions, and firms that are important for parsing the contribution of each underlying factor.

In the model, firms produce goods by combining tasks in different proportions. Tasks, in turn, are produced using labor, with more skilled workers having a comparative advantage in more complex tasks. The task-based production function is the solution to the within-firm problem of assigning workers to tasks with the goal of maximizing production. I study the properties of this production function and show that it provides an intuitive and parsimonious way to model heterogeneity in skill intensity across firms, via differences in task requirements for different goods.

Next, I construct a long-run model of imperfectly competitive labor markets and study the determinants of between-firm wage differentials and labor market sorting. In the model, workers have preferences over employers. Firms can set wages below the marginal product of labor, extracting rents from infra-marginal employees that enjoy working there. Firm-level wage premia arise if some firms have higher entry costs than others, such that they locate at different points of the labor supply curve, or if they have worse amenities, such that wages compensate for undesirable workplaces. Additionally, when firms differ in task requirements, they pay more to worker types that they use relatively more intensively. Thus, AKM regressions of log wages, where firms vary only in a proportional term (e.g., firm A pays 10 percent more to all workers relative to firm B), are in general misspecified in this model. Nevertheless, I show that a decomposition of the variance of log wages based on AKM regressions may still be informative about parameters governing between-firm wage dispersion and labor market sorting.

I derive two results on how this economy responds to structural shocks. First, the model admits a balanced growth path where technical progress is conceptualized as shifts towards more complex tasks in the production of all goods (a form of skill-biased technical change) accompanied by increased productivity. If skill levels and minimum wages rise in tandem, the shape of the wage distribution is not affected. Second, substitution in consumption creates an additional channel through which wages are affected by imbalances in the race between supply, demand, and institu-

tions. Imbalances have direct effects on the costs of labor: for example, a higher minimum wage makes low-skilled workers more expensive relative to high-skilled workers. The ensuing changes in the prices of different goods drive substitution in consumption. In the minimum wage example, consumers substitute away from products that are intensive in low-complexity tasks, which are produced by firms using mostly low-skilled labor. Finally, because the set of tasks being produced in that economy has changed, marginal productivity gaps are affected. Without substitution between goods, a plot of the impact of the minimum wage on quantiles of the wage distribution is a decreasing curve; with substitution, it becomes less steep and can change to a U-shaped curve.

To study the quantitative performance of this framework, I analyze trends in wage inequality and labor market sorting in Rio Grande do Sul state, Brazil using a parsimonious parameterization of the model's primitives. Wage inequality has fallen in that state, following minimum wage hikes and accelerated gains in schooling. Labor markets are also becoming more assortative, as measured using the leave-out estimator of Kline, Saggio and Sølvsten (2018). This combination makes for an interesting case study where trends in inequality and minimum wages are mirror images of what has happened in the US, while changes in sorting go in the same direction. I employ a minimum distance estimator that targets levels and changes in (i) mean log wage gaps between educational groups, (ii) within-group variances of log wages, (iii) decompositions of log wages using AKM regressions, (iv) measures of how binding the minimum wage is, and (v) reduced form estimates of minimum wage spillovers (causal effects on quantiles of the wage distribution where the minimum wage does not bind) obtained using the methodology of Autor, Manning and Smith (2016). The model can closely match these targets, despite being heavily over-identified.

I use the estimated model to create counterfactuals that isolate the role of supply, demand, and minimum wages in explaining changes in inequality and sorting. With monopsonistic labor markets and firm heterogeneity, these shocks change both the magnitude of firm-level wage premia and which workers earn them, in addition to affecting marginal productivities of labor. These additional effects are illustrated by the role of the demand shock in Rio Grande do Sul, Brazil. The estimated shock includes three components: a drift towards more complex tasks for all goods (i.e., skill-biased technical change), a reduction in the entry cost gap between goods, and a similar convergence in productivity. That shock increases wages for college-educated workers, as it would in most models of labor demand with skill-biased technical change. However, its overall effect on the variance of log wages is negative; indeed, it accounts for almost 40 percent of the overall decline in inequality. That result follows from reductions in cross-firm wage dispersion for all worker groups, particularly those with more education. These reductions are caused by changes in both worker-to-firm sorting patterns and the magnitude of wage premia. I also find that the demand shock is the

main contributor to the observed increase in the correlation between worker and firm fixed effects in AKM regressions.

Minimum wages are the most important factor behind decreased inequality. This shock alone accounts for more than 60 percent of the change in the variance of log wages. It has no effects, however, on the share of the variance attributed to firm effects or the correlation of worker effects and firm effects in the AKM decomposition. The increase in the relative supply of high school and college workers reduces the mean log wage gaps between these workers and those with less education. But its effects on the total variance of log wages is negligible.

The paper is organized as follows. The next section presents the task-based production function. The third section describes the labor market model and provides analytical results. The fourth section contains the quantitative exercise. The last section concludes with a discussion of two directions for further research: adding capital to the task-based production function and using this framework to study the inequality effects of international trade.

2   The task-based production function

Task-based models of comparative advantage are increasingly used to model wage inequality. Acemoglu and Autor (2011) show these models are better suited than the "canonical" constant elasticity of substitution (CES) model of labor demand to study inequality trends in the US. Teulings (2000, 2003) shows that substitution patterns implied by assignment models make them particularly suitable for studying minimum wages. Costinot and Vogel (2010) develop a task-based model to study the consequences of trade integration and offshoring, finding that it offers new perspectives relative to workhorse models of international trade.

In this section, I show an additional advantage of the task-based approach: it allows for intuitive, tractable, and parsimonious modeling of firm heterogeneity, whereby firms have production functions with imperfect substitution and differ in their demand for skill.

The production structure in this paper is built upon four assumptions. First, final goods embody a set of tasks that vary in complexity, combined in fixed proportions. Second, tasks cannot be traded. Third, workers are perfect substitutes in the production of any particular task, but with different productivities. Fourth, some worker groups have comparative advantage in the production of complex tasks relative to others.[1]

---

[1]There exists a parallel between the task-based production function developed here and models of hierarchical

I start this section by defining the production function and solving the managerial problem of assigning workers to tasks. The second subsection discusses cost minimization and shows how this structure generates differences in skill intensity between firms. The third subsection derives and explains distance-dependent substitution. The final subsection presents the parametric version that is employed in the quantitative exercises of this paper. All proofs are in Appendix A.

2.1 Setup, definitions, and the assignment problem

Workers in this economy are characterized by their type $h \in \{1, \ldots, H\}$ and the amount of labor efficiency units they can supply, $\varepsilon \in \mathbb{R}_{>0}$. Workers use their labor to produce *tasks* which are indexed by their *complexity* $x \in \mathbb{R}_{>0}$. All labor types are perfect substitutes in the production of any particular task, but their productivities are not the same:

Definition 1. *The **comparative advantage function** $e_h : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ denotes the rate of conversion of worker efficiency units of type h into tasks of complexity x. It is continuously differentiable and log-supermodular:* $h' > h \Leftrightarrow \frac{d}{dx}\left(\frac{e_{h'}(x)}{e_h(x)}\right) > 0 \ \forall x.$

To fix ideas, consider two workers, whom I will refer to as Alice and Bob. Alice, characterized by h, ε, can use a fraction $r \in [0,1]$ of her time to produce $r\varepsilon e_h(x)$ tasks of complexity x. Bob (h', ε'), who belongs to a lower type $(h' < h)$, can still produce more of those tasks than Alice, so long as his quantity of efficiency units is high enough relative to hers $(\varepsilon' > \varepsilon e_h(x)/e_{h'}(x))$. But Alice has a comparative advantage: moving towards more complex tasks increases her productivity relative to Bob's.

The interpretation of task complexity depends on how worker groups are defined. In the quantitative exercise of this paper, workers are grouped by educational achievement, and thus more complex tasks are those that benefit from formal education (or intrinsic characteristics that correlate with formal education). The assumption that all tasks are ordered in a single dimension of complexity is strong. Autor, Levy and Murnane (2003), for example, have a multidimensional definition of task complexity; in their case, manual versus analytic and routine versus non-routine. For a quantitative model of the impact of technological change on wage inequality with multi-dimensional tasks, see Lindenlaub (2017).

Because workers in the same group differ only in a proportional productivity shifter, the sum of

---

firms (Garicano, 2000; Garicano and Rossi-Hansberg, 2006; Antràs, Garicano and Rossi-Hansberg, 2006; Caliendo and Rossi-Hansberg, 2012), once one reinterprets tasks in my model as "problems" in those models. The key difference is that I ignore costs of information transmission within the firm, adding tractability by simplifying the assignment of workers to problems/tasks.

efficiency units of each type is a sufficient statistic for analyzing production. Thus, throughout this section, definitions and results are in terms of total efficiency units of each type available to the firm, which I denote by $l = \{l_1, \ldots, l_H\}$ (bold-faced symbols denote vectors over worker types throughout the paper). The distinction between labor efficiency units and workers will be relevant in the next section, when discussing labor markets and the wage distribution.

There is a discrete number of final consumption goods, $g = 1, \ldots, G$. Each good is produced by combining tasks in fixed proportions:

Definition 2. *The **blueprint** $b_g : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a continuously differentiable function that denotes the density of tasks of each complexity level x required for the production of a unit of consumption good g. Blueprints satisfy $\int_0^\infty b_g(x)/e_H(x)dx < \infty$ (production is feasible given a positive quantity of the highest labor type).*

Tasks cannot be traded; firms must use their internal workforce to produce them. The justification for this assumption is that there are unmodeled costs that make task exchange between firms unprofitable, in the spirit of Coase (1937).[2] I assume that firms are allowed to split worker's time across tasks in a continuous way by choosing assignment functions $m_h : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$, where $m_h(x)$ denotes the intensity of use of efficiency units of labor type h on tasks of complexity x. The only restriction imposed on $m_h(\cdot)$ is that these functions are right continuous.[3] That formulation of the assignment problem is very general, allowing firms to use multiple worker types to produce the same task, the same worker type in disjoint sets of tasks, and discontinuities in assignment rules.

Given a blueprint $b(\cdot)$ and l efficiency units of labor, firms choose these assignment functions with the goal of maximizing output. In this problem, they are subject to two constraints: producing the required amount of tasks of each complexity level x and using no more than l_h units of labor of type h.

Definition 3. *The **task-based production function** $f : \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{>0} \times \{b_1(\cdot), \ldots, b_G(\cdot)\} \to \mathbb{R}_{\geq 0}$ is*

---

the value function of the following assignment problem:

$$f(\boldsymbol{l};b_g) = \max_{q \in \mathbb{R}_{\geq 0}, \{m_h(\cdot)\}_{h=1}^{H} \subset RC} q$$

$$s.t. \quad qb_g(x) = \sum_h m_h(x)e_h(x) \quad \forall x \in \mathbb{R}_{>0}$$

$$l_h \geq \int_0^\infty m_h(x)dx \quad \forall \in \{1,\ldots,H\}$$

where q is output and m_h is an assignment function denoting the density of labor efficiency units of type h used in the production of each task x. RC is the space of right continuous functions $\mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$.

The definition of the production function assumes positive input of the highest worker type. This assumption simplifies proofs and ensures the well-behaved derivatives, while not being restrictive for the applications in this paper. In general, 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.

Comparative advantage implies that the optimal assignment of workers to tasks is assortative:

Lemma 1 (Optimal allocation is assortative). *For every combination of inputs $(\boldsymbol{l},b_g(\cdot))$, there exists a unique set of $H-1$ complexity thresholds $\bar{x}_1(\boldsymbol{l},b(\cdot)) < \cdots < \bar{x}_{H-1}(\boldsymbol{l},b_g(\cdot))$ that define the range of tasks performed by each worker type in an optimal allocation:*

$$m_h(x) = \begin{cases} q\dfrac{b_g(x)}{e_h(x)} & \text{if } x \in [\bar{x}_{h-1},\bar{x}_h) \\[2mm] 0 & \text{Otherwise} \end{cases}$$

where I omit the dependency on inputs $(\boldsymbol{l},b_g(\cdot))$ and set $\bar{x}_0(\cdot) = 0$, $\bar{x}_H(\cdot) = \infty$ to simplify notation. Thresholds satisfy:

$$\frac{e_{h+1}(\bar{x}_h)}{e_h(\bar{x}_h)} = \frac{f_{h+1}}{f_h} \quad h \in \{1,\ldots,H-1\} \tag{1}$$

where $f_h = f_h(\boldsymbol{l},b_g(\cdot)) = \frac{d}{dl_h}f(\boldsymbol{l},b_g(\cdot))$ denotes marginal product of labor h, which is strictly positive.

Lower types specialize in low complexity tasks and vice-versa. Equation (1) means that the shadow cost of using neighboring worker types is equalized at the task that separates them. This result is useful for obtaining compensated labor demands, as described in the next subsection.[4]

---

[4]In general, the task-based production function and its derivatives do not have simple closed-form representations.

2.2 Compensated labor demand and sorting of workers to firms

To study the properties of this production function, I start by considering its implications in a competitive labor market, where the cost of acquiring efficiency units of each type is given by $\boldsymbol{w} = \{w_1, \ldots, w_H\}$. When firms choose labor quantities by minimizing production costs, marginal productivity ratios equal wage ratios. It then follows from Equation (1) that:

$$\frac{e_{h+1}(\bar{x}_h)}{e_h(\bar{x}_h)} = \frac{w_{h+1}}{w_h}$$

Because the ratio on the left-hand side is strictly increasing in $\bar{x}_h$, this expression pins down all task thresholds as functions of wage ratios and comparative advantage functions. Since neither are firm-specific, thresholds are common across firms in competitive economies.

The compensated labor demand is then given by:

$$l_h(q, b_g, \boldsymbol{w}) = q \int_{\bar{x}_{h-1}(\boldsymbol{w})}^{\bar{x}_h(\boldsymbol{w})} \frac{b_g(x)}{e_h(x)} dx \tag{2}$$

Figure 1 illustrates how differences in blueprints reflect into differences in the internal workforce composition of firms. The graphs at the top show the compensated labor demand integral above. The heavy, continuous line is the blueprint, which varies across graphs (becoming more intensive in high complexity tasks from left to right). The vertical dashed lines are the thresholds, which are common for all graphs. The colored areas are the labor demand integrals. The compensated labor demand is shown again in the bottom row, in the form of blueprint-specific wage distributions within the firm (weighted by efficiency units).

If labor markets are not competitive, as in labor market model described in the next section, thresholds might differ across firms. Firms using different blueprints will still differ in the skill composition of their internal workforce, though possibly less so than in the competitive benchmark.

The concept of firms in this model is significantly different from that in the literature on labor market sorting (Shimer, 2005; Eeckhout and Kircher, 2011; Gautier, Teulings and van Vuuren, 2010; Gautier and Teulings, 2015; Lise, Meghir and Robin, 2016; Grossman, Helpman and Kircher, 2017; Lindenlaub and Postel-Vinay, 2017; de Melo, 2018; Eeckhout and Kircher, 2018). Most models in this literature focus on sorting of workers to *jobs* (or, equivalently, to firms that employ

---

If one needs to evaluate output and marginal productivities as a function of labor inputs, first solve the system of H compensated labor demand equations (2) on q and the $H - 1$ thresholds. Next, use equation (1) to calculate marginal productivity gaps. Finally, use the constant returns relationship $q = \sum_h l_h f_h$ to normalize marginal productivities.

Figure 1: Compensated labor demand in competitive labor markets
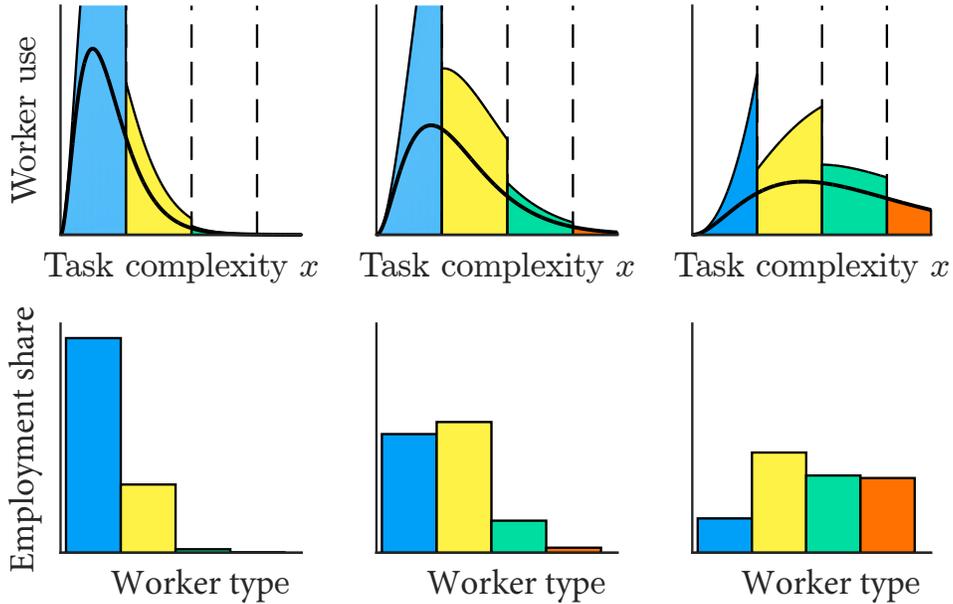
Notes: Graphs on the top show the compensated labor demand integrals for different blueprints. The vertical dashed lines are task thresholds, common for all firms in competitive labor markets. The solid continuous curve is the blueprint showing task requirements of each complexity level. The colored areas are compensated labor demands for each type, which are integrals of task requirements divided by the efficiency of labor at each complexity level. Graphs on the bottom display the resulting employment shares corresponding to each blueprint.

exactly one worker). Even in the ones with a concept of large firms, such as Eeckhout and Kircher (2018), any degree of within-firm wage dispersion is a sign of inefficiencies introduced by search frictions; if markets are competitive, each firm hires workers of a single type. I contribute to this literature by introducing a more realistic concept of firms as bundles of jobs (tasks in my model), coupled with a technology to acquire workers. In addition to having welfare implications, this distinction is relevant for quantitative studies where model predictions are matched to firm-related moments, such as the between-firm share of wage inequality or variance decompositions from AKM regressions.

