NBER WORKING PAPER SERIES

SUPPLY, DEMAND, INSTITUTIONS, AND FIRMS:
A THEORY OF LABOR MARKET SORTING AND THE WAGE DISTRIBUTION

Daniel Haanwinckel

Working Paper 31318
http://www.nber.org/papers/w31318

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023, Revised June 2024

I would like to thank David Card, Fred Finan, Pat Kline, Maurizio Mazzocco, and Andrés Rodríguez-Clare for their guidance and support. The paper benefited from comments from Ben Faber, Thibault Fally, Cecile Gaubert, Jinyong Hahn, Brian Kovak, Thibaut Lamadon, Thomas Lemieux, Nicholas Li, Juliana Londoño- Vélez, Magne Mogstad, Piyush Panigrahi, Raffaele Saggio, Andres Santos, Yotam Shem-Tov, Avner Strulov-Shlain, José P. Vásquez-Carvajal, Christopher Walters, four anonymous referees, and participants at several seminars and conference presentations. I also thank Lorenzo Lagos and David Card for providing code to clean the RAIS data set, and Gustavo de Souza for providing task content data for Brazilian occupations. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution
Daniel Haanwinckel
NBER Working Paper No. 31318
June 2023, Revised June 2024
JEL No. J24,J31,J38

ABSTRACT

This paper builds a general equilibrium framework with firm and worker heterogeneity, monopsony power, and task-based production to quantify the long-run effects of education, labor demand shocks, and minimum wage. I take it to Brazilian data and find that, between 1998 and 2012: (i) supply and demand shocks increased the sorting of high-wage workers to high-wage firms; (ii) endogenous entry of high-wage firms boosted the effect of rising schooling attainment on mean log wages by 33%; (iii) the impacts of the rising minimum wage on inequality and sorting would have been much stronger without the co-occurring supply and demand shocks.

Daniel Haanwinckel
Department of Economics
University of California, Los Angeles
8242 Bunche Hall
Los Angeles, CA 90095
and NBER
haanwinckel@econ.ucla.edu

A data appendix is available at http://www.nber.org/data-appendix/w31318

1 Introduction

Brazil experienced a dramatic reduction in wage inequality between the mid-1990s and the early 2010s. In a literature review, Firpo and Portella (2019) point to three shocks as plausible causes of that phenomenon: an increased supply of skilled labor due to rising educational attainment, labor demand shocks that favored unskilled workers (mostly due to the 2000s commodities boom), and large real increases in the federal minimum wage. Understanding the labor market effects of these shocks is important for not only those interested in the Brazilian case but also those seeking to remedy rising wage inequality in other contexts.

To that end, this paper develops a tractable framework that describes how supply, demand, and minimum wage jointly determine the long-run wage distribution in imperfectly competitive labor markets. I employ matched employer-employee data to test its theoretical predictions and to structurally estimate a local labor markets model of the Brazilian economy. Finally, I simulate counterfactual scenarios based on the estimated model to quantify the individual impacts of each shock, as well as their interactions.

Current academic literature employs two separate frameworks to study the labor market effects of those shocks. Supply and demand factors are typically examined under the assumption of perfect competition, using models with representative firms (e.g., Bound and Johnson, 1992; Card and Lemieux, 2001) or assignment models based on comparative advantage (e.g., Teulings, 1995; Acemoglu and Autor, 2011). In such models, inequality trends reflect changes in productivity gaps between workers. By contrast, leading quantitative models of the minimum wage, such as those developed by Flinn (2006) and Engbom and Moser (2022), are imperfectly competitive. Those models emphasize the contribution of cross-firm wage differentials between equally productive workers (henceforth, firm wage premiums) to overall wage inequality.

Although the use of different frameworks for different shocks facilitates tractability, it also imposes restrictions on causal pathways. In competitive models, supply and demand factors cannot affect wages through firm wage premiums or *sorting*, defined in this paper as the assortativeness between worker skill and the firm wage premium they earn at their current employer. But those channels may be quantitatively important. For example, Card, Heining and Kline (2013) and Song et al. (2018) show that long-run changes in sorting account for significant shares of the overall increase in wage inequality in West Germany and the US, respectively. If those changes in sorting are driven by supply and demand factors, competitive models may provide an incomplete account of their labor market effects. On the minimum wage side, the leading models impose strong restrictions on how productivity gaps between workers may change by assuming perfect substitutability between worker types, ruling out changes in technologies firms may use, or disallowing cost pass-throughs.