2.3 Substitution patterns and distance-dependent complementarity

The task-based structure might appear exceedingly flexible at first glance, due to the infinite-dimensional blueprints and efficiency functions. Proposition 1 extends the results in Teulings

(2005) and shows that, on the contrary, there are strong constraints on substitution patterns.[5] Locally, the $H \times (H-1)/2$ partial elasticities of complementarity or substitution depend only on factor shares and at most $H-1$ scalars ρ_h — the same number of elasticity parameters in an equally-sized nested CES structure. However, unlike with a CES, there is a straightforward way to impose further restrictions on the number of parameters (both elasticities and productivity shifters for each worker type), via parametrization of blueprints and efficiency functions.

Proposition 1 (Curvature of the production function). *The task-based production function is concave, has constant returns to scale, and is twice continuously differentiable with strictly positive first derivatives. 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 $h < h'$:*

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

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

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

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

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

The curvature of the task-based production function reflects division of labor within the firm. Suppose that, initially, a firm only employs Alice, who belongs to the highest type H. In that case, output is linear in the quantity of labor bought from Alice. Adding another worker, Bob, of a lower type increases Alice's productivity, because she can now specialize in complex tasks while Bob takes care of the simpler ones. At that point, decreasing returns to Alice's hours reflect a reduction in those gains from specialization.

The impact of adding a third worker, Carol, on the marginal productivities of Alice and Bob depends on Carol's skill level (in terms of comparative advantage), relative to Alice's and Bob's:

---

[5]Teulings (2005) derives elasticities of complementarity for a similar model, but using parametric efficiency functions and taking a limit where the number of worker types grows to infinity.

Figure 2: Distance-dependent complementarity
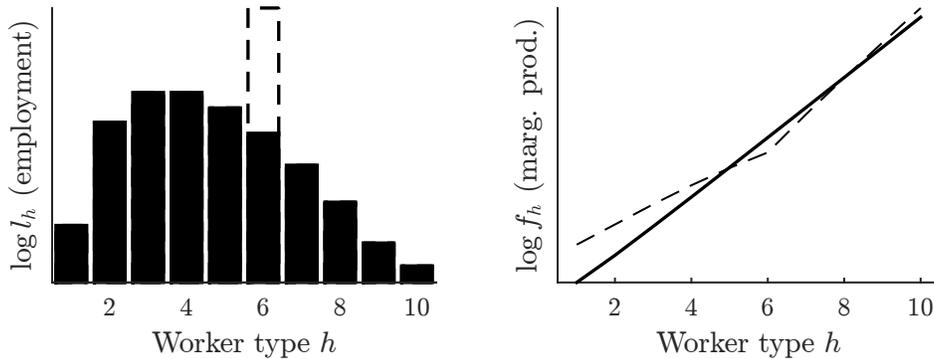
Notes: This figure shows the impact of adding workers of a given type on the marginal productivity of all types. In this example, the initial labor endowments of the firm are shown as solid bars in the graph on the left, and the increase in labor of type 6 is the dashed bar. The solid line in the graph on the right shows initial marginal productivities for each labor type, and the dashed line shows marginal productivities after the shock. Nearby types are substitutes to labor type 6, while types far away are complements.

Corollary 1 (Distance-dependent complementarity). *For a fixed h, the partial elasticity of complementarity is strictly increasing in h' for $h' \geq h$ and strictly decreasing in h' for $h' \leq h$.*

Close types perform similar tasks and are net substitutes; distant types perform very different tasks and are complements. The distance-dependent complementarity pattern is illustrated in Figure 2. The left panel shows baseline log employment by worker type (in solid bars) and a shock to employment of workers of type 6 (bar with dashed contour). The right panel shows baseline log marginal productivities (solid line) and marginal productivities after the employment shock (dashed line). Workers of type 6 suffer the largest relative decline in marginal productivity, followed by neighbor types 7 and 5. Marginal productivities increase for types that are further away, both low-skilled and high-skilled.

Distance-dependent complementarity has important implications for modeling minimum wages, a point made by Teulings (2003):

Corollary 2 (Effects of the minimum wage). *Consider a competitive economy with a representative task-based production function where the minimum wage only binds for workers of the lowest type $h = 1$. Then, a marginal increase in the minimum wage reduces all wage gaps $w_{h'}/w_h$ with $h' > h$.*

A minimum wage that is only binding for the lowest type will have spillover effects along the wage distribution, reducing wage gaps between any pair of worker types. This is in contrast to the CES case, where a small increase in the minimum wage would raise wages for the lowest type but keep wage gaps between other types constant. This property is useful for matching reduced form

estimates of minimum wage spillovers (see Figure 7).

2.4 Parametric example

Consider the following parametrization, which I use in the quantitative exercises of this paper:

Example 1 (Exponential-Gamma parametrization)**.**

$$e_h(x) = \exp(\alpha_h x) \qquad\qquad -1 = \alpha_1 < \alpha_2 < \cdots < \alpha_{H-1} < \alpha_H = 0$$

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

The exponential function is a straightforward way to generate log-supermodularity and is used in other models of comparative advantage (e.g. Krugman (1985); Teulings (1995)). Differences in the α_h coefficients determine the degree of comparative advantage between any two worker types. The expression for blueprints is the probability density function of a Gamma distribution divided by a "productivity" term z_g. Doubling z_g divides the quantity of tasks needed per unit of output by two, effectively doubling physical productivity.

Appendix B presents the mapping between marginal productivity gaps and task thresholds in this parametrization, as well as formulas for compensated labor demand integrals in terms of incomplete Gamma functions. These formulas are useful because they dispense with numerical integration, improving computational performance. Incomplete Gamma functions are readily available in software packages commonly used by economists.

The parameter θ_g is related to average task complexity. All else equal, firms with higher θ_g require more complex tasks and employ more skilled workers (in terms of comparative advantage). Increases in task complexity over time, modeled as changes in θ_g, provide an intuitive way to model skill-biased technical change because higher complexity is linked to increasing returns to skill (measured as the worker group h). The shape parameter k_g determines the dispersion of tasks. If two firms differ only in this parameter, the one with the smallest k_g has fatter tails. Thus, differences in k_g in the cross-section translate into some firms being more specialized than others in their hiring patterns.

This approach allows for modeling firm-level differences in skill intensity, skill dispersion, and productivity with a small number of parameters, while ensuring sensible substitution patterns within all firms. To understand the economic content behind those parametric restrictions, consider an

example of two firms in the retail sector. One, with a low θ_g, is a small local shop, while the other, with a high θ_g, is a large online retailer. In the first one, most tasks are of low complexity, measured in terms of how they benefit from schooling: stocking shelves, operating the register, and cleaning. In those tasks, workers with little formal education can substitute for others with a college degree. Because workers with a college degree cost much more, that first firm mostly hires less educated workers. In contrast, the online retailer is intensive in tasks such as web design, system administration, and business analytics, where college-graduated workers usually perform much better. This is why those firms find it profitable to use a more skilled workforce.

3   Markets and wages

This section builds a labor market model with monopsonistic firms and free entry. The first subsection lays out the structure of the economy. The second subsection describes the functioning of labor markets, solves the problem of the firm, and shows an important property of the model: goods encapsulate firm heterogeneity in skill intensity and wages. The third subsection derives analytical results on what determines wage differentials between firms and how the wage distribution changes over time.

This is the point of departure from other task-based assignment models of comparative advantage such as Teulings (2005), Costinot and Vogel (2010), and Acemoglu and Autor (2011). The contributions of the previous section fit inside that literature: new formulas for elasticities of complementarity and substitution, along with the convenient exponential-gamma parametrization. This section introduces more significant deviations. First, aggregate demand for tasks is CES in all papers in that literature, but not in this model. This has implications for comparative statics. For example, the introduction of a minimum wage always decreases wage gaps in Teulings (2000), but here the same shock might cause wage polarization. Second, labor markets are not competitive. And third, workers in this model differ not only in comparative advantage but also in absolute advantage, generating realistic distributions of log wages that include bunching at the minimum wage.

3.1   Factors, goods, technology, and preferences

Consider an economy with $N = \{N_1, \dots, N_H\}$ workers of each type h, and a large number of entrepreneurs. Entrepreneurs own entrepreneurial talent, whose total stock in the economy is T

and which is used to create firms. The model is static.

There are G final goods in this economy, interpreted as differentiated varieties within an industry.[6] An entrepreneur j may set up a firm producing one good $g \in \{1, \ldots, G\}$ or not enter at all. Setting up a firm requires a fixed cost F_g, paid in units of entrepreneurial talent. Once that cost is paid, the entrepreneur receives the blueprint b_g and a random draw of workplace amenities a_j from a good-specific distribution with strictly positive support and a finite mean $\bar{a}_g$. The role of workplace amenities will be explained below. Hiring and production decisions are done after the amenities draw is observed, as discussed in the next subsection.

I assume that there is a competitive market for entrepreneurial talent and that entrepreneurs can form coalitions to insure against idiosyncratic risk associated with the draw of firm amenities a_j. These assumptions allow me to abstract from the distribution of entrepreneurial talent and to pin down firm entry by equating expected profit and entry costs for each good g:

$$E_{a_j|g}\left[\pi_g(a_j)\right] = F_g p_T = F_g \quad \forall g$$

where $\pi_g(a_j)$, defined below, denotes profits achieved by a firm with amenities a_j producing good g. The second equality follows from assuming that entrepreneurial talent is the numeraire in this economy. This choice of numeraire is valid because firms have positive profits, as I will show below, and so the price of entrepreneurial talent cannot be zero. A positive price for entrepreneurial talent also implies that all of it is used up in equilibrium:

$$\sum_g J_g F_g = T \tag{3}$$

where J_g is total entry of firms producing good g. When there is a single good $g = 1$ in this economy, the number of firms is fixed at $J_1 = T/F_1$. But with multiple goods, the number of firms producing each good might respond to shocks.

The utility function of entrepreneurs, U^E, is a constant elasticity aggregator of consumption $Q_1, \ldots, Q_G$. Preferences of worker i of type h, captured by U_{hi}^L, depend on both consumption and the firm j

---

[6]This is my preferred interpretation because, in many contexts, changes in inequality happen within industries (see e.g. Card, Heining and Kline (2013) and Song et al. (2018)). Consistent with this interpretation, the next section shows that the estimated elasticity of substitution for the two goods in the quantitative exercise is large. The model can also be used for studying between-industry phenomena, such as between-industry sorting in Abowd et al. (2018).

where she is employed:

$$U^E\left(\{Q_g\}_{g=1}^{G}\right) = C\left(\{Q_g\}_{g=1}^{G}\right)$$

$$U_{hi}^L\left(\{Q_g\}_{g=1}^{G}, j\right) = C\left(\{Q_g\}_{g=1}^{G}\right)\left[a_j \exp\left(\eta_{ij}\right)\right]^{\frac{1}{\beta_h}}$$

$$\text{where } C\left(\{Q_g\}_{g=1}^{G}\right) = \left[\sum_{g=1}^{G} Q_g^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{1-\sigma}}$$

$$\text{and } \eta_{ij} \sim \text{Extreme Value Type I}$$

Firms matter to workers not only due to their overall level of amenities a_j, but also because of an idiosyncratic component η_{ij}. This component captures match-specific features such as distance to the workplace or personal relationships with the manager or other coworkers. The parameters β_h measure the importance of consumption relative to these non-pecuniary elements. Higher β_h implies that the market for labor of type h is more competitive. The details of the labor market are discussed in the next subsection.

Markets for goods are competitive. Thus, any equilibrium will feature prices p_g equal to the marginal cost of good g at all firms producing that good. There is a price index $P = \left[\sum_{g=1}^{G} p_g^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$ such that consumption level C costs $C \times P$. Because $C(\cdot)$ is homothetic, aggregate consumption is only a function of prices and aggregate income.

Continuing with the example from Section 2.4, the small local shop and the large online retailer are interpreted as differentiated varieties in the retail sector, with elasticity of substitution σ. In addition to task requirements, these firms might differ in entry costs and the average level of amenities. The online retailer might require substantial capital investment or managerial input to set up, justifying high entry costs F_g. If $\bar{a}_g$ is higher for the large retailers, then they are also more desirable workplaces on average.

3.2 Labor markets, the problem of the firm, and equilibrium

Labor markets are based on the model of Card et al. (2018), where firms compete monopsonistically for labor. Each worker is characterized by its type $h \in \{1, \dots, H\}$ and a quantity of efficiency units of labor ε. The distribution of efficiency units of labor across workers of type h is continuous with

15

density $r_h(\cdot)$ and support over the real line.[7] Throughout this section, it is important to distinguish between quantities of workers, denoted by n, and quantities of labor, denoted by l. Worker earnings are denoted by y, while prices for labor are denoted by w.

Labor regulations prevent firms from paying a total compensation of less than $\underline{y}$ to any worker. I refer to $\underline{y}$ as the minimum wage; the model has no variation in hours worked, so earnings and wages are interchangeable. Workers with low ε might have a marginal product of labor lesser than $\underline{y}$ at some firms, in which case hiring those workers would be unprofitable. Thus, I allow firms to reject workers with productivity below some minimum value $\underline{\varepsilon}_{hj}$, generating involuntary non-employment.

3.2.1 Firm-level labor supply and labor costs

There are separate labor markets for each worker group h. The timing of each of these labor markets is as follows:

1. Each firm j posts a price per labor efficiency unit w_{hj} and a rejection cutoff $\underline{\varepsilon}_{hj}$.

2. Workers observe all w_{hj} and $\underline{\varepsilon}_{hj}$. Based on that information, they choose firms that maximize their indirect utility. If no firm is chosen, the worker earns zero income.

3. Firms observe (h, ε) of workers who applied to them (but not idiosyncratic preference shifters η_{ij}) and hire those with $\varepsilon > \underline{\varepsilon}_{hj}$.

4. Production occurs and hired workers are paid $y = \max\{w_{hj}\varepsilon, \underline{y}\}$. Rejected workers, if any, earn zero income.

To study worker choices in step 2, consider the indirect utility of a worker i characterized by (h, ε), if this worker chooses firm j:

$$
V_{ih}(\varepsilon, j) = \underbrace{\frac{\mathbf{1}\{\varepsilon \geq \underline{\varepsilon}_{hj}\} \max\{\varepsilon w_{hj}, \underline{y}\}}{P}}_{\text{Consumption}} \left[a_j \exp(\eta_{ij})\right]^{\frac{1}{\beta_h}}
$$

$$
= \begin{cases} \dfrac{1}{P} \exp\left(\beta_h \log\left(\max\{\varepsilon w_{hj}, \underline{y}\}\right) + \log a_j + \eta_{ij}\right)^{\frac{1}{\beta_h}} & \text{if } \varepsilon \geq \underline{\varepsilon}_{hj} \\ 0 & \text{otherwise} \end{cases}
$$

---

[7] I employ LogNormal distributions of ε in the quantitative exercise. Counterfactual exercises require a parametric assumption for $r_h(\cdot)$, which is used to obtain the number of workers driven to unemployment because of the minimum wage and the distribution of ε in that unobserved population.

where the indicator function denotes that worker i earns positive income at firm j only if i's endowment of labor efficiency units is at least $\underline{\varepsilon}_{hj}$.

Because η_{ij} is drawn from a Type I Extreme Value distribution, the probability of a worker (h, ε) choosing a particular firm j is given by:

$$P\left(j = \underset{j' \in \{1,\dots,J\}}{\arg\max} V_{ih}(\varepsilon, j')\right) = \mathbf{1}\{\varepsilon > \underline{\varepsilon}_{hj}\} a_j \left(\frac{\max\{\varepsilon w_{hj}, \underline{y}\}}{\omega_h(\varepsilon)}\right)^{\beta_h}$$

$$\text{where } \omega_h(\varepsilon) = \left(\sum_{j'} \mathbf{1}\{\varepsilon > \underline{\varepsilon}_{hj'}\} a_{j'} \max\{\varepsilon w_{hj'}, \underline{y}\}^{\beta_h}\right)^{\frac{1}{\beta_h}}$$

The "inclusive value" $\omega_h(\varepsilon)$ is a measure of demand for skills in this model. A high value means that many firms are posting high wages for type h and willing to hire that particular ε, despite the minimum wage. That makes those workers harder to attract for any individual firm because they have good outside options at other firms. Mechanically, $\omega_h(\varepsilon)$ has an allocative role similar to that of wages in competitive models: it is a cost shifter that firms take as given and that ensures labor market clearing.

The number of workers choosing a particular firm and the resulting supply of labor are increasing functions of posted wages, conditional on rejection cutoffs:

$$n_h(w_{hj}, \underline{\varepsilon}_{hj}, a_j) = N_h a_j \int_{\underline{\varepsilon}_{hj}}^{\infty} \left(\frac{\max\{\varepsilon w_{hj}, \underline{y}\}}{\omega_h(\varepsilon)}\right)^{\beta_h} r_h(\varepsilon) d\varepsilon \tag{4}$$

$$l_h(w_{hj}, \underline{\varepsilon}_{hj}, a_j) = N_h a_j \int_{\underline{\varepsilon}_{hj}}^{\infty} \exp(\varepsilon) \left(\frac{\max\{\varepsilon w_{hj}, \underline{y}\}}{\omega_h(\varepsilon)}\right)^{\beta_h} r_h(\varepsilon) d\varepsilon \tag{5}$$

Finally, total labor costs are given by:

$$C_h(w_{hj}, \underline{\varepsilon}_{hj}, a_j) = N_h a_j \left[\int_{\underline{\varepsilon}_{hj}}^{\frac{\underline{y}}{w_{hj}}} \frac{\underline{y}^{\beta_h+1}}{\omega_h(\varepsilon)^{\beta_h}} r_h(\varepsilon) d\varepsilon + \int_{\frac{\underline{y}}{w_{hj}}}^{\infty} \frac{(\varepsilon w_{hj})^{\beta_h+1}}{\omega_h(\varepsilon)^{\beta_h}} r_h(\varepsilon) d\varepsilon\right] \tag{6}$$

In these expressions, I omit the dependency of $\omega_h(\varepsilon)$ on the own firm's posted wage w_{hj} because, with monopsonistic competition, $\omega_h(\varepsilon)$ is taken as given by firms.

3.2.2 Problem of the firm

Firms maximize profit by choosing posted wages and rejection cutoffs:

$$\pi_g(a_j) = \max_{\boldsymbol{w}_j, \boldsymbol{\epsilon}_j} p_g f\left(\boldsymbol{l}(\boldsymbol{w}_j, \boldsymbol{\epsilon}_j, a_j), b_g\right) - \sum_{h=1}^{H} C_h(w_{hj}, \boldsymbol{\epsilon}_{hj}, a_j)$$

The following Lemma shows that this problem has intuitive solutions and that the model admits a representative firm for each good:

Lemma 2. *The solution of the problem of the firm is interior and characterized by the following first order conditions:*

$$p_g f_h\left(\boldsymbol{l}(\boldsymbol{w}_j, \boldsymbol{\epsilon}_j, a_j), b_g\right) \frac{\beta_h}{\beta_h + 1} = w_{hj} \qquad\qquad h = 1, \ldots, H \qquad\qquad (7)$$

$$p_g f_h\left(\boldsymbol{l}(\boldsymbol{w}_j, \boldsymbol{\epsilon}_j, a_j), b_g\right) \underline{\epsilon}_{hj} = \underline{y} \qquad\qquad h = 1, \ldots, H \qquad\qquad (8)$$

Additionally, firms producing good g choose the same wages $\boldsymbol{w}_g$ and rejection criteria $\underline{\boldsymbol{\epsilon}}_g$. Output and employment are linear in firm amenities: $q_j = \frac{a_j}{\bar{a}_g} \bar{q}_g$ and $l_j = \frac{a_j}{\bar{a}_g} \bar{l}_g$, where $\bar{q}_g$ and $\bar{l}_g$ denotes mean output and mean labor demand for all firms producing good g, respectively.

The first order conditions represent trade-offs along two different margins: workers above the minimum wage and workers around the rejection threshold. To build intuition on the optimality condition on wages, denote by l_{hj}^+ the sum of efficiency units at firm j supplied by workers earning more than the minimum wage. A proportional increase in posted wages $d \log w_{hj}$ brings in $(\beta d \log w_{hj}) l_{hj}^+$ labor units, generating $(\beta d \log w_{hj}) l_{hj}^+ p_g f_h(\cdot)$ in additional revenues. Labor costs increase for two reasons. First, the firm pays $(\beta d \log w_{hj}) l_{hj}^+ w_{hj}$ for the additional labor purchased. Second, a higher wage increases the wage bill for current workers by $d \log w_{hj} l_{hj}^+ w_h$. Setting added revenues equal to additional costs yields Equation 7.

Equation 8 is the first order condition on the rejection cutoffs. A lower cutoff brings in additional workers with $\varepsilon = \underline{\epsilon}_{hj}$, each of which increases revenues by $p_g f_h \underline{\epsilon}_{hj}$. When firms chose thresholds optimally, that additional revenue equals the minimum wage $\underline{y}$, which is the cost of labor at that margin.

Figure 3 illustrates how workers are divided in three groups according to their level of efficiency units. Those to the left of $\underline{\epsilon}_{hj}$ are rejected. Those with $\varepsilon > \underline{y}/w_{hj}$ earn the wage posted by the firm times their quantity of labor units. Finally, those in the intermediate range earn the minimum

Figure 3: Choice of rejection criterion and bunching at the minimum wage
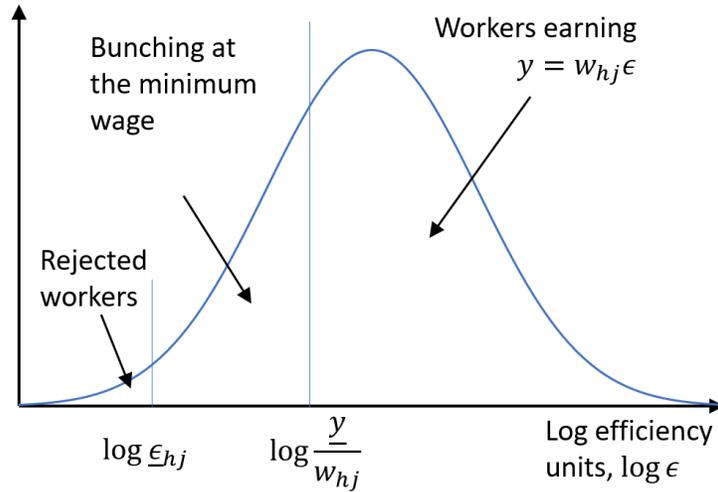
Notes: This figure shows thresholds in the distribution of efficiency units of labor ε that determine whether worker are rejected by firm j, are employed receiving the minimum wage, or employed receiving the posted wage times the number of efficiency units. The horizontal axis is in log scale. The blue line shows the distribution of efficiency units, which is LogNormal in this illustration (as well as in the quantitative exercise). When there is a single good in the economy, such that there is a representative firm, the distribution of log wages for workers of type h is a truncated normal with a peak at the minimum wage. The mass of this peak is given by the area between the two vertical lines in this graph.

wage. The first order conditions imply that these two thresholds are proportional to each other in all firms choosing labor inputs optimally, with their ratio being given by $1 + 1/\beta_h$. Log wage histograms simulated from the model have peaks at the minimum wage corresponding to the mass of workers between the two vertical lines. Bunching at the minimum wage is often observed in the data (DiNardo, Fortin and Lemieux, 1996; Harasztosi and Lindner, 2018) but is not a common feature in models of wage inequality.

Lemma 2 also shows that firms producing the same good are equal in wages and input intensities. Dispersion in amenities within good only scales the firm up or down. This result simplifies the analysis of between-firm wage differentials and sorting in this model by restricting the sources of these patterns to differences in blueprints, entry costs, or mean amenities $\bar{a}_g$. It also simplifies the

expression for $\omega(\varepsilon)$, making the computation of labor demands feasible:

$$\omega_h(\varepsilon) = \left(\sum_g J_g \mathbf{1}\left\{ \varepsilon > \underline{\varepsilon}_{hg} \right\} \bar{a}_g \max\left\{ \varepsilon w_{hg}, \underline{y} \right\}^{\beta_h} \right)^{\frac{1}{\beta_h}} \tag{9}$$

3.2.3 Equilibrium

An equilibrium of this model is a set of prices $\{p_g\}_{g=1}^G$, aggregate consumption $\{Q_g\}_{g=1}^G$, firm entry $\{J_g\}_{g=1}^G$, and choices by representative firms $\{w_g, \varepsilon_g\}_{g=1}^G$ such that:

1. Markets for goods clear:

$$Q_g = \left[\frac{p_g}{P}\right]^{-\sigma} \frac{I}{P} = J_g \bar{q}_g \quad \forall g \tag{10}$$

$$\text{where } I = T + \sum_{g=1}^G J_g \sum_{h=1}^H C_h(w_{hg}, \underline{\varepsilon}_{hg}, \bar{a}_g)$$

2. For all g, firm choices solve the set of equations (7) and (8).

3. Entrepreneurs have zero ex-ante expected profits:

$$E_{a_j|g}\left[\pi_g(a_j)\right] = p_g f\left(l(w_g, \varepsilon_g, \bar{a}_g), b_g\right) - \sum_{h=1}^H C_h(w_{hg}, \underline{\varepsilon}_{hg}, \bar{a}_g) = F_g \quad \forall g \tag{11}$$

4. The market for entrepreneurial talent clears (Equation 3).

Labor market clearing is implied by the definition of $\omega_h(\varepsilon)$, which ensures that the number of job applicants to all firms (calculated using Equation 4) is equal to total number of workers N_h.

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 where the choice variables are firm-specific task thresholds, firm-level output, and prices for each good. The procedure below describes how to calculate that system of equations:

1. Start with values for mean output $\bar{q}_g$ and task thresholds $\bar{x}_g = \{\bar{x}_{1g}, \dots, \bar{x}_{Hg}\}$ 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 2 in the text, or Equation 15 in Appendix B if using the

exponential-Gamma parametrization).

3. Find marginal products of labor f_{hg} via the non-arbitrage conditions (1) 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. Obtain $\omega_h(\varepsilon)$ using expression 9.

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

   (a) The deviation between $\bar{l}_{hg}$ found in step 2 and that implied by the labor supply curve (5).

   (b) The deviation between profits and entry costs in Equation 11.

That system of equations can be solved using standard numerical procedures, with the restrictions that $\bar{q}_g > 0$, $p_g > 0$, and $0 \le \bar{x}_{1g} \le \bar{x}_{2g} \le \cdots \le \bar{x}_{Hg}$ $\forall g$. 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 $h = 2, \ldots, H-1$.

3.3 Determinants of the wage distribution

The key outcome of the analysis is how wages differ between groups, within groups, and across firms. We know from the labor market structure that log earnings of a worker i of type h at a firm producing good g take the form $\log y_{ihg} = \max\{\log w_{hg} + \log \varepsilon_i, \log \underline{y}\}$. Variation in wages between worker groups is driven by differences in w_{hg}. Within-group variation of log wages has three components: the dispersion of efficiency units, differences in mean log wages across goods for the same worker type, and censoring by the minimum wage.

The following proposition provides intuition about how wages vary across firms producing different goods:

Proposition 2. *1. If $b_g(x) = b(x)/z_g$ and $\frac{F_g}{\bar{a}_g}$ is common across goods, then there are no firm-level wage premia:*

$$\log y_{ihg} = \max\left\{\lambda_h + \log \varepsilon_i, \log \underline{y}\right\}$$

for scalars $\lambda_1, \ldots, \lambda_H$.

2. *If there is no minimum wage ($\underline{y} = 0$), $\beta_h = \beta$, and $b_g(x) = b(x)/z_g$, then wages are log additive in worker type and firm type:*

$$\log y_{ihg} = \lambda_h + \frac{1}{1+\beta} \log \left(\frac{F_g}{\bar{a}_g} \right) + \log \varepsilon_i$$

3. *If there is no minimum wage, $\beta_h = \beta$, 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:*

$$y_{ih'g'}/y_{ihg'} > y_{ih'g}/y_{ihg}$$

The first part of Proposition 2 shows that wage dispersion for similar workers exists only if there are differences in the shapes of blueprints (such that firms differ in skill intensity) or in the ratio of entry costs to mean amenities. Notably, differences in physical productivity across goods (denoted by z_g above) are not enough to generate wage differentials between firms. The reason is that, if the ratio of entry costs to firm amenities is the same, differences in physical productivity lead to additional entry and reduced marginal utility of consumption for the good with more productivity, up to the point where marginal revenue product of labor is equalized across firms.

The second part highlights the role of entry costs in generating wage differences across firms. The zero profits condition implies that firms producing goods with higher entry costs need to operate at larger scale. To hire more workers, these firms need to post higher wages, unless the differences in entry costs are exactly offset by differences in mean amenities.

The third part of Proposition 2 shows how heterogeneity in skill intensity generates differential wage gaps across firms. Firms using some factors more intensively than others must pay a relative premium to that factor. Thus, in general, the model cannot generate log-additive wages as in Abowd, Kramarz and Margolis (1999), except when factor intensities do not vary. Equal skill intensities are ensured by the conditions imposed in (2).

The inability of this model to generate log-additive wages and sorting simultaneously echoes some results in the literature on labor market sorting, such as those in Eeckhout and Kircher (2011). But it is possible that skill-intensive firms pay a positive wage premium for all worker types if those firms have high entry costs relative to amenities. The quantitative exercise shows that this flexibility is necessary for fitting the data.

To provide a concrete example of how firms differ in equilibrium, consider the Exponential-Gamma parametrization introduced in Subsection 2.4. Under that parametrization, goods are fully described by five scalars: blueprint complexity θ_g, blueprint shape k_g, blueprint productivity z_g, mean amenities $\bar{a}_g$, and the ratio of entry costs to mean amenities $F_g/\bar{a}_g$. These scalars map directly into five empirical measures for the set of firms producing that good, respectively: mean worker education, dispersion in worker education, share of total workforce employed by those firms, mean firm size, and firm-level wage premia.

The next step in the analysis is understanding how the wage distribution changes over time, given shocks to labor supply, labor demand, and minimum wages. As a starting point, the following proposition considers a case in which the supply of skill, demand for task complexity, and minimum wages rise in tandem:

Proposition 3 (Race between technology, education, and minimum wages). *Start with a baseline economy characterized by parameters $\left(\{e_h, N_h, \beta_h\}_{h=1}^{H}, \{b_g, F_g, \bar{a}_g\}_{g=1}^{G}, T, \sigma, \underline{y} \right)$ and 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. *$b_g'(x) = \frac{1}{(1+\Delta_1)(1+\Delta_2)} b_g \left(\frac{x}{1+\Delta_1} \right) \ \forall g$: Production requires fewer tasks, but the composition of tasks moves towards increased complexity.*

4. *$\underline{y}' = \underline{y}$: The minimum wage stays constant relative to the price of entrepreneurial talent.*

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 $P' = P/(1+\Delta_2)$.[8]

Proposition 3 delineates technological progress in this economy. Production becomes more efficient by using tasks that are more complex. At the same time, the skill of workers increase, changing the set of tasks where skill differences are relevant. If minimum wages remain as impor-

---

[8]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 $z_g' = (1+\Delta_2)z_g$.

tant, then there is a uniform increase in living standards. Wage differences between worker groups and between firms for the same group remain stable.

If these transformations are not balanced, then relative prices and the allocation of resources might change. The overall effect on the wage distribution and sorting is difficult to study analytically because they interact through four different channels: (i) changes in the economy-wide measures of skill scarcity, $\omega_h(\cdot)$; (ii) changes in relative consumption, which are tantamount to changes in the distribution of firm types; (ii) changes in the employment composition of each firm, conditional on type; and (iv) changes in firm-specific wage premia. Since all of these channels are potentially important, the best way to disentangle the role of each shock is through a quantitative application of the model.

It is possible, however, to obtain some intuition about how firm heterogeneity might amplify or attenuate the impact of specific shocks on the wage distribution, relative to a framework with a representative firm. Shocks that affect the price for skills will have differential effects on the cost of goods that are produced using different sets of tasks. Those changes in cost are passed through to consumers. As they substitute towards cheaper goods, the aggregate set of tasks being produced by this economy shifts towards more or less complex tasks. That shift acts as a secondary demand shock, leading to further adjustments in the price for skills.

The following proposition isolates the effect of that secondary shock by considering what happens when there is a change in cost for a particular firm, in a simplified version of the model:

Proposition 4 (Changes in relative output affects the returns to skill). *Consider a competitive version of this economy ($\beta_h = \infty$, $F_g = 0$) with two goods ($G = 2$), no minimum wages ($\underline{y} = -\infty$), and where relative labor demand Q_2/Q_1 is an exogenous parameter rather than the outcome of consumer optimization. Assume good $g = 2$ is relatively more intensive in high-complexity tasks, such that $b_2(x)/b_1(x)$ is increasing. Then, an increase in Q_2/Q_1 causes increases in wage gaps w_{h+1}/w_h.*

The full effect of a shock to labor supply, technical change, or minimum wages combines a direct effect and this secondary demand effect. Consider for example a minimum wage hike. As discussed in the presentation of the task-based production function, minimum wages decrease all wage gaps in this economy when there is a representative firm. But with two firms, minimum wages increase the cost of the low-skill good relative to the high-skill good. The secondary demand shock is in the direction of increased wage inequality, as in Proposition 4 above. The overall effect can be either an attenuated decline in wage gaps or wage polarization, whereby wages increase for low and high worker groups relative to intermediate groups. The next section of this paper shows that

24

this channel can be quantitatively important and help fit reduced-form estimates of the effects of minimum wage shocks.

Proposition 4 is conceptually related to papers where structural shocks change the composition of jobs in the economy, causing additional effects on the wage distribution through this channel. Examples of such papers are Kremer and Maskin (1996), Acemoglu (1999), and Mak and Siow (2018), which study supply shocks; Acemoglu and Restrepo (2018), which studies automation; Acemoglu (2001), which studies minimum wages; and Sampson (2014) and Davis and Harrigan (2011), which study trade liberalization.

4   Quantitative exercise: wage inequality and sorting in Rio Grande do Sul, Brazil

While the literature covering countries such as the US and Germany often tries to explain increasing wage disparities in recent decades, the academic debate in Brazil attempts to rationalize a decline in inequality starting in the 1990s. The most salient facts in the Brazilian context are significant increases in both the minimum wage and educational achievement, following policies aimed at universal primary schooling in the 1980s and 1990s and expansion in access to college-level education in the 2000s. In this section, I study the Brazilian context as a proof of concept for the model. In the first step of the analysis, I take a sparsely parameterized, over-identified version of the model to the data and study whether it can rationalize changes from 1998 through 2012. In the second step, I use the estimated model to generate counterfactuals that isolate the individual impact of supply, demand, and minimum wages.

Throughout the analysis, I restrict my attention to the formal sector in the southernmost state of Brazil, Rio Grande do Sul. Similarly to many other developing countries, a substantial share of the Brazilian workforce is informal (employed at firms that evade regulations such as payroll taxes and minimum wages). It would be ideal to include formal and informal workers in the analysis because the model has general equilibrium effects. That is impossible, however, for data limitations. Households surveys measure employment and wages in the informal sector, but I require matched employer-employee data to gauge between-firm wage dispersion for similar workers and labor market assortativeness.

To partially address problems related to informality, I restrict my analysis to Rio Grande do Sul, the largest state in Southern Brazil. Except for the Southeast, the Brazilian South is the region with

the lowest rates of employment informality in the country. The Southeast is less interesting for this exercise, however, because higher wages in that region make the minimum wage shock less relevant.[9] I discuss the implications of ignoring the informal sector before presenting the results of the counterfactual exercises. In addition, Appendix C.2 contains a thorough analysis of inequality and education patterns using a different data source that includes the informal sector.