A descriptive analysis of the Brazilian case shows that these restrictions may be consequential. I use matched employer-employee data to calculate labor market statistics for 151 *microregions* comparable to US commuting zones. Those statistics include several measures of wage inequality, minimum wage bindingness, and formal employment rates for 1998 and 2012. I also use the methodology detailed by Kline, Saggio and Sølvsten (2018) to obtain reduced-form estimates of the importance of firm wage premiums and sorting, based on two-way fixed effects regressions in the tradition of Abowd, Kramarz and Margolis (1999).

Many of my descriptive findings align with previous work on Brazil: the fall in inequality is large, widespread, and associated with the reduced dispersion of firm wage premiums (Alvarez et al., 2018). However, I also document a new fact not readily explained by existing theoretical approaches: assortative matching rises in most regions. Although papers such as Engbom and Moser (2022) allow for the minimum wage to impact sorting, it acts in the opposite direction. Thus, the new finding challenges the hypothesis that the evolution of Brazilian labor markets can be easily understood through the lenses of existing theoretical frameworks: a competitive supply-demand model on the one hand and an imperfectly competitive minimum wage model on the other.

Motivated by that challenge, I develop a new framework to investigate whether the transformations observed in Brazilian labor markets can be parsimoniously explained by supply, demand, and minimum wage shocks and, if so, to determine what role each of them plays. It features rich worker and firm heterogeneity, a task-based model of production, monopsony power based on idiosyncratic worker preferences, general equilibrium in the market for goods, and free entry of firms. The distinguishing feature of my framework is that it combines the two theoretical perspectives mentioned above by allowing all shocks to affect wages via changes in labor productivity, the dispersion of firm wage premiums, and sorting.

This unified approach provides novel insights into how these shocks affect wage inequality. The first insight is a new explanation for why increases in the supply of skilled labor may have limited effects on the aggregate skill wage premium, or may even widen it (Blundell, Green and Jin, 2021; Carneiro, Liu and Salvanes, 2022). This phenomenon is typically explained using models of endogenous innovation, which creates non-convexities in the aggregate production function (Acemoglu, 1998, 2007). My framework features no such non-convexities. Instead, the aggregate skill premium can rise when the supply shock leads to the creation of skill-intensive, high-wage firms, and the gains in firm premiums for skilled workers reallocated to those new firms outweigh decreases in productivity differentials by skill.[1]

I also show that combining monopsony power, firm heterogeneity, and task-based production can lead to qualitative changes in minimum wage effects. Workers of different skill levels may be complements at high-wage firms that use a broad set of tasks in production but substitutes in low-wage firms specialized in simple tasks. When the minimum wage reallocates unskilled labor from low- to high-wage firms, productivity gaps between skilled and unskilled workers may widen within the destination firms. However, there may be no corresponding compression of the same gaps at the origin firms. As a result, the impact of minimum wages may be negative in the middle of the wage distribution and positive at the top, contrasting with the smooth inequality-reducing effects predicted by competitive task-based models (Teulings, 2000). I also discuss how the reallocation effects and firm wage-setting responses predicted by this framework may affect the interpretability of reduced-form empirical designs commonly employed in minimum wage literature.

With the objective of performing policy counterfactuals, I estimate a parsimonious parameterization of the framework using a simultaneous equation nonlinear least squares procedure. Conceptually, the exercise resembles Katz and Murphy (1992) or Krusell et al. (1999), who use supply/demand models to explain rising wage inequality in the US. I target an array of endogenous outcomes at the region-time level: wage inequality between and within three educational groups, the variance of firm effects, the covariance of firm and worker effects, minimum wage bindingness metrics, and formal employment rates by education type. Although over-identified, the model fits the data well. I interpret the quality of fit as demonstrating that, at least in the Brazilian context, secular trends in wage inequality, the dispersion of firm wage premiums, and sorting can be largely explained by supply, demand, and minimum wage.