The data source used in this section is the RAIS (*Relação Anual de Informações Sociais*), a confidential matched employer-employee dataset administered by the Brazilian Ministry of Labor. Firms are mandated by law to report to RAIS, and in doing so provide information about their employees. The dataset I utilized contains information about both the firm (including legal status, economic sector, and the municipality in which it is registered) and each worker it formally employs (including education, age, earnings in December, contract hours, and hiring and separation dates).

Because I am interested in equilibrium effects, the sample I use has few restrictions. I select adults of both genders between 18 and 54 years of age, who are not currently in school, and who are working in December having been hired in November or earlier. I only consider one job per worker per year. The resulting data set has 1,494,186 workers and 148,203 firms in 1998, and 2,398,391 workers and 238,545 firms in 2012. For each worker, I calculate the hourly wage based on their monthly earnings and contract hours, before winsorizing the bottom and top percentiles of the wage distribution. Summary statistics are provided in Table 6, located in Appendix C.1.

4.1 Target moments

Figure 4 demonstrates the evolution of wages in the *Rio-Grandense* economy. The top left panel shows that, from 1998 to 2012, real wages have increased for all deciles of the log wage distribution, and particularly so for the lowest deciles. Almost all commonly used measures show a reduction in inequality: upper-tail or lower-tail percentile gaps (top-right panel), differences in mean log wage between workers with secondary education (that is, those complete high-school and college dropouts) and less educated workers, and the variance of log wages — for the sample as a whole and within each educational group. The single exception is the gap between secondary and tertiary education (workers that had completed college and beyond), which rose until 2006 and subsequently remained stable through to the end of the period studied. In Appendix C.2, I show that wage inequality trends are similar in a different data set that includes informal workers.

---

[9]I use a single state in the South because the estimator of variance components of Kline, Saggio and Sølvsten (2018) performs better in well-connected labor markets (in terms of worker transitions between firms). Table 5 shows that the sample size is large enough to generate precise estimates.

Figure 4: Measures of wage dispersion in Rio Grande do Sul, Brazil
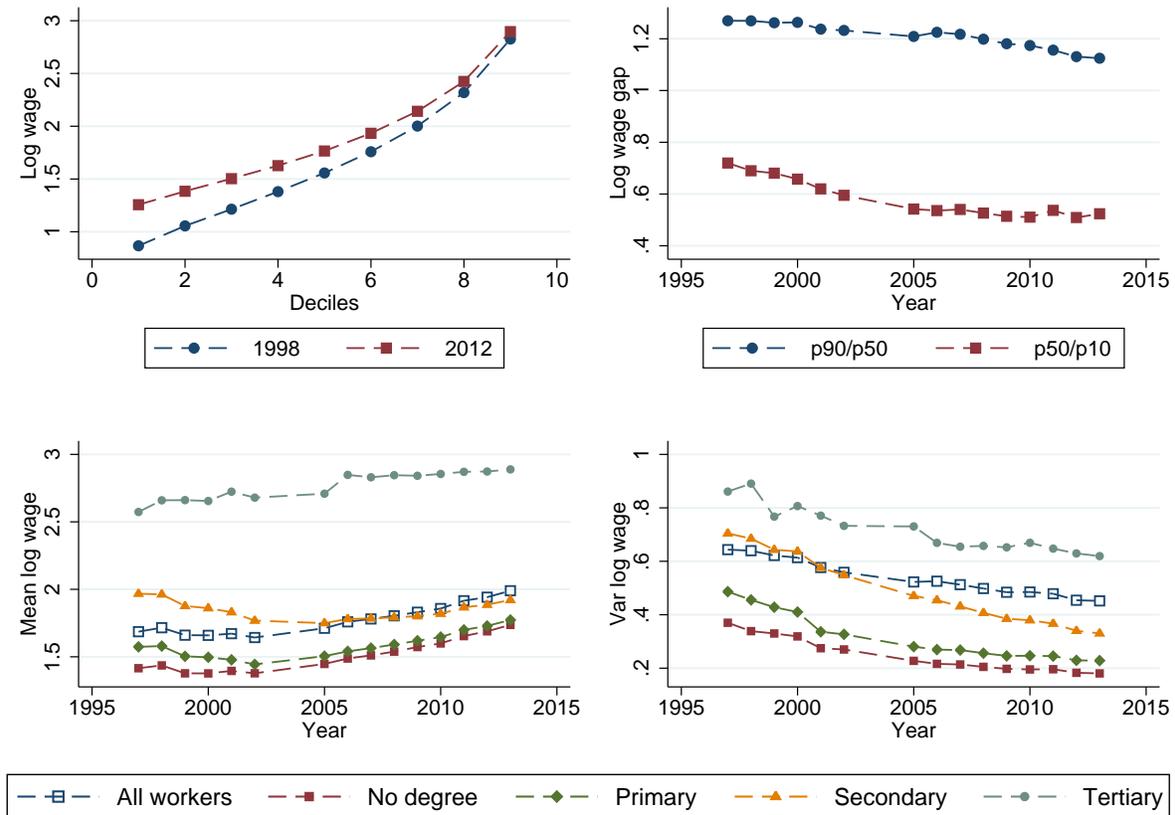
Notes: RAIS data for the formal sector in Rio Grande do Sul, Brazil. The top left graph shows deciles of the log wage distribution in 1998 and 2012. The top right graph shows the evolution of the 90 to 50 and 50 to 10 percentile gaps from 1997 through 2013. The bottom left and bottom right graphs show means and variances of log wages, respectively, for the whole sample and each educational group. Data for 2003 and 2004 are not available.

Figure 5: Changes in educational achievement and minimum wages
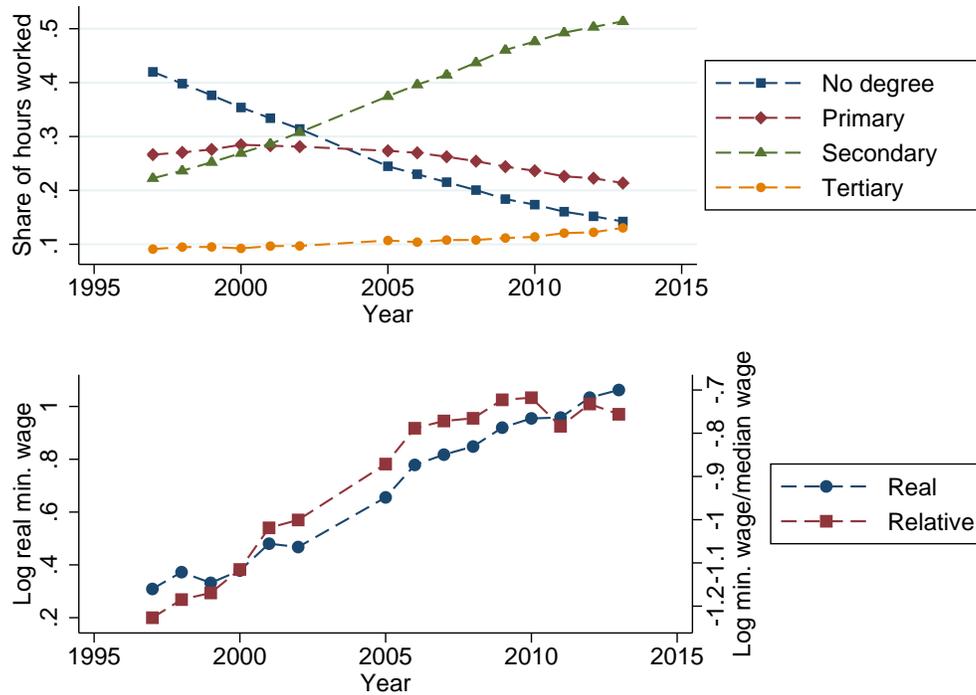
Notes: RAIS data for the formal sector in Rio Grande do Sul, Brazil. The top graph shows, for each year from 1997 through 2012, the share of hours worked by employees in each educational group. The bottom graph shows the evolution of minimum wages in the same years, both in real terms and relative to the median wage in that year. Data for 2003 and 2004 are not available.

The estimation procedure will target levels from 1998 and changes from 1998-2012 in between-group mean log wage gaps and in the variance of log wages within each group.

The literature studying wage inequality in Brazil highlights two candidate explanations for these patterns: increased educational achievement and minimum wages. Figure 5 shows that both factors are relevant in Rio Grande do Sul. The first graph displays the fraction of hours worked by employees in each educational group. The pattern is striking: workers with less than a complete primary education (that is, less than eight years of schooling) supply 40 percent of the hours in 1998, but only around 15 percent in 2012. On the other hand, the group with a complete secondary education (high school and college dropouts) increased its participation level by almost 30 percentage points. Moreover there is a substantial increase in college completion in relative terms (from 9.4 percent to 12.2 percent), though they remain a fraction of the formal workforce.

In a strict sense, these are not changes in the supply of labor, but are instead the observed employ-

ment shares for each educational group. Even with the exogenous labor supply in the model, these two concepts differ because the minimum wage creates involuntary non-employment. Figure 11 in Appendix C.2 shows similar trends in the share of all adults belonging to each of these educational groups, regardless of whether they participate in the labor force or not. That fact indicates that the source of changes in schooling achievement of the workforce are changes in the education levels for the whole population, not changes in selection patterns into employment.

The bottom graph in Figure 5 shows the large and steady increase in the national minimum wage in Brazil. The same figure shows that the minimum wage increased much faster than median wages in Rio Grande do Sul until 2006. The estimation procedure will target levels and changes in five measures that capture the degree to which the minimum wage is binding: the shares of workers in each educational group earning up to the minimum wage plus 25 log points, along with the minimum wage relative to mean log wages.

In addition, I will employ reduced form estimates of minimum wage spillovers as additional targets in estimation. The objective is to impose discipline on substitution patterns for the estimated model, adding credibility to counterfactual exercises.

I use the methodology developed by Autor, Manning and Smith (2016) to estimate the following equation using data for all Brazilian states in the period studied:

$$
\begin{aligned}
\log y_{st}(p) - \log y_{st}(50) =& \beta_1(p) \left[\log \underline{y}_t - \log y_{st}(50) \right] + \beta_2(p) \left[\log \underline{y}_t - \log y_{st}(50) \right]^2 \\
& + \zeta_{0s}(p) + \zeta_{1s}(p) \times time_t + \zeta_2(p) \times (time_t)^2 + u_{st}(p) \qquad (12)
\end{aligned}
$$

where $y_{st}(p)$ is the p-eth percentile of the real wage distribution in state s at time t; $\underline{y}_t$ is the national minimum wage at time t; $\zeta_{0s}(p)$ and $\zeta_{1s}(p)$ are state-quantile fixed effects and linear trends, respectively; $\zeta_2(p)$ is a national quadratic trend; and $u_{st}(p)$ is the residual.

This expression parameterizes the impact of the "effective minimum wage" $\underline{y}_t - \log y_{st}(50)$ — the minimum wage relative to the median wage in any given state and year — on any quantile p of the wage distribution, again relative to the median. The quadratic specification accounts for possibly non-linear effects of the effective minimum wage. The regression includes state-percentile fixed effects and linear trends to account for state-level changes in the shape of the wage distribution that are unrelated to the minimum wage. It also includes a national quadratic trend for each percentile, accounting for flexible changes in the shape of the wage distribution that are common across states. I use this trend instead of year effects because the statutory minimum wage is set at the federal level in Brazil.

Table 1: Reduced form estimates of minimum wage spillovers

Quantile	Levels		Differences	
	OLS	IV	OLS	IV
10	0.584	0.427	0.641	0.540
	(0.062)	(0.068)	(0.050)	(0.052)
20	0.369	0.246	0.389	0.321
	(0.049)	(0.043)	(0.036)	(0.029)
30	0.204	0.158	0.241	0.167
	(0.073)	(0.054)	(0.034)	(0.022)
40	0.106	0.025	0.119	0.052
	(0.032)	(0.031)	(0.029)	(0.023)
60	-0.051	0.044	-0.084	-0.019
	(0.037)	(0.029)	(0.041)	(0.024)
70	0.091	0.259	-0.037	0.067
	(0.095)	(0.060)	(0.059)	(0.053)
80	0.113	0.281	0.015	0.028
	(0.108)	(0.088)	(0.079)	(0.063)
90	0.230	0.282	0.113	-0.011
	(0.085)	(0.093)	(0.073)	(0.068)
N	378	378	351	351
Cragg-Donald F		11.50		43.24

Notes: Each cell in this table corresponds to the marginal effects of the "effective minimum wage" (log statutory minimum wage minus median log wage) on quantiles of the wage distribution relative to the median log wage, coming from separate (quantile-specific) regressions. Each observation is a state-year and the regression is weighted by total hours worked. All years from 1996 through 2013 are included except 2002, 2003, 2004 and 2010, years in which data is not available for some states. Marginal effects are calculated at the median log wage for the whole sample (hours weighted). Regressions in levels include state fixed effects, state linear trends, and a national quadratic trend. Regressions in differences include state fixed effects and a national linear trend. Standard errors are clustered by state (27 clusters).

Autor, Manning and Smith (2016) argue that the effective minimum might correlate with the residual term because median wages are used to construct both the independent and the dependent variables. I follow their approach to solve this problem. Specifically, I use an instrument set composed of the log real minimum wage, the square of the log real minimum wage, and an interaction of the log real minimum wage with the average median real wage in state s for the whole period.

Table 1 shows ordinary least squares and instrumental variables estimates of the marginal effect of minimum wages over different quantiles of the wage distribution. I estimate specifications in levels and in differences. The specification in differences presents much stronger first stages (measured by the Cragg-Donald (1993) F statistic). In addition, it shows no spillovers in the upper tail, a criterion that has been used for model selection when studying the impact of minimum wages on the wage distribution (e.g. Autor, Katz and Kearney (2008) and Cengiz et al. (2018)). For these reasons, it is my preferred specification.

The estimates show spillovers that are economically and statistically significant up to percentile 40. Spillovers on the upper tail are small and indistinguishable from zero. These estimates are larger than what Autor, Manning and Smith (2016) found for the US, consistent with the fact that the minimum wage is more binding in Brazil and that only a small fraction of the workforce is in possession of a tertiary education.

Finally, I use the panel structure of the matched employer-employee data to gauge the degree of wage differentials across firms for similar workers. I begin with a log-additive specification for the wage of worker i at time t:

$$\log y_{it} = v_i + \psi_{J(i,t)} + \delta_t + u_{it}$$

where v_i is worker i's fixed effect, ψ_j is firm j's fixed effect, $J(i,t)$ represents the firm employing worker i at time t, δ_t is a time effect, and u_{it} is a residual that is uncorrelated with all fixed effects. I am primarily interested in the following decomposition of the variance of log wages:

$$\text{Var}(\log y_{it}) = \text{Var}(v_i) + \text{Var}(\psi_{J(i,t)}) + 2\text{Cov}(v_i, \psi_{J(i,t)}) + \text{Var}(\delta_t) + 2\text{Cov}(v_i + \psi_{J(i,t)}, \delta_t) + \text{Var}(u_{it})$$
$$(13)$$

This decomposition has been used to quantify the relevance of firm-level wage premia for similar workers and the degree of labor market sorting (Abowd, Kramarz and Margolis, 1999). If the log wage above is interpreted as a structural economic model, a positive covariance term means that high wage workers are matched to high wage firms, increasing the total variance of log wages. On the other hand, if wages are not log-additive (as in my model), it is unclear what this decomposition uncovers (Eeckhout and Kircher, 2011).

Table 2: Variance decomposition from two-way fixed effects model

Component	1998		2012		Change	
Var $(\log y_{it})$	0.675	100%	0.544	100%	-0.131	100%
Var (v_i) (worker effects)	0.391 (0.003)	57.9%	0.324 (0.001)	59.5%	-0.067 (0.003)	51.1%
Var $(\psi_{J(i,t)})$ (firm effects)	0.139 (0.002)	20.6%	0.071 (0.001)	13.1%	-0.067 (0.003)	51.5%
2Cov $(v_i, \psi_{J(i,t)})$	0.073 (0.003)	10.8%	0.103 (0.001)	18.9%	0.030 (0.003)	-22.6%
Other terms	0.072	10.7%	0.046	8.5%	-0.026	20.0%
$Corr(v_i, \psi_{J(i,t)})$	0.157		0.337		0.180	

Notes: This table shows the variance decomposition described in Equation (13) obtained using the Leave-out estimator of Kline, Saggio and Sølvsten (2018). Numbers in parentheses are asymptotic standard errors. The estimation labeled 1998 uses data for two years, 1997 and 1999. Similarly, the 2012 estimation uses data for 2011 and 2013. The decomposition includes both workers who move between firms and stayers. Each worker-year observation has the same weight. See Appendix C.3 for details and sample sizes.

However, even when the log-additive regression cannot be interpreted as a structural economic model, the variance decomposition may still provide information about the structural parameters governing imperfect competition and sorting. Thus, I use two elements of the variance decomposition as targets in the estimation procedure: the share of the variance of log wages accounted for by firm effects, estimates of the share of the variance of log wages accounted for by firm fixed effects, Var $(\psi_{J(i,t)})$ /Var (y_{it}); and the correlation between worker effects and firm effects.

Estimating the variance decomposition (13) is not a trivial task. Andrews et al. (2008) show that a simple "plug-in" estimator of the covariance composition using estimates of the fixed effects from ordinary least squares (OLS) regressions is biased. These authors provide a correction method that assumes the homoskedastic residuals. More recently, Kline, Saggio and Sølvsten (2018) show that heteroskedasticity causes bias and proposed a leave-out estimator that corrects for it. These authors also discuss how to conduct inference on the variance decomposition terms. I use the latter estimator (henceforth denoted by KSS) because the variance of residuals vary systematically across worker groups both in the data and in the model. Appendix C.3 provides details about the procedure.

Table 2 shows the KSS estimates of variance components. The variance of both firm and worker

effects decline over time, helping to explain the fall of wage inequality. There is also a sizable and statistically significant increase in the covariance of worker effects and firm effects. As a result, the correlation between worker effects and firm effects increases substantially.

To summarize, the estimation exercise attempts to match a total of 36 moments that provide a broad picture of inequality trends, including between-firm wage dispersion for similar workers, and constraints on the impact of minimum wages on the wage distribution.