Armed with the estimated model, I measure each shock's labor market impact. Consistent

[1] This mechanism is comparable to that of Acemoglu (1999) but differs in that it is not based on search frictions. In addition, firms in my model are large and simultaneously employ many worker types, with within-firm imperfect substitution between skill levels. This generates smooth labor market responses to supply shocks instead of the discrete regime changes predicted by Acemoglu (1999).

with previous work, demand shocks and the rising minimum wage are the leading causes of the decline in wage inequality in Brazil's formal sector. Supply and demand shocks increase measured sorting, the former due to compositional effects and the latter through changes in structural differences between firm types.

I also find significant interactions that are only detectable with a unified framework. When acting in isolation, the minimum wage compresses the lower tail of the wage distribution and reduces the degree of assortative matching, as found in previous work. However, when the minimum wage is accompanied by supply and demand transformations, its inequality-reducing effects are only half as strong, and the sorting effects disappear.

Finally, I conduct two decomposition exercises that demonstrate the quantitative relevance of the new theoretical pathways. In the first exercise, I show that general equilibrium effects—most importantly, increased entry of high-wage firms—amplify the effects of rising schooling achievement on mean log wages by 33%. The second exercise finds that, different from the predictions of Engbom and Moser (2022), the minimum wage has a minimal impact on workers' wages beyond the bottom two deciles of productivity. One reason for that difference is that, whereas wage-posting responses are positive for mid-skill workers in their model, they are negative in mine, as the average marginal productivity of those workers falls following the reallocation of low-skill workers to high-wage firms. The other reason is that, in my model, reallocation effects are limited by local labor market boundaries. In contrast, workers may reallocate nationally in Engbom and Moser (2022) following a minimum wage increase.

The paper proceeds as follows. The next section details how this work builds upon and contributes to different strands of literature. The third section contains a descriptive analysis of the Brazilian data. The fourth section presents the task-based model of production and some of its implications in partial equilibrium. The fifth section describes the complete general equilibrium framework and discusses its predictions concerning the effects of supply, demand, and minimum wage. The sixth section contains the quantitative exercises. The final section concludes with directions for further research.

2 Literature and contribution

This paper's framework can rationalize a large set of empirical facts documented in recent years. It can explain why the contribution of firm wage premiums and sorting to wage inequality may change in the long run (Card, Heining and Kline, 2013; Song et al., 2018; Alvarez et al., 2018). Sorting originates from differences in demand for skills between firms, as documented by Deming and Kahn (2018). Because firms use production functions featuring complementarity between worker types, the framework rationalizes changes in within-firm wages in response to shifts in its internal workforce composition, such as those documented by Jäger and Heining (2022). Minimum wage can cause positive employment effects, reallocation of workers from low- to high-wage firms (Dustmann et al., 2021), spillovers (Fortin, Lemieux and Lloyd, 2021), and changes in how selective firms are when hiring (Butschek, 2022). Minimum wages may also precipitate changes in the types of firms operating in the economy (Rohlin, 2011; Aaronson et al., 2018) and relative consumer prices (Harasztosi and Lindner, 2019). Including all those potential channels lends credibility to the model's quantitative predictions.

On the theoretical side, my task-based model of production builds upon the work of Sattinger (1975) and Teulings (1995), among many others. I derive new formulas for elasticities of complementarity between worker types and provide computationally efficient parameterization. But the core contribution to this literature is characterizing task-based production in an environment with monopsony power and heterogeneous firms. I show that the optimal assignment of workers to tasks may differ between firms and find support for that prediction in the data. I also discuss how substitution patterns differ between firms and why that matters for comparative statics.

The second strand of literature I build upon concerns monopsony models of labor markets based on idiosyncratic worker preferences for firms. I embed the model developed by Card et al. (2018) into a general equilibrium framework with task-based production, firm entry, endogenous participation decisions, and minimum wages. I show how firm heterogeneity in skill intensity and wage premiums emerge from differences in production technologies available to entrepreneurs when they create firms. Within the monopsony literature, my paper resembles the work of Lamadon, Mogstad and Setzler (2022), whose model also generates realistic firm wage premiums and sorting patterns. They allow for worker reallocation across regions and richer forms of firm heterogeneity but do not model within-firm complementarities between worker types, endogenous participation decisions, firm entry, or minimum wages.