4.2 Fitting the model

4.2.1 Parameters

I use a simple, parsimonious version of the model to fit these moments. I employ the exponential-Gamma parametrization of comparative advantage functions and blueprints described in Section 2.4. The complete list of estimated parameters is presented in Table 4.

The model has four worker types, assumed to be observable and linked to the four educational groups. I assume that the dispersion of efficiency units within each group has the same variance, $S_h^2 = S^2$. This assumption means that the model must fit differences in within-group wage dispersion based solely on minimum wages and wage dispersion between firms. In addition to S^2, I estimate comparative advantage α_h (except for normalizations $\alpha_1 = -1$, $\alpha_4 = 0$) and the relative taste for consumption β_h, which is also the elasticity of job applicants to the firm. I calibrate the elasticity of the higher type to 4, corresponding to a wage mark-down of 20%. That number is used as a reasonable benchmark in Card et al. (2018), based on their own literature reviews and those in Manning (2011).[10]

The estimated model has only two goods (and, equivalently, two firm types), making it parsimonious and simplifying its interpretation. I make further restrictions on parameters related to goods. The shape parameter of blueprints is assumed to be constant over time and common for

---

[10]I calibrate one of the β_h because, without that normalization, their levels are weakly identified. The main source of identifying variation for the β_h are levels and changes in within-group variance of log wages, because in the model these parameters affect wage differentials between firms. However, the parameter set being estimated also includes the difference in entry costs between the two firm types, which also affect between-firm wage differentials for similar workers in the model. Under correct specification, all β_h can still be identified jointly with the entry cost gap because they imply the size of bunching at the minimum wage. Figure 6, however, shows that the distribution of log wages is not perfectly specified and that the model cannot exactly match the observed bunching. Thus, using a calibrated value for either one of the β_h or the entry cost gaps can lead to more reliable estimates. I chose β_4 for that purpose because, among Brazilian workers, those with college are probably the most similar to workers in developed countries where the estimates cited in Manning (2011) were obtained.

Table 3: Target moments and model fit

Target	Data		Model	
	1998	Change	1998	Change
Group 1: Wage inequality (RMSE = 0.025)				
Mean log wage gaps:				
Primary / No degree	0.144	-0.104	0.201	-0.085
Secondary / Primary	0.383	-0.226	0.416	-0.259
Tertiary / Secondary	0.695	0.292	0.739	0.302
Within-group variance of log wages:				
No degree	0.338	-0.156	0.328	-0.147
Primary	0.456	-0.226	0.463	-0.209
Secondary	0.685	-0.345	0.670	-0.354
Tertiary	0.891	-0.261	0.878	-0.271
Group 2: The role of firms (RMSE = 0.028)				
Variance decomposition from two-way fixed effects model				
Share of firm effects	0.206	-0.075	0.179	-0.079
Corr. worker and firm effects	0.157	0.180	0.110	0.177
Group 3: Minimum wages (RMSE = 0.065)				
Log minimum wage relative to mean log wage				
All workers	-1.342	0.435	-1.315	0.417
Share of workers with $\log y \leq \log \underline{y} + 0.25$				
No degree	0.055	0.092	0.101	0.166
Primary	0.044	0.112	0.084	0.144
Secondary	0.028	0.098	0.048	0.117
Tertiary	0.006	0.014	0.021	0.001
Minimum wage spillovers:				
p10		0.540		0.440
p20		0.322		0.226
p30		0.167		0.130
p40		0.052		0.050
p60		-0.019		-0.033
p70		0.067		-0.063
p80		0.028		-0.089
p90		-0.011		-0.118

Notes: This table shows the 36 moments targeted by the estimation procedure, along with their model-based equivalents predicted by the estimated parameters. RMSE means the root mean squared error for all of the moments in each group.

Table 4: Estimated parameters

Parameters				
General:				
σ (substitution between goods)	*2*			
S^2 (variance of efficiency units)	0.447			
$\underline{y}$, 1998 (log min. wage)	0.161			
$\underline{y}$, 2012	0.493			
Worker type-specific:	No Degree	Primary	Secondary	Tertiary
α_h (comparative advantage)	*-1*	-0.595	-0.314	*0*
β_h (taste for consumption/elast. of worker supply)	60.55	7.65	5.16	*4*
Goods:	$g = 1$	$g = 2$		
θ_g, 1998 (initial blueprint complexity)	0.378	1.195		
k_g (blueprint shape)	116.18			
$\log z_g$, 1998 (initial blueprint productivity)	-2.192	*0*		
$\log F_g/\bar{a}_g$, 1998 (entry cost to mean amenities)	-6.653	*0*		
Demand shock:				
$d\log\theta$ (skill-biased technical change)	1.151			
$d\log z_2/z_1$ (change in productivity gap)	-1.901			
$d\log((F_2/\bar{a}_2)/(F_1/\bar{a}_1))$ (change in entry cost gap)	-4.197			

Notes: This table shows the 16 parameters being estimated in the quantitative exercise. Numbers that are greyed-out and italicized are normalizations or calibrations.

both goods. In addition, because I am not interested in predictions about firm size or the number of firms, I restrict my attention to the ratio of entry costs to mean amenities — more specifically, the log difference of this ratio between firm types. Finally, the elasticity of substitution between goods is calibrated to $\sigma = 2$. This parameter is technically identified from minimum wage spillovers; specifically, from the secondary impact of that shock on wages discussed in Proposition 4. In this particular case, however, that effect is small relative to the imprecision in the estimated minimum wage spillover (see Figure 7). Thus, that parameter is weakly identified in practice, and calibration is a more transparent approach.[11]

I now introduce the definition of the demand shock as a combination of three components. First, following the tradition in this literature, I allow for a skill-biased technical shock that increases productivity gaps between worker groups. In the task-based production function, this can be mod-

---

[11]I assume that goods are net substitutes because the model-based curves in Figure 7 suggest that the higher the substitution, the closer the predicted spillovers in the upper tail are to reduced-form estimates. Still, I chose a relatively small value for the elasticity of substitution as a way of being conservative about the extent of changes in the composition of firm types following structural shocks.

eled as a proportional increase in task complexities $d \log \theta$, as shown in Proposition 3. The second component is a possible change in the relative productivity of firm types, $d \log z_2/z_1$. This component affects the overall employment share of each firm type. The third component is a change in the relative ratio of entry costs to amenities, $d \log((F_2/\bar{a}_2)/(F_2/\bar{a}_2))$. That relative ratio is a key determinant of cross-firm wage dispersion for similar workers.

The last parameters in the model are minimum wages relative to the price of entry inputs (the numeraire in the model). Minimum wages are added as a free parameter procedure because prices for goods, entrepreneurial talent, and labor are unobserved, making a direct calibration of that parameter impossible.

4.2.2 Estimation procedure

The model is estimated by minimizing the least squares distance between the observed moments in Table 3 and the model equivalent of these moments. Each of the three groups of moments have the same importance in estimation. To evaluate the loss function, I need a function that maps the vectors of parameters to the vectors of moments predicted by two equilibria of the model, one corresponding to 1998 and another corresponding to 2012. To find these equilibria given a set of candidate models, I use a slightly modified set of equilibrium conditions: instead of imposing that the number of job applicants equals the total number of workers of each type (which is unobserved in the data), I impose instead that the number of employed workers in the model equals observed employment.[12]

The first group of moments in Table 3, along with the measures of how binding the minimum wage is in the third group, are calculated from the model-implied wage distributions for each worker type. Minimum wage spillovers are calculated as follows. First, I take each of the estimated equilibria and calculate a counterfactual where the log minimum wage increases by 10^{-6} relative to the entry input. Next, consistent with the reduced form specification (Equation 12), I calculate the changes in each quantile p over the change in the minimum wage, both of which are relative to the median wage: $\Delta[y(p) - y(50)]/\Delta[y - y(50)]$. Finally, I use the average of spillovers across both periods (1998 and 2012) as the model-based analogues of the reduced form estimates.

I simulate the model-predicted variance decomposition from an AKM regression using a large firms

---

[12]This approach is computationally more efficient than adding parameters for total labor supply N_h and additional moments for observed employment. They are not completely equivalent because the latter can account for sampling error in observed employment shares by worker group. Given the large sample size, this difference is unlikely to be relevant in practice.

assumption. When firms are large, the internal distributions of worker types and wages in a firm of type g are approximately the same as the model-predicted distributions for that type. Thus, firms of the same type in the model are statistically identical except for size. I also need a stylized model of employment dynamics, because the model described in the previous section is static. Define a worker type i as a tuple (h_i, ε_i). First, I assume that there is mobility between firm types for all worker types i that are not rejected by either firm. Second, I assume that ε_{it} is a combination of a portable human capital component (the worker fixed effect) plus a transient effect, with the relevance of the transient effect calibrated so that the share of the variance of log wages attributable to the residual is of similar size to that in the data.[13]

Given these assumptions, the variance decomposition is calculated from a simulation of the model where there are only two large firms in the economy, one for each firm type. First, I discretize the distribution of efficiency units ε, so that there is a discrete number of worker types (h_i, ε_i). Then, I simulate a data set where each observation is a combination of worker type i and firm type g. This data set contains indicators for firm and worker type, worker earnings, and weights constructed from the model-predicted mass of workers of each type i employed by firms of each type g. Next, I estimate firm and worker effects via a weighted ordinary least squares regression of log earnings on firm and worker type dummies.[14] Finally, I calculate the variance decomposition components using those estimated fixed effects.

The parametric space is the real line for parameters defined in logs and the positive real line for most others. The exceptions are comparative advantages $\alpha_h \in (0,1)$ and $\beta_h \in (1,100)$. Standard errors will be provided in a future version of this paper.

### 4.2.3 Estimates, goodness of fit, and discussion

Table 4 shows the estimated parameters and Table 3 illustrates the moments predicted by the estimated model. Overall, the model fits the data well, particularly for the first two groups of moments (levels and trends in wage inequality and components of the variance decomposition based on the two-way fixed effects regressions). That the model fits within-group variances well is particularly interesting given the restriction that the variance of labor efficiency units is the same for all groups and both periods.

---

[13]Formally, I set $\varepsilon_{it} = \varepsilon_i^F + \varepsilon_{it}^T$, with the transitory component ε_{it}^T orthogonal to the fixed component ε_i^F. The variances of each of these components is calibrated so that (i) the variance of ε_{it} is the parameter S^2; and (ii) the estimated share of the residuals in the model-based variance decomposition is 0.9.
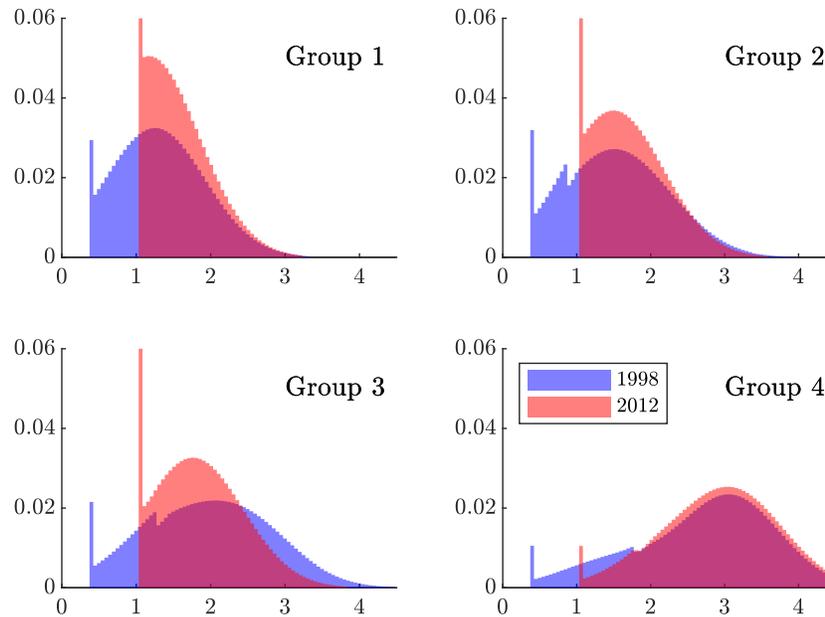
[14]OLS estimates are consistent in this regression because the number of firm fixed effects being estimated does not grow with the sample size.

**Figure 6:** Distribution of log wages, data and model

(a) Data
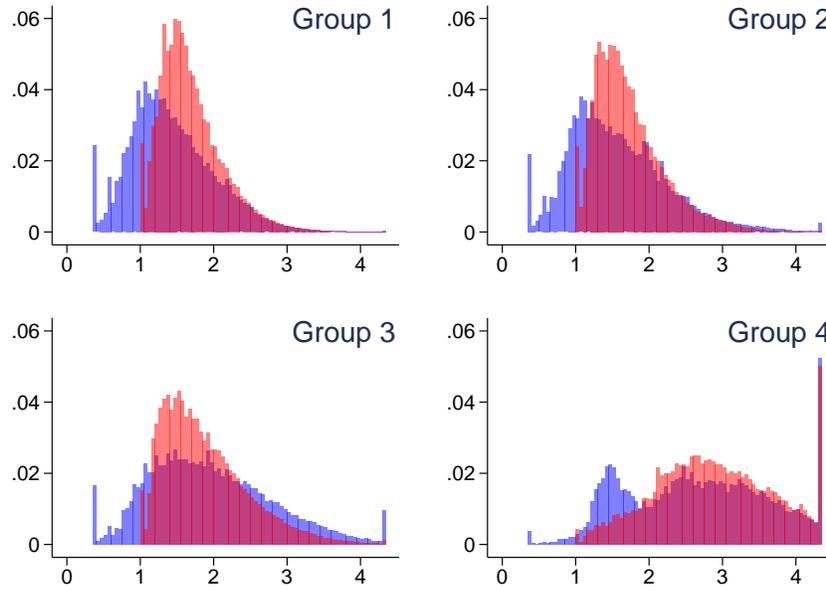
(b) Simulation from estimated model

Notes: This figure shows histograms of log wages using 0.05-sized bins, separately by educational group (No degree, Primary, Secondary, and Tertiary) and time (1998 in blue, 2012 in red). Panel (a) shows data from RAIS, Rio Grande do Sul, Brazil, hours-weighted. Panel (b) shows histograms predicted by the estimated model.

Figure 7: Minimum wage spillovers
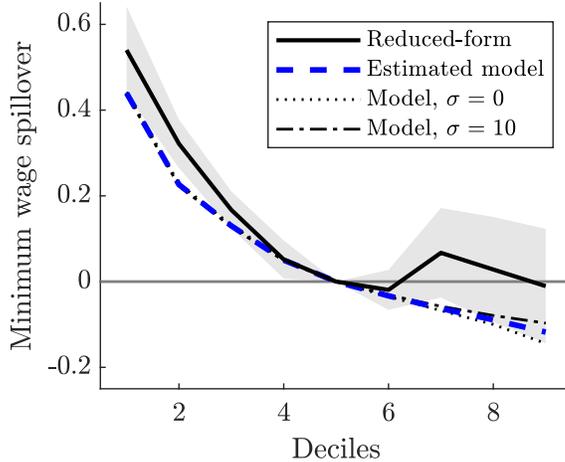
Notes: This figure shows the impact of changes in the "effective minimum" (statutory minimum wage minus median wage) on different deciles of the wage distribution, relative to the mean. The solid line shows marginal effects from the instrumental variable estimation of Equation (12), following Autor, Manning and Smith (2016). The shaded area represents 95 percent confidence intervals for these estimates. The dashed line shows spillovers predicted from the estimated model. The dash-dot line and the dotted line demonstrate the spillovers from a model similar to that of the estimated one, with the exception that substitution between goods is either ruled out (that is, $\sigma = 0$) or amplified ($\sigma = 10$) relative to the baseline estimation ($\sigma = 2$).

Figure 7 shows that minimum wage spillovers from the model are similar to the reduced form estimates, though the model understates positive spillovers in the lower tail and overstates negative spillovers in the upper tail. The same figure also shows how the elasticity of substitution parameter substitution between goods affects comparative statics, following the logic introduced in Proposition 4. More substitution between good leads to a less negative spillovers in the upper tail of the distribution. Those differences, however, are small.[15]

The fit of the model can be verified visually by comparing observed wage distributions with model-generated ones in Figure 6. Overall, the model captures the most salient features of the data. The fit is worse close to the minimum wage, with the model over-predicting bunching in this area.

Now I discuss the estimated parameters shown in Table 4 and the implied equilibria in the model. I start with worker-related parameters. Comparative advantage α_h is increasing in education, as expected. These parameters are identified from changes in mean log wage gaps. The taste for consumption β_h is estimated to be decreasing with education. Workers with no complete degree have a high taste for consumption, implying that labor markets are close to competitive for that

---

[15] In this exercise, productivity gaps are adjusted so that employment shares of each firm type remain the same after the change in σ. As a result, all other moments are the ones shown in Table 3.

Figure 8: Wages and employment by firm type and worker type
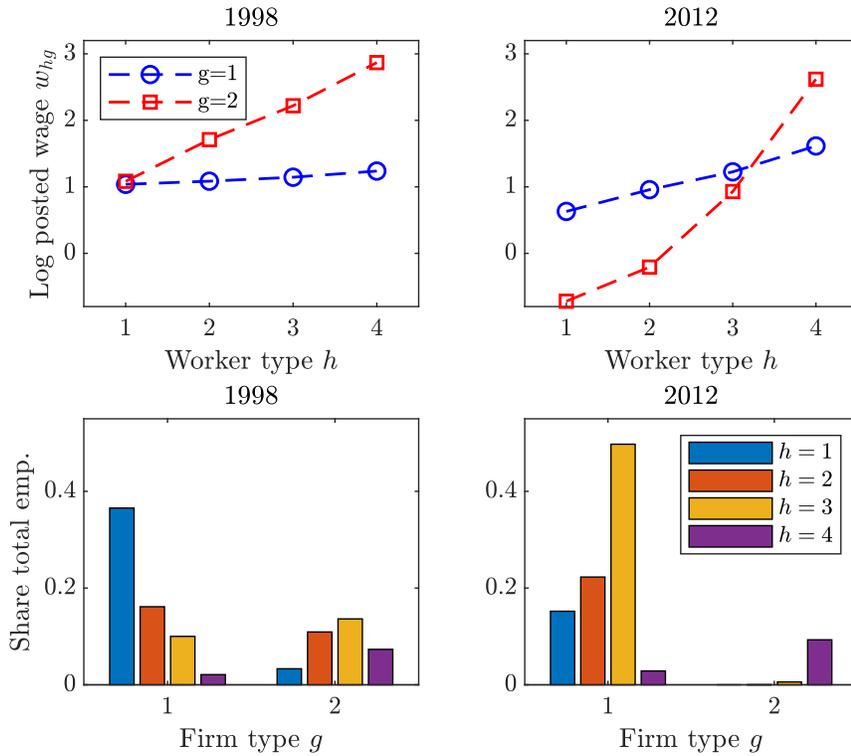
Notes: The top panel shows wages for each worker type h posted by firms producing good g. The bottom panel shows the distribution of employment in the economy between worker types and firm types. For each year, all bars sum to one.

worker group. For workers with more education, the implied wage mark-downs relative to marginal productivities range from 12 percent (for workers with basic schooling) to 20 percent (for workers with a college degree). The variance of efficiency units S^2 is 0.447, close to the total variance of log wages in 2012. Both S^2 and β_h are identified essentially from levels and changes in the variance of log wages within worker groups.

Good $g = 2$ is more intensive in task complexity, has higher fixed costs, and requires fewer tasks in production than good $g = 1$. The estimated skill-biased component of the demand shock is positive, stretching the distribution of task requirements for both goods to the right. The demand shock also includes convergence between goods in entry costs and productivity.

The resulting employment patterns, along with the mean log wages for workers of each type employed at firms producing each good, are shown in Figure 8. Firms producing the second good are more skill intensive in both periods. In the first period, firms producing $g = 2$ pay higher wages to all worker types except the least educated one. The size of the wage premium is increasing in

worker type. These differences stem from a combination of increased demand for skilled workers along with a higher ratio of entry costs to amenities in the second firm type relative to the first, as described in Proposition 2. These differences explain both the substantial share of firm effects in the variance of log wages and why the estimated firm effects from the log-additive model correlate with worker effects.

In the second period, firms producing $g = 2$ still pay more to skilled workers. However, they post lower wages for less skilled workers, reflect extremely low labor demand for them. The combination of these effects explains why firm fixed effects become less relevant as a share of the total variance of log wages, while at the same time the correlation between worker and firm fixed effects increases.

I interpret the demand shock as follows. In the first period, the first good, $g = 1$ represents a "back-ward" technology intensive in low-complexity tasks. The second good, $g = 2$ represents a "modern" technology that has higher returns to education. The demand shock replaces the "backward" technology to another that uses more complex tasks and is closer to the modern one in entry costs and physical productivity. At the same time, the "modern" technology is also affected by skill-biased technical change, becoming even more specialized in high-complexity tasks. Following a concurrent increase in the supply of education, firms producing the modern good find it profitable to fully specialize in college-educated workers.

4.3   Disentangling the role of supply, demand, and minimum wages

In this final step, I use the model to generate counterfactuals that isolate the role of supply, demand, and minimum wage shocks. The first counterfactual scenario has all parameters from the estimated model in 1998, except for labor supply —- which changes to the 2012 levels – and minimum wages, which adjust relative to the numeraire of the model so that it remains constant relative to mean log wages.[16] Next, I move from this scenario to another where the demand shocks occurred, and the minimum wage is still kept constant relative to mean log wages. Finally, the last step measures

---

[16]It is easy to observe the number of employed workers in each educational group, but it is not obvious how to measure total labor supply N_h in the model due to involuntary non-employment. I use a structural approach to deal with this issue. The estimation procedure looks for one equilibrium in the model for each period such that the model predictions (including share of employment by education) match the data. After the equilibrium and its corresponding set of parameters is identified, the parametric assumption of the distribution of efficiency units allows the researcher to extrapolate the number of unemployed workers, thus obtaining the total labor supply N_h. Thus, the change imposed in the first counterfactual is moving from the estimated N_h for the 1998 equilibrium to the one coming from the 2012 equilibrium.

Table 5: Model-based decomposition: supply, demand, and minimum wages.

Moment	All changes	Supply	Demand	Minimum wage
Variance of log wages	-0.201	0.004	-0.079	-0.125
Mean log wage gaps:				
Primary / No degree	-0.085	-0.081	0.102	-0.106
Secondary / Primary	-0.259	-0.088	-0.100	-0.070
Tertiary / Secondary	0.302	-0.010	0.424	-0.112
Within-group variance of log wages:				
No degree	-0.147	-0.007	-0.032	-0.108
Primary	-0.209	-0.034	-0.064	-0.111
Secondary	-0.354	-0.038	-0.225	-0.091
Tertiary	-0.271	0.073	-0.327	-0.017
Variance decomposition from two-way fixed effects model				
Share of firm effects	-0.079	0.081	-0.193	0.033
Corr. worker and firm effects	0.177	-0.080	0.281	-0.024

Notes: The meaning of each column is as follows: *All changes* denotes changes predicted by the estimated model, comparing the 1998 equilibrium to the 2012 equilibrium. *Supply* shows differences between the 1998 equilibrium and a counterfactual equilibrium using 1998 parameters, except for the number of workers of each type (changed to the 2012 value) and minimum wages (adjusted so that the minimum wage to mean log wage remains constant). *Demand* shows the differences between the first counterfactual and a second counterfactual where the demand shock is imposed, while still holding minimum wages stable relative to mean log wages. *Minimum wage* shows differences between the second counterfactual and the 2012 equilibrium.

the difference from this second counterfactual to the estimated model in the second period, which includes changes in all dimensions.

Before showing the results of this model-based decomposition, it is worth noting that this exercise requires a strong assumption: each of these components is entirely exogenous, such that moving either one in isolation is a meaningful counterfactual. One example of a deviation would be an endogenous technical response to increased availability of qualified labor. Another example is labor supply responses to changes in wages following demand and minimum wage shocks. The labor supply example is particularly relevant for two reasons. First, in contrast to other model-based exercises, I do not restrict the sample to male workers, a group that usually has a low elasticity of labor supply. Second, the informal sector provides an "outside option" for workers in the formal sector, and the gap in attractiveness between formal and informal jobs might be a complicated function of the labor supply, technical change, and minimum wages (Haanwinckel and Soares, 2016). A model-based account of the informal sector is possible, but it is beyond the scope of this

paper.

Even with these limitations, this exercise can still shed light on the causes of decreased inequality in Brazil, because it incorporates equilibrium effects while studying a wide range of shocks. There are some model-based decompositions of wage inequality in Brazil, but they focus on smaller sets of shocks and causal mechanisms relative to this paper. Engbom and Moser (2018) develop a model of between-firm wage dispersion and focus on the role of the minimum wage, but do not account for equilibrium effects of changes in educational achievement and technical change. Mak and Siow (2018) build a model of within-firm complementarities between workers and use it to study the role of labor supply. That paper does not feature between-firm wage dispersion and minimum wages. Finally, (Ferreira, Firpo and Messina, 2017) performs a decomposition exercise that features a broad array of explanatory factors and includes both the formal and informal sectors. That exercise, however, does not account for equilibrium effects of changes in labor supply, demand, and institutions; while helpful as a descriptive tool and for ruling out some possible explanations, it does not offer measurements of the causal effect of each shock.

With these caveats in mind, I discuss the results of the model-based decomposition. The supply shock reduces mean log wage gaps among workers with less than tertiary education, but have no effect on the returns to college relative to high school. These effects are consistent with a straightforward supply-demand intuition, since the supply of high school and college educated workers increases substantially. The effects on within-group variances of log wages is small. Finally, the overall effect on the total variance of log wages is null; even though there are reductions in between-group log wage gaps, there is also an endowment effect that increases the overall dispersion of wages in the economy. This result is similar to the conclusions in Ferreira, Firpo and Messina (2017). The supply shock also causes a moderate increase in the share of the variance of log wages explained by firm effects, along with a moderate decline in the correlation of worker and firm fixed effects.

As described in the previous subsection, the estimated demand shock is a combination of skill-biased technical change (modeled as a drift towards more complex tasks) and convergence between firm types in entry costs and productivity. With a representative firm, increased task complexity should widen all between-group wage gaps. Following this logic, there is a large increase in wages for college-educated workers. However, high school workers lose relative to those with only primary education. The reason for this "wage polarization" effect is the loss of firm-level wage premia for high-school workers. Without the minimum shock, most workers in this group are employed by "modern" firms where their skills are valuable. As described in the previous section, the demand

shock induces "modern" firms to fully specialize in college-educated workers. The high school workers are then reallocated to "backward" firms, where their comparative advantage relative to less-educated workers is smaller.

The demand shock also causes reductions in cross-firm wage dispersion, which reflect in an overall decline in wage inequality. There are two separate channels leading to these reductions. First, because the modern firms specialize in college-educated workers following the increase in task complexity, there is effectively no cross-firm wage dispersion for workers up to high school in the second period. That explains decreased within-group inequality for these workers following the demand shock. In addition, the reduction in the entry cost gap between goods reduced the size of the cross-firm wage premia for college educated workers, with corresponding consequences for the variance of log wages within that group. The combined effect of these changes is a sizable reduction in overall inequality, accounting for almost 40 percent of the total decline in the variance of log wages.

The demand shock also makes the labor market more assortative, as measured by the correlation between worker and firm fixed effects in the AKM regressions. This result follows from full specialization in college-educated workers in the modern firms, along with the fact that the wage premia for these workers remain positive (though smaller in magnitude).

The minimum wage is the most important factor explaining declining wage inequality in this context. It has sizable effects in both between-group and within-group inequality. I also find that minimum wages have small impacts on the share of the variance of log wages explained by firm effects and in the correlation between worker and firm effects. These results are similar to those in Engbom and Moser (2018).

5 Conclusion

This paper demonstrated that a task-based model of production is a useful tool for modeling firm heterogeneity in imperfectly competitive labor markets. The theory reveals two kinds of novel interactions between the supply-demand-institutions framework commonly employed to study wage inequality and the applied microeconomic literature examining imperfect competition in labor markets using matched employer-employee data. First, shocks to labor supply, labor demand, and minimum wages affect between-firm wage dispersion and labor market sorting. Second, adding firm heterogeneity and imperfect competition to the traditional approach leads to aggregation issues that

might change how those shocks affect the wage distribution in a qualitative sense.

The application of the model using Brazilian data showed that the channels introduced in the theory are relevant for explaining the evolution of inequality and labor market sorting. It also showed that demand-side shocks, including skill-biased technical change, play essential roles in the labor market transformations in Brazil, despite the observed declines in wage inequality.

I conclude this section by discussing two directions for further research. First, a version of the model with capital provides a tool for modeling different forms of technical changes, as well as capital-skill complementarity. Acemoglu and Autor (2011) employ a task-based structure to model routine-biased technical change resulting from price reductions for types of capital particularly effective at tasks executed by mid-skill workers. The same idea can be introduced in this framework. A simple way to do so is to include different vintages of capital, all of which perfectly substitute for labor at individual tasks but with varying schedules of productivity. This is simple from a modeling perspective because capital vintages behave as worker types do, thus requiring no modification of the model. Each vintage is then a substitute for some worker types and a complement to others. A different approach is to assume capital and labor are imperfect substitutes in the production of tasks, and that this elasticity of substitution decreases in task complexity. This provides a microfoundation for capital-skill complementarity, a pattern that has been documented in the labor demand literature (Hamermesh, 1996) and that is potentially relevant for explaining trends in wage inequality (Krusell et al., 1999).

The model can also be used to quantify the impact of trade shocks on wages. Economists have increasingly used models with heterogeneous firms when studying the inequality effects of trade. Some papers (Sampson, 2014; Burstein, Morales and Vogel, 2016; Burstein and Vogel, 2017) show, using competitive models, that trade liberalization can affect the returns to skill because it favors firms that are more skill-intensive. Others (Helpman, Itskhoki and Redding, 2010; Davis and Harrigan, 2011; Helpman et al., 2017) highlight that trade opening can also increase between-firm wage dispersion for similar workers when labor markets are not competitive. The model developed in this paper can combine both perspectives, while also accounting for the ex-ante worker heterogeneity in productivity, firm-to-worker sorting, and other shocks to wage inequality coming from the traditional supply-demand-institutions framework.

References

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Abowd, John M., Francis Kramarz, Sébastien Pérez-Duarte, and Ian M. Schmutte. 2018. "Sorting Between and Within Industries: A Testable Model of Assortative Matching." *Annals of Economics and Statistics*, , (129): 1–32.

Acemoglu, Daron. 1999. "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence." *The American Economic Review*, 89(5): 1259–1278.

Acemoglu, Daron. 2001. "Good Jobs versus Bad Jobs." *Journal of Labor Economics*, 19(1): 1–21.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. Vol. 4, Part B, , ed. David Card and Orley Ashenfelter, Chapter 12, 1043–1171. Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Artificial Intelligence, Automation and Work." National Bureau of Economic Research Working Paper 24196.

Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser. 2018. "Firms and the Decline in Earnings Inequality in Brazil." *American Economic Journal: Macroeconomics*, 10(1): 149–89.

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3): 673–697.

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2006. "Offshoring in a Knowledge Economy." *The Quarterly Journal of Economics*, 121(1): 31–77.

Ashenfelter, Orley C., Henry Farber, and Michael R Ransom. 2010. "Labor Market Monopsony." *Journal of Labor Economics*, 28(2): 203–210.

Autor, David H. 2014. "Skills, education, and the rise of earnings inequality among the other 99 percent." *Science*, 344(6186): 843–851.

Autor, David H., Alan Manning, and Christopher L. Smith. 2016. "The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment." *American Economic Journal: Applied Economics*, 8(1): 58–99.

Autor, David H, Frank Levy, and Richard J Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics*, 118(4): 1279–1333.

Autor, David H, Lawrence F Katz, and Melissa S Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics*, 90(2): 300–323.

Bound, John, and George Johnson. 1992. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *The American Economic Review*, 82(3): 371–392.

Burstein, Ariel, and Jonathan Vogel. 2017. "International Trade, Technology, and the Skill Premium." *Journal of Political Economy*, 125(5): 1356–1412.

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel. 2016. "Changes in between-group inequality: computers, occupations, and international trade."

Caliendo, Lorenzo, and Esteban Rossi-Hansberg. 2012. "The Impact of Trade on Organization and Productivity." *The Quarterly Journal of Economics*, 127(3): 1393–1467.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics*, 36(S1): S13–S70.

Card, David, and John E DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics*, 20(4): 733–783.

Card, David, and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *The Quarterly Journal of Economics*, 116(2): 705–746.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics*, 128(3): 967.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2018. "The Effect of Minimum Wages on Low-Wage Jobs: Evidence from the United States Using a Bunching Estimator." Centre for Economic Performance, LSE CEP Discussion Papers dp1531.

Chade, Hector, Jan Eeckhout, and Lones Smith. 2017. "Sorting through Search and Matching Models in Economics." *Journal of Economic Literature*, 55(2): 493–544.

Coase, R H. 1937. "The Nature of the Firm." *Economica*, 4(16): 386–405.

Costinot, Arnaud, and Jonathan Vogel. 2010. "Matching and Inequality in the World Economy." *Journal of Political Economy*, 118(4): 747–786.

Cragg, John G., and Stephen G. Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory*, 9(2): 222–240.

Davis, Donald R., and James Harrigan. 2011. "Good jobs, bad jobs, and trade liberalization." *Journal of International Economics*, 84(1): 26 – 36.

de Melo, Rafael Lopes. 2018. "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence." *Journal of Political Economy*, 126(1): 313–346.

De Negri, João Alberto, Paulo Furtado de Castro, Natalia Ribeiro de Souza, and Jorge Saba Arbache. 2001. "Mercado Formal de Trabalho: Comparação entre os Microdados da RAIS e da PNAD." IPEA Texto para Discussão 840.

DiNardo, John, Nicole M Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64(5): 1001–1044.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying Sorting—In Theory." *The Review of Economic Studies*, 78(3): 872–906.

Eeckhout, Jan, and Philipp Kircher. 2018. "Assortative Matching With Large Firms." *Econometrica*, 86(1): 85–132.

Engbom, Niklas, and Christian Moser. 2018. "Earnings Inequality and the Minimum Wage: Evidence from Brazil." Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute Working Papers 7.

Ferreira, Francisco H. G., Sergio P. Firpo, and Julián Messina. 2017. "Ageing Poorly?: Accounting for the Decline in Earnings Inequality in Brazil, 1995-2012." Inter-American Development Bank IDB Publications (Working Papers) 8220.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, 108(5): 874–904.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *The Quarterly Journal of Economics*, 121(4): 1383–1435.

Gautier, Pieter A, and Coen N Teulings. 2015. "Sorting and the Output Loss due to Search Frictions." *Journal of the European Economic Association*, 13(6): 1136–1166.

Gautier, Pieter A, Coen N Teulings, and Aico van Vuuren. 2010. "On-the-Job Search, Mismatch and Efficiency." *The Review of Economic Studies*, 77(1): 245–272.

Grossman, Gene M, Elhanan Helpman, and Philipp Kircher. 2017. "Matching, Sorting, and the Distributional Effects of International Trade." *Journal of Political Economy*, 125(1): 224–264.

Haanwinckel, Daniel, and Rodrigo R. Soares. 2016. "Workforce Composition, Productivity, and Labor Regulations in a Compensating Differentials Theory of Informality." Institute for the Study of Labor (IZA) IZA Discussion Papers 9951.

Hamermesh, Daniel S. 1996. *Labor demand.* princeton University press.

Harasztosi, Péter, and Attila Lindner. 2018. "Who Pays for the Minimum Wage?" Unpublished manuscript.

Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding. 2010. "Inequality and Unemployment in a Global Economy." *Econometrica*, 78(4): 1239–1283.

Helpman, Elhanan, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J Redding. 2017. "Trade and Inequality: From Theory to Estimation." *The Review of Economic Studies*, 84(1): 357.

Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2011. "Frictional Wage Dispersion in Search Models: A Quantitative Assessment." *The American Economic Review*, 101(7): 2873–2898.

Katz, Lawrence F, and Kevin M Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, 107(1): 35–78.

Kline, P., R. Saggio, and M. Sølvsten. 2018. "Leave-out estimation of variance components." *ArXiv e-prints*.

Kremer, Michael, and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill." National Bureau of Economic Research Working Paper 5718.

Krugman, Paul. 1985. "A 'Technology Gap' Model of International Trade." *Structural Adjustment in Developed Open Economies*, , ed. Karl Jungenfelt and Douglas Hague, 35–61. London:Palgrave Macmillan UK.

Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante. 1999. "Capital-skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica*, 68(5): 1029–1053.

Lee, David S. 1999. "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?" *The Quarterly Journal of Economics*, 114(3): 977–1023.

Lentz, Rasmus, and Dale T Mortensen. 2010. "Labor market models of worker and firm heterogeneity." *Annu. Rev. Econ.*, 2(1): 577–602.

Lindenlaub, Ilse. 2017. "Sorting Multidimensional Types: Theory and Application." *The Review of Economic Studies*, 84(2): 718–789.

Lindenlaub, Ilse, and Fabien Postel-Vinay. 2017. "Multidimensional Sorting Under Random Search." Unpublished manuscript.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2016. "Matching, sorting and wages." *Review of Economic Dynamics*, 19: 63 – 87. Special Issue in Honor of Dale Mortensen.

Mak, Eric, and Aloysius Siow. 2018. "Occupational Choice, Matching and Earnings Inequality." Unpublished manuscript.

Manning, Alan. 2011. "Imperfect Competition in the Labor Market." In *Handbook of Labor Economics*. Vol. 4, Part B, , ed. David Card and Orley Ashenfelter, Chapter 11, 973–1041. Elsevier.

Sampson, Thomas. 2014. "Selection into Trade and Wage Inequality." *American Economic Journal: Microeconomics*, 6(3): 157–202.

Shimer, Robert. 2005. "The Assignment of Workers to Jobs in an Economy with Coordination Frictions." *Journal of Political Economy*, 113(5): 996–1025.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2018. "Firming Up Inequality." Forthcoming, The Quarterly Journal of Economics.

Teulings, Coen N. 1995. "The Wage Distribution in a Model of the Assignment of Skills to Jobs." *Journal of Political Economy*, 103(2): 280–315.

Teulings, Coen N. 2000. "Aggregation Bias in Elasticities of Substitution and the Minimum Wage Paradox." *International Economic Review*, 41(2): 359–398.

Teulings, Coen N. 2003. "The Contribution of Minimum Wages to Increasing Wage Inequality." *The Economic Journal*, 113(490): 801–833.

Teulings, Coen N. 2005. "Comparative Advantage, Relative Wages, and the Accumulation of Human Capital." *Journal of Political Economy*, 113(2): 425–461.

A Proofs

Section 2: 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 $\{q, \{m_h\}_{h=1}^H\}$ that satisfy all constraints in the assignment problem.

Lemma 3. *If there exists a candidate solution $\{q, \{m_h(\cdot)\}_{h=1}^H\}$ 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 $\{q', \{m_h'(\cdot)\}_{h=1}^H\}$ 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 $\{q', \{m_h'(\cdot)\}_{h=1}^H\}$ identical to $\{q, \{m_h(\cdot)\}_{h=1}^H\}$, except for:

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

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

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

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

I need to prove that $\{q', \{m_h'(\cdot)\}_{h=1}^H\}$ 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 4. *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 $\tilde{u}_h = \int_{\underline{x}}^{\underline{x}+\delta} e_H(x)/e_h(x)dx < \infty$ 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 $m'_h(x) = m_h(x) + ge_H(x)/e_h(x)$ 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 4, we know that any optimal solution must not have any slack. The same Lemma implies that any candidate solution satisfying the conditions in Lemma 3 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 1) is provided in the next section of this Appendix.

$\square$

Proof of Proposition 1: curvature of the task-based production function and non-arbitrage condition (Equation 1)

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 l^1 and 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 $\bar{l} = \alpha l^1 + (1 - \alpha)l^2$, 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 $f(\bar{l}, b) \geq \bar{q}$. For constant returns, note that, given $\alpha > 1$, output αq^1 is attainable with inputs αl^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 $(l, b_g(\cdot))$, 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 $z : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{>0} \to \mathbb{R}^H$ is given by:*

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

Lemma 5. *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 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 l also exist and are continuous. $\qquad\square$

Lemma 6. *The Jacobian of the excess labor demand function regarding $(q, \bar{x}_1, \ldots, \bar{x}_{H-1})$, when evaluated at a point where $z(\cdot) = \mathbf{0}_{H \times 1}$, has non-zero determinant.*

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

$$J = \begin{bmatrix}
\frac{l_1}{q} & q\frac{b_g(\bar{x}_1)}{e_1(\bar{x}_1)} & 0 & 0 & \cdots & 0 & 0 \\[6pt]
\frac{l_2}{q} & -q\frac{b_g(\bar{x}_1)}{e_2(\bar{x}_1)} & q\frac{b_g(\bar{x}_2)}{e_2(\bar{x}_2)} & 0 & \cdots & 0 & 0 \\[6pt]
\frac{l_3}{q} & 0 & -q\frac{b_g(\bar{x}_2)}{e_3(\bar{x}_2)} & q\frac{b_g(\bar{x}_3)}{e_3(\bar{x}_3)} & \cdots & 0 & 0 \\[4pt]
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\[4pt]
\frac{l_{H-1}}{q} & 0 & 0 & 0 & \cdots & -q\frac{b_g(\bar{x}_{H-2})}{e_{H-1}(\bar{x}_{H-2})} & q\frac{b_g(\bar{x}_{H-1})}{e_{H-1}(\bar{x}_{H-1})} \\[6pt]
\frac{l_H}{q} & 0 & 0 & 0 & \cdots & 0 & -q\frac{b_g(\bar{x}_{H-1})}{e_H(\bar{x}_{H-1})}
\end{bmatrix}$$

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 $b(x), e_h(x) > 0\ \forall x, h$. $\qquad\square$

Lemmas 5 and 6 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 $q(l) = f(l, b_g(\cdot))$ and $\bar{x}_h(l)$. Because 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 1, 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:

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

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 $\xi_{h,h',\mathfrak{h}}$. 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 Corollary 2: Effect of minimum wages

In a competitive economy where the minimum wage binds for only the lowest worker type, a marginal increase in the minimum wage causes unemployment of that lowest type up to the point where marginal revenue product of labor equals wages. That is exactly the same comparative statics as reducing the supply of the least skilled labor type by that corresponding amount. The

proportional effect of that change on marginal productivities of all labor types is given by the negative of the elasticity of complementarity times the share of labor costs attributable to the lowest worker types. Distant-dependent complementarity constrains that effect to be decreasing over the wage distribution (because the elasticities of complementarity are increasing).

Section 3: Markets and wages

Proof of Lemma 2: Firm problem and representative firms

The most difficult part of this proof is showing that the solution is interior and characterized by first order conditions. I start by implicitly defining a function $\underline{\varepsilon}_h(l_{hj}, w_{hj}, a_j)$ that maps desired labor $l_{hj} > 0$ to the required cutoff point given a choice of posted wage and firm amenities:

$$l_{hj} = l_h\left(w_{hj}, \underline{\varepsilon}_h(l_{hj}, w_{hj}, a_j), a_j\right)$$

That function is not defined for all combinations of (l_{hj}, w_{hj}, a_j) because, given w_{hj}, $l_{hj} = 0$ is always reachable (by setting $\underline{\varepsilon}_h \to \infty$) but there is an upper bound to l_{hj} as $\underline{\varepsilon}_h \to -\infty$. For my purposes, it is sufficient that it is defined around optimal values of l_{hj} and w_{hj}. The implicit function $\underline{\varepsilon}_h(l_{hj}, w_{hj}, a_j)$ is differentiable, but the derivative is not continuous where there are discontinuities in $\omega_h(\varepsilon)$. Furthermore, it is strictly decreasing in l_{hj} and strictly increasing in w_{hj}.

Now redefine the problem of the firm in terms of a choice of labor inputs given a labor cost function implied by the labor market microstructure:

$$\pi_g(a_j) = \max_{l_j} p_g f\left(\mathbf{l}_j, b_g(\cdot)\right) - \sum_{h=1}^{H} \tilde{C}_h\left(l_{hj}, a_j\right)$$

$$\text{where } \tilde{C}_h\left(l_{hj}, a_j\right) = \min_{w_{hj}} C_h\left(w_{hj}, \underline{\varepsilon}_h(l_{hj}, w_{hj}, a_j), a_j\right)$$

Totally differentiating the minimand in the cost minimization problem regarding w_{hj} yields:

$$l_{hj}^{+}\left[(\beta_h + 1) - \frac{y}{\underline{\varepsilon}_h(l_{hj}, w_{hj}, a_j)} \frac{\beta}{w_{hj}}\right]$$

where l_{hj}^{+} is the amount of efficiency units supplied by workers who earn more than the minimum wage. This derivative is strictly increasing in w_{hj}, which means that the cost-minimizing wage

56

$w_h(l_{hj}, a_j)$ is defined by the first order condition:

$$\frac{w_h(l_{hj},a_j)\underline{\varepsilon}_h(l_{hj},w_h(l_{hj},a_j),a_j)}{\underline{y}} = \frac{\beta}{\beta+1}$$

This expression shows that wages and cutoffs move in opposite directions. To satisfy increased labor demand, firms both increase posted wages and decrease cutoffs.

The marginal cost of labor for all $l_{hj} > 0$ can then be found by using the envelope theorem on the cost minimization problem, yielding $\underline{y}/\underline{\varepsilon}_h(l_{hj}, w_h(l_{hj},a_j),a_j)$. It can be shown that $l_{hj} \to 0 \Rightarrow \underline{\varepsilon}_h(l_{hj}, w_h(l_{hj},a_j),a_j) \to \infty$. So marginal costs are arbitrarily close to zero as labor demand decreases.

The economic intuition for why labor costs go to zero as labor demand becomes small is as follows. With low labor requirements, firms can set arbitrarily large cutoffs and low wages. There are a few workers for whom the idiosyncratic component of preferences η_{ij} is so large that they still choose that firm. These workers mostly cost the minimum wage, given the low posted wage, but they provide increasing quantities of efficiency units as the cutoff increases.

To rule out corner solutions, note that the marginal product of the highest type $h = H$ is bounded below due to the feasibility constraint on blueprints — these workers are always capable of producing a positive amount of output by themselves, without exploring any comparative advantage gains. So there is positive employment for those types. In addition, Proposition 1 states that the marginal product of labor is strictly positive even when $l_{hj} = 0$, as long as there is positive employment of the higher type. That rules out corner solutions for the other types.

To obtain the first order conditions in terms of w_{hj} and $\underline{\varepsilon}_{hj}$, one can take the corresponding derivatives of (5) and (6). Alternatively, one can obtain the optimal cutoff by equating the marginal cost of labor above to the marginal revenue product of labor, and then use the first order condition on cost minimization to find the mark-down rule.

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 $C_h\left(w_{hj}, \underline{\varepsilon}_{hj}, a_j\right)$ and $l_h\left(w_{hj}, \underline{\varepsilon}_{hj}, a_j\right)$. 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 2: Wage differentials across firms

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

Lemma 7. *If $b_g(x) = b(x)/z_g$ for a blueprint $b(\cdot)$ and scalar $z_g > 0$, then $f(l, b_g(\cdot)) = z_g f(l, b(\cdot))$.*

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 $f(l, b(\cdot))$ by definition. ∎

Now I proceed to the proof of each statement of Proposition 2 separately.

Proof of part 1: From Lemma 7, $f_h(l, b_g(\cdot)) = z_g f_h(l, b(\cdot))$. Also note $l(w_g, \epsilon_g, \bar{a}_g) = \bar{a}_g l(w_g, \epsilon_g, 1)$ and $C(w_g, \epsilon_g, \bar{a}_g) = \bar{a}_g C(w_g, \epsilon_g, 1)$, 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 (11) imposing the conditions from this proposition:

$$p_g z_g f_h(l(w_g, \epsilon_g, 1), b(\cdot)) \exp(\underline{\epsilon}_{hg}) = \underline{y} \qquad \forall h, g$$

$$p_g z_g f_h(l(w_g, \epsilon_g, 1), b(\cdot)) \frac{\beta_h}{\beta_h + 1} = w_{hg} \qquad \forall h, g$$

$$\bar{a}_g \left[p_g z_g f(l(w_g, \epsilon_g, 1), b(\cdot)) - \sum_{h=1}^{H} C_h(w_g, \epsilon_g, 1) \right] = \bar{a}_g \tilde{F} \qquad \forall g$$

To see that these equations imply a representative firm for the economy, plug in $\epsilon_g = \epsilon$, $w_g = \lambda = \{\lambda_1, \ldots, \lambda_H\}$, and $p_g = p/z_g$ for common ϵ, λ, 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: $\underline{\epsilon}_{hg} = 0$. 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:

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

Also, rewrite the zero profit condition as:

$$F_g=p_g z_g f\left(\boldsymbol{l}\left(\boldsymbol{w}_g,\cdot,\bar{a}_g\right),b(\cdot)\right)-\sum_{h=1}^{H}C_h\left(\boldsymbol{w}_g,\cdot,\bar{a}_g\right)$$

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

I claim that $\boldsymbol{w}_g=\left(F_g/\bar{a}_g\right)^{1/(\beta+1)}\boldsymbol{\lambda}$ 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:

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

$$=\frac{\beta+1}{z_g f\left(\boldsymbol{\ell},b(\cdot)\right)}\left(\frac{F_g}{\bar{a}_g}\right)^{\frac{1}{\beta+1}}$$

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

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

Which again has no dependency on g, showing that the claimed solution solves the problem for all firms. $\qquad\square$

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.

$\square$

Proof of Proposition 3: Race between technology, education, and minimum wages

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 $\bar{x}_h^g$ denote task thresholds for firm g in the original equilibrium. Thresholds $(1+\Delta_1)\bar{x}_h^g$ lead to exactly the same unit labor demands, except for a proportional reduction:

$$\int_{(1+\Delta_1)\bar{x}_{h-1}^g}^{(1+\Delta_1)\bar{x}_h^g} \frac{b_g'(x')}{e_h'(x')}dx' = \int_{(1+\Delta_1)\bar{x}_{h-1}^g}^{(1+\Delta_1)\bar{x}_h^g} \frac{1}{(1+\Delta_1)(1+\Delta_2)} \frac{b_g(x'/(1+\Delta_1))}{e_h(x'/(1+\Delta_1))}dx' = \frac{1}{1+\Delta_2}\int_{\bar{x}_{h-1}^g}^{\bar{x}_h^g} \frac{b_g(x)}{e_h(x)}dx$$

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.

Proof of Proposition 4: 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 8. *Let b and b' denote blueprints such that their ratio $b'(x)/b(x)$ is strictly increasing. Then:*

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

Proof. Fix l, let $q = f(l,b)$ and $q' = f(l,b')$. Now construct $b''(x) = b'(x)q'/q$. From Lemma 7, it

follows that $f(l, b'') = q$ and $f_h(l, b'') = f_h(l, b') \; \forall h$. 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 (1) 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:

61

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

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

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 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 $f_h(L, B(\cdot))$, 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 $f_h(L, B'(\cdot))$. But $B(x, Q_1', Q_2')/B(x, Q_1, Q_2)$ is increasing in x if $Q_2'/Q_1' > Q_2/Q_1$. Thus, Lemma 8 implies that wage gaps increase as stated in the Proposition.

B Numerical implementation

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

The exponential-Gamma parametrization is helpful because it provides a simple closed form solu-

tion for thresholds and the labor demand integrals. Let:

$$e_h(x) = \exp(\alpha_h x) \qquad\qquad -1 = \alpha_1 < \alpha_2 < \cdots < \alpha_{H-1} < \alpha_H = 0$$

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

Then:

$$\bar{x}_h\left(\frac{f_{h+1}}{f_h}\right) = \frac{\log f_{h+1}/f_h}{\alpha_{h+1} - \alpha_h} \tag{14}$$

$$\ell_{hg}(\bar{x}_{h-1}, \bar{x}_h) = \int_{\bar{x}_{h-1}}^{\bar{x}_h} \frac{b_g(x)}{e_h(x)} dx$$

$$= \begin{cases} \dfrac{1}{z_g \Gamma(k_g)} \left(\dfrac{k_g}{\Upsilon_{hg}\theta_g}\right)^{k_g} \left[\gamma\left(\Upsilon_{hg}\bar{x}_h, k_g\right) - \gamma\left(\Upsilon_{hg}\bar{x}_{h-1}, k_g\right)\right] & \text{if } \Upsilon_{hg} \neq 0 \\[3mm] \dfrac{1}{z_g \Gamma(k_g)} (1-k_g)^{-k_g} \left[\gamma(\bar{x}_{h-1}(1-k_g)/\theta_g, k_g) - \gamma(\bar{x}_h(1-k_g)/\theta_g, k_g)\right] & \text{otherwise} \end{cases}$$

$$\tag{15}$$

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

When $\Upsilon_{hg} < 0$, one needs to use the standard holomorphic extension of the incomplete Gamma function. This extension is readily available in many numerical software packages. In Matlab, Equation 15 can be implemented using the *gammainc* function; it will handle both positive and negative values of $\Upsilon_{hg} < 0$.

If using complex numbers is not convenient, one can employ instead the following power series representation, which only requires real number computations:

$$\ell_{hg}(\bar{x}_{h-1}, \bar{x}_h) = \frac{1}{z_g \Gamma(k_g+1)} \left(\frac{k_g}{\theta_g}\right)^{k_g} \sum_{m=1}^{\infty} \frac{\Upsilon_{hg}^m}{(1+m/k_g)m!} \left[(-\bar{x}_h)^{k_g+m} - (-\bar{x}_{h-1})^{k_g+m}\right]$$

That series converges slowly if Υ_{hg} is large in magnitude. But Υ_{hg} is bounded below at -1. Thus, one can use the power series representation for negative Υ_{hg} and Equation 15 for other values.