More broadly, this paper relates to models that quantify the effects of changing supply of and demand for skills. Within that literature, it is closest to those where supply/demand shocks alter the composition of jobs in the economy. Some work in that tradition, such as Kremer and Maskin (1996) and Lindenlaub (2017), abstract from the role of firm wage premiums. Others, such as Helpman et al. (2017), Shephard and Sidibe (2019), and Lise and Postel-Vinay (2020), feature imperfect competition and firm wage premiums but assume workers are perfect substitutes within firms (or that each firm hires only one worker). In such models, labor market imperfections are the only reason for observing skills dispersion within a firm type. By contrast, firms in my model hire multiple types of workers to benefit from the division of labor, even when labor markets are competitive. Accurate firm-worker sorting patterns are important for capturing the part of the effects of supply/demand shocks that derive from endogenous firm entry and changing prices.

Finally, I describe how my framework differs from quantitative models of minimum wages developed in recent years. Engbom and Moser (2022) build a model with on-the-job search in the style of Burdett and Mortensen (1998). Similar to my study, they estimate their model using Brazilian data and match moments from two-way fixed effects decompositions. Because their model features search frictions, it is better suited to studying job ladders and transitions into and out of unemployment. However, it abstracts from non-wage amenities and assumes perfect substitutability between worker types.

Berger, Herkenhoff and Mongey (2024) and Hurst et al. (2022) build monopsonistic minimum wage models with imperfect substitution across labor types. Berger, Herkenhoff and Mongey (2024) include cross-firm differences in productivity and allow for variation in markdowns depending on the firm size relative to the market. Hurst et al. (2022) abstract from firm heterogeneity but include search frictions and a putty-clay technology that allows them to distinguish between short- and long-run minimum wage effects. They also study how minimum wage can be paired with transfers to achieve redistribution goals.

As a tool for evaluating minimum wages, my framework is unique in four ways. First, substitution patterns between worker types depend on whether they are close or distant in terms of skill and also on the task demands of the firm employing them. Second, it allows for cost pass-throughs and endogenous changes in the composition of firms operating in the economy. Third, it measures how minimum wages interact with educational trends and many types of labor demand shocks. Fourth, it includes an estimation procedure based on regional and time variation. That procedure showcases the model's tractability (because each iteration of the estimation procedure requires solving for equilibria more than 15 thousand times) and its ability to explain cross-sectional variation in features such as the minimum wage spike.

It also allows for measuring how minimum wage effects differ between local labor markets, which may be important in contexts with significant regional heterogeneity.

3 Wage inequality and sorting in Brazil

In this section, I present descriptive statistics that motivate the theoretical framework. I use two data sources. The first is the RAIS (*Relação Anual de Informações Sociais*), a confidential linked employer-employee dataset maintained by the Brazilian Ministry of Labor. Firms are mandated by law to report to the RAIS at the establishment level. The dataset contains information about both the establishment (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).

The other data come from the Brazilian censuses of 1991, 2000, and 2010. From them, I obtain statistics for the overall population, such as the number of adults in each educational group and the proportion of those who hold formal jobs. I also extract from the Census the share of workers in agriculture, manufacturing, or other sectors.[2]

I focus on individuals between 18 and 54 years of age. In the RAIS, I select individuals in that age range who are working in December, having been hired in November or earlier. If a worker has more than one job in the same year, I only keep the highest-paying one.

All the statistics are calculated at the local level. I use the concept of "microregion" as defined by the Brazilian Statistical Bureau (IBGE). Microregions group nearby, economically connected municipalities ("IBGE", 2003). They are commonly used to define local labor market models in Brazil (e.g., Costa, Garred and Pessoa, 2016; Ponczek and Ulyssea, 2021).[3]

I use a local labor markets approach for two reasons. First, regional variation helps identify key parameters of the structural model. Second, local labor markets more closely map theory to empirics. If firm-worker sorting is measured nationally, it will largely reflect ge-

[2]The 1998 outcomes are interpolated using the 1991 and 2000 Censuses. The 2012 outcomes are extrapolated using 2000 and 2010. The interpolations and extrapolations are linear for formal employment rates and sectoral shares, and linear in logs for population counts.

[3]Using data for 2000 and 2010, Dix-Carneiro and Kovak (2017) calculate that less than 5% of workers lived in one region and worked in another. That number, combined with their average size, makes Brazilian microregions analogous to commuting zones in the US. After combining some microregions to ensure that their boundaries remain constant throughout the study period, my base sample features 486 microregions.

Table 1: Evolution of wage inequality measures and sorting

	1998	2012
Panel A: Variances of log wages in base years		
All workers	0.745	0.553
Less than secondary	0.410	0.241
Secondary	0.684	0.355
Tertiary	0.702	0.624
Panel B: Mean log wage gaps in base years		
Secondary / less than secondary	0.498	0.168
Tertiary / secondary	0.965	1.038
Panel C: Variance decomposition using three-year panels		
Variance of worker effects	0.454	0.368
Variance of establishment effects	0.126	0.054
2×Covariance worker, estab. effects	0.105	0.092
Correlation worker, establishment effects	0.224	0.315

Notes: Panels A and B display average wage inequality measures for the base years of 1998 and 2012. Panel C shows the average outcomes of region-specific log wage decompositions based on Equation (1), using the estimator provided by Kline, Saggio and Sølvsten (2018). All numbers are averaged over regions using the total number of formal workers in both base years as weights.

ographical barriers in addition to the supply-demand-minimum wage dynamics emphasized by the framework. I return to this point at the end of the paper when I compare my results to previous work studying the Brazilian case.

The final sample is restricted to microregions with at least 15,000 workers in the RAIS data in 1998 and 2012 and at least 1,000 formal workers in each of the three educational groups defined below.[4] That leaves a set of 151 microregions encompassing 73% of the adult population. Appendix Table D1 presents the consequent sample sizes.

Differing from the pattern in many high-income countries, wage inequality has been downward trending in Brazil since the 1990s. The first two panels in Table 1 report the evolution of several inequality metrics calculated at the microregion level and averaged nationally using total formal employment in both base years as weights (this means that region weights are constant over time). Almost all metrics are declining, some of them dramatically. The one exception is the college premium, which widened in 47 out of 151 regions. Because those regions tend to be larger, the average college premium increased.

[4] My structural estimation procedure requires a low level of measurement error in formal employment rates by educational group and minimum wage bindingness. Those restrictions also yield better estimates of the contribution of firm wage premiums and sorting to local wage inequality.

I gauge the contribution of firm wage premiums and sorting using region-specific variance decompositions based on two-way fixed effects regressions of log wages (henceforth AKM regressions after Abowd, Kramarz and Margolis, 1999). The log wage of worker i in region r at time τ is written as:

$$\log y_{i,r,\tau} = \nu_{i,r} + \psi_{J(i,r,\tau)} + \delta_{r,\tau} + u_{i,r,\tau}$$

where $\nu_{i,r}$ is the worker fixed effect, ψ_j is establishment j's fixed effect, $J(i,r,\tau)$ denotes the establishment employing worker i in region r at time τ, $\delta_{r,\tau}$ is a region-time effect, and $u_{i,r,\tau}$ is a residual. Then, the within-region variance of log wages can be written as follows:

$$\begin{aligned} \text{Var}\left(\log y_{i,r,\tau}|r\right) = {} & \text{Var}\left(\nu_{i,r}|r\right) + \text{Var}\left(\psi_{J(i,r,\tau)}|r\right) + 2\text{Cov}\left(\nu_{i,r}, \psi_{J(i,r,\tau)}|r\right) \\ & + \text{Var}\left(\delta_{r,\tau}|r\right) + 2\text{Cov}\left(\nu_{i,r} + \psi_{J(i,r,\tau)}, \delta_{r,\tau}|r\right) + \text{Var}\left(u_{i,r,\tau}|r\right) \end{aligned} \quad (1)$$

If wages differ substantially across establishments for similar workers, the variance of establishment effects may be large, adding to overall wage dispersion. If high-wage workers are more likely to work at high-wage establishments, then the first covariance term will be positive, further boosting inequality. Based on this logic, the correlation between establishment and worker fixed effects is often used as a simple measure of labor market sorting.

Estimating the variance decomposition (1) is not trivial. I use the method developed by Kline, Saggio and Sølvsten (2018) (henceforth KSS), which is not subject to the limited mobility bias discussed by Andrews et al. (2008). I run the KSS model separately for each microregion and period, using three-year panels centered on either 1998 or 2012. Because that procedure requires a leave-one-out connected set, small establishments are under-represented in that sample. Appendix D.2 provides details about the procedure.

Decomposition results appear in Panel C of Table 1. The variances of both worker and establishment effects decline on average. Together, they account for more than 80% of the average decline in the variance of log wages. In comparison, the covariance term displays a minor average reduction, explaining less than 7% of the inequality fall. Thus, it accounts for a larger share of the variance of log wages in 2012. The measured correlations between worker and establishment effects increase in most microregions (104 out of 151).[5,6]

[5] The KSS estimate of the *correlation* between worker and establishment effects is not guaranteed to be unbiased. In the structural estimation exercise, I target the unbiased covariance estimates rather than the correlations.

[6] Alvarez et al. (2018) and Engbom and Moser (2022) also find that establishment effects explain a signifi-

Table 2: Trends in schooling achievement and minimum wage bindingness

	1998	2012
Panel A: Share of adults by education group		
Less than secondary	0.699	0.493
Secondary	0.229	0.383
Tertiary	0.072	0.124
Panel B: Minimum wage bindingness		
Log minimum wage minus mean log wage	-1.418	-0.922
Log minimum wage minus log median wage	-1.220	-0.719
Share up to log minimum wage + 0.3	0.086	0.212

Notes: All numbers are averaged over regions using the total number of formal workers in both base years as weights.

The interpretability of AKM decompositions relies on categorizing establishments as high- or low-wage. However, in many economic models of sorting, including this paper's, wages are not log-additive in worker and establishment components: Some establishments may pay some worker types more and other worker types less. Still, indirect inference can be used to extract identifying information from the AKM decomposition. I employ this strategy in this paper.

Now I consider the potential explanations for the falling inequality in Brazil. The most conspicuous are increased educational achievement and rising minimum wages. Table 2 shows the magnitude of those shocks. Panel A displays the average share of adults in each of three educational groups: less than secondary (that is, a level of achievement lower than completing high school, or between zero and ten years of schooling), secondary (combining complete high school and college dropouts, or between 11 and 14 years or schooling); and tertiary (complete college or more). The pattern is striking: In the span of 14 years, the share of adults completing high school or further education increases by 20 percentage points (a 68% increase). This represents the outcome of educational reforms and policies traceable to the 1980s, including minimum government expenditure requirements on education, construction of schools, cash transfers conditional on school enrollment, and vouchers for tertiary education.

Panel B shows that the minimum wage became more binding over the study period. The

cant fraction of the decline in wage inequality in Brazil. However, they find that the covariance term also falls, such that there is no increase in measured sorting. The key difference between my approach and theirs is that whereas my decompositions are performed at the local labor market level, they use national models. National-level sorting can fall if, for example, gains in educational achievement are stronger in areas with low-wage firms.

Brazilian national minimum wage increased by 66 log points in real terms (93.7%) between December 1998 and December 2012, which increased the "bite" of the minimum wage into the wage distribution regardless of the bindingness metric used. The apparent compression of wage distribution is shown in Appendix Figure D1.

A third factor emphasized in the Brazilian case is labor demand shocks associated with international trade. During the study period, Brazilian regions were still adapting to the trade liberalization of the early 1990s, which, according to Dix-Carneiro and Kovak (2017), had long-lasting impacts. During the 2000s, the "rise of China" led to significant changes in terms of trade. Costa, Garred and Pessoa (2016) study that shock and also find evidence of differential labor market impacts at the microregion level. Trade liberalization seemingly benefitted skilled workers, while the commodities boom benefited unskilled workers.

These transformations are not easily explained using existing quantitative frameworks. One could be tempted to conclude that rising education and demand for commodities increase the relative productivity of unskilled workers, while the minimum wage further reduces markdowns for unskilled workers and reallocates some of them to high-wage firms (Engbom and Moser, 2022). But that simple