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 labor available to the firm. The choice variables can be either $(q, \bar{x}_1, \ldots, \bar{x}_{H-1})$ or $f_1, \ldots, f_H$. 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   Appendix to the quantitative exercise

C.1   Summary statistics

Descriptive statistics for the RAIS dataset are presented in Table 6. Statistics are presented for the whole sample and separately by schooling group.

C.2   Wage inequality and schooling trends using PNAD data

In this Appendix, I analyze the robustness of the main facts presented in Section 4.1 using an alternative data source, the PNAD survey. I proceed in three steps. First, I compare wage inequality and schooling trends for formal sector workers in the two datasets. Second, I expand the sample to include both formal and informal workers to check whether these trends are restricted to the formal sector. Third, I look at schooling achievement for Brazilian adults regardless of their workforce participation status, as a way of investigating whether increased schooling achievement among employed workers reflects changes in selection patterns into employment or fundamental changes in access to schooling for the whole population.

The PNAD is a household survey with national coverage administered by the by the Brazilian Statistical Bureau (IBGE). Jointly with the Census, it is one of the primary sources of nationally representative data on a series of topics that include labor market participation, earnings, and education. It contains thorough information on employment status, including whether workers had a signed "labor card" — that is, whether the employment relationship is formally registered.

This Appendix analyzes PNAD data from 1998 through 2012. The sample I use includes adults 18 through 54 years old that are not in school, the same criterion imposed on RAIS data. I use public use software developed by PUC-Rio's Datazoom project to read the data, make it compatible across years, and deflate income variables. More information about the resulting dataset is available at

Table 6: Summary statistics, RAIS data.

	All Workers	No degree	Primary	Secondary	Tertiary
Panel A: 1998					
Age	33.867	34.838	32.164	32.054	38.716
	(9.338)	(9.601)	(9.296)	(8.615)	(7.700)
Female	0.411	0.303	0.365	0.536	0.626
	(0.492)	(0.459)	(0.482)	(0.499)	(0.484)
Log wage	1.759	1.447	1.596	2.006	2.725
	(0.828)	(0.593)	(0.688)	(0.840)	(0.938)
Public sector	0.274	0.195	0.190	0.335	0.622
	(0.446)	(0.396)	(0.392)	(0.472)	(0.485)
Monthly hours	179.374	185.830	183.650	173.885	158.111
	(26.578)	(19.033)	(20.340)	(29.749)	(39.292)
Number of workers	1,494,186	574,904	394,990	364,376	159,916
Panel B: 2012					
Age	34.501	38.682	34.015	32.554	37.727
	(9.890)	(9.865)	(10.513)	(9.329)	(8.642)
Female	0.452	0.327	0.361	0.476	0.636
	(0.498)	(0.469)	(0.480)	(0.499)	(0.481)
Log wage	1.978	1.692	1.732	1.903	2.909
	(0.701)	(0.434)	(0.487)	(0.597)	(0.776)
Public sector	0.192	0.138	0.109	0.152	0.512
	(0.393)	(0.344)	(0.311)	(0.359)	(0.500)
Monthly hours	179.376	186.569	185.134	182.107	153.702
	(27.319)	(17.788)	(19.368)	(22.342)	(42.728)
Number of workers	2,398,391	350,704	517,748	1,189,063	340,876

This table presents summary statistics (means and standard deviations, in parenthesis) for the RAIS data. The sample includes adults in Rio Grande do Sul state from 18 to 54 years of age who are not in school and who are employed in December, having been hired in November or earlier. Wages are in 2010 Brazilian Reais and are winsorized at the top and bottom 1 percent of the wage distribution in each year.

Figure 9: Measures of wage dispersion, PNAD data, formal sector

Wage inequality, PNAD data, formal workers
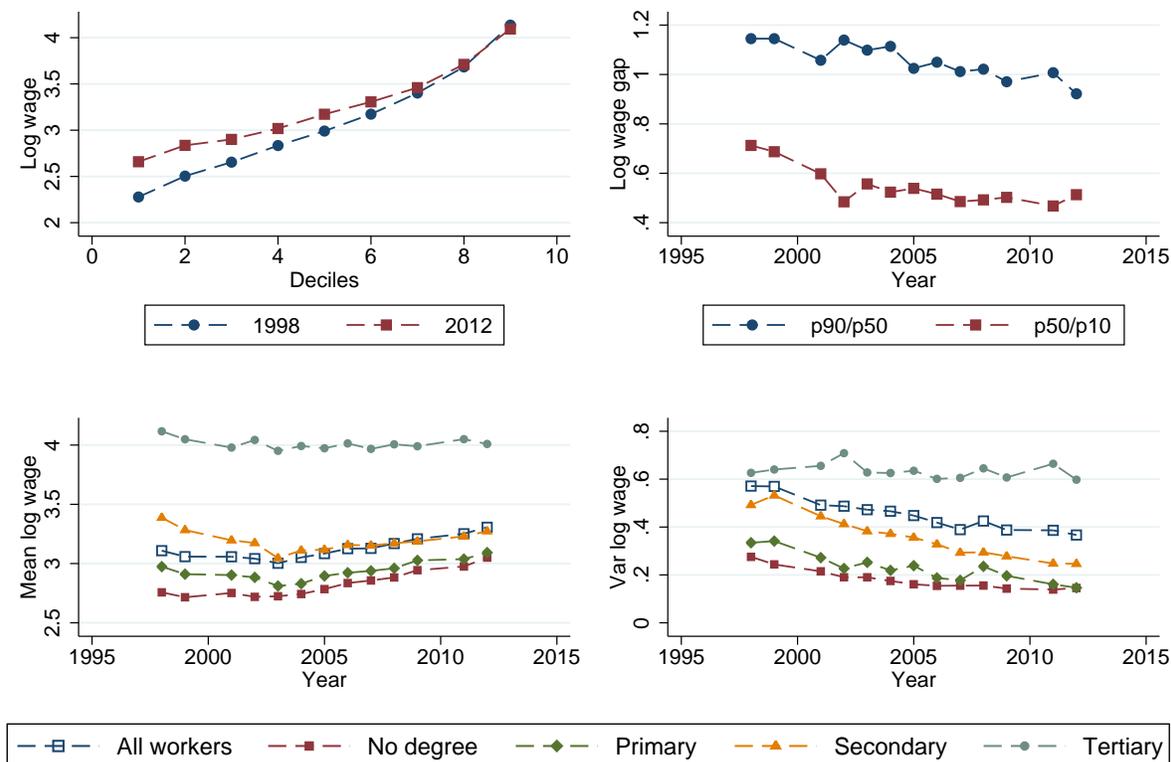
Notes: PNAD data, Rio Grande do Sul, Brazil. Formal sector employees only (including public sector). Observations are weighted by sampling weights multiplied by hours worked.

Datazoom's website.[17]

C.2.1 Comparing RAIS data and PNAD data for formal sector workers

Figure 9 replicates Figure 4 using PNAD data instead of RAIS data. The PNAD sample is constructed to match the RAIS sample, including only formal employees. Overall, the patterns are broadly similar: they show decreased wage inequality along different dimensions. There are two significant differences. First, the mean log wage gap between college and high school workers is stable from 1998 to 2012 in PNAD, but increasing in RAIS. Second, variances of log wages within

---

groups and for the whole sample are larger with the RAIS data for 1998, but not for 2012. Thus, RAIS shows larger reductions in inequality using this measure.

The first panel in Figure 11 replicates the evolution of schooling achievement of formal employees, shown in Figure 5. Again, the overall patterns are broadly similar: there is a substantial decline in the share of hours supplies by workers without any educational degree, accompanied by a similarly large increase in the percentage of hours supplied by workers with complete high school (secondary). There is also an increase in hours supplied by workers with college degrees. There are small changes in the shares; in particular, the PNAD shows a higher fraction of college-educated workers.

There are three reasons for differences between the PNAD and RAIS. First, the RAIS is a census of formal employees, while PNAD is a small sample of that population. While the latter is designed to be representative, it might under-sample some workers with very high or very low earnings. Second, RAIS data are reported by firms, while PNAD data are reported by workers. That might lead to differences if, e.g., workers with high wages under-report in the PNAD or firms misreport the education of workers. Third, there are differences in the primitive questions used to construct wages and years of schooling in each dataset. De Negri et al. (2001) compares PNAD data and RAIS data and provides a detailed account of those differences. The first two reasons suggest that, when assessing inequality trends in the formal sector, RAIS data are probably more reliable than PNAD data.

C.2.2 Inequality trends for the whole workforce

Figure 10 is constructed similarly to Figure 9 above, but the data includes both formal and informal workers. I use a broad definition of the informal sector that includes domestic and self-employed workers. There are no substantial changes in qualitative patterns once informal workers are taken into account. The amount of wage dispersion is higher for the whole sample than for the restricted sample, especially in the lower tail of the wage distribution. One possible candidate for these differences is the presence of the binding minimum wage.

Differences in schooling achievement between the formal sample and the full sample can be observed by comparing the first two panels in Figure 11. Formal sector workers are a selected subsample with higher education levels. However, trends for the whole sample are, again, similar to those obtained from the formal sample.

Figure 10: Measures of wage dispersion, PNAD data, all workers

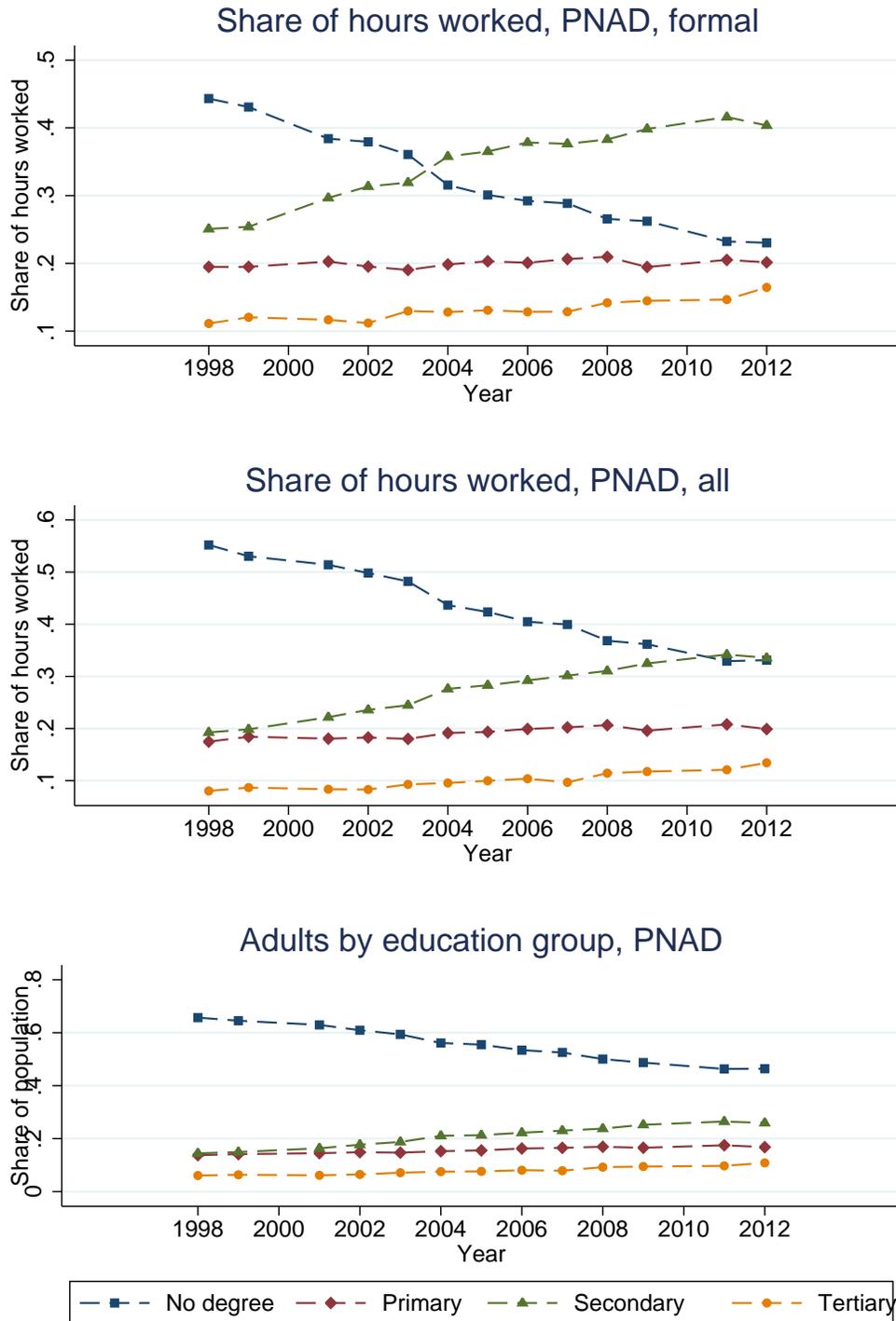Wage inequality, PNAD data, all workers
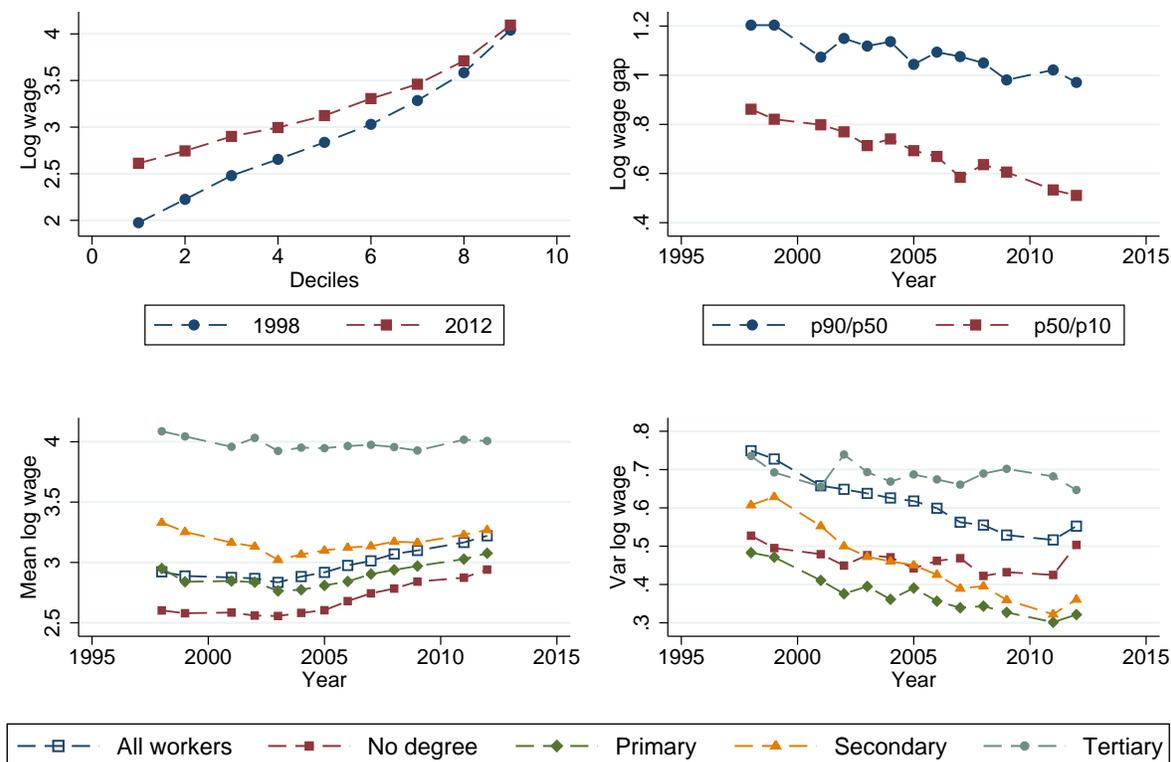
Notes: PNAD data, Rio Grande do Sul, Brazil. All employees (including public sector and informal sector). Observations are weighted by sampling weights multiplied by hours worked.

C.2.3 Changes in relative labor supply

The first two panels of Figure 11, along with Figure 5 in the main text, show shares of hours worked supplied by each schooling group. One might wonder whether these could reflect changes in selection patterns into employment over time (coming, e.g., from business cycle fluctuations) instead of changes in labor supply. The third panel in Figure 11 shows that this is not the case. That graph shows the share of adults out of school, aged 18 through 54, in each educational group — regardless of whether they are employed, looking for jobs, or not in the labor force. The changes in educational achievement from that figure are similar in magnitude to those in the second and first panels. The levels are different, though, suggesting selection into employment by education.

Figure 11: Changes in educational achievement, PNAD data

Share of hours worked, PNAD, formal

Share of hours worked, PNAD, all

Adults by education group, PNAD

—■— No degree   —♦— Primary   —▲— Secondary   —●— Tertiary

Notes: PNAD data, Rio Grande do Sul, Brazil. In the first two panels, the sample includes employed workers and observations are weighted by sampling weights multiplied by hours worked. In the third panel, the sample is composed of all adults 18-54 who are not in school, weighted by sampling weights.

Table 7: Sample sizes in variance decomposition exercise

	1998	2012
Person-year observations	1,618,478	2,570,016
Workers	809,239	1,285,008
Firms	31,107	65,466
Movers	174,299	334,206

This table presents the number of observations in the largest leave-one-out connected set.

C.3 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 the 1998 period, I use data for two years: 1997 and 1999. I use non-consecutive years to increase the number of firm-to-firm transitions.

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 7 presents the size of that largest connected set in each period.

I implement the variance decomposition using the code provided by KSS.[18] There are some implementation choices required in this estimation, stated below:

- Dealing with controls (year fixed effects): "Partialled out" prior to estimation (option 1 in the resid_controls argument).

- Computation of local linear regressions: stratified by grids, separate for movers and stayers (option 2 in the subsample_llr_fit argument).

- Sample selection: includes both movers and stayers (option 0 in the restrict_movers argument).

- Algorithm: Random projection method (option "JLL" in type_of_algorithm option, with epsilon=0.005).

---

[18]Currently available at `https://github.com/rsaggio87/LeaveOutTwoWay`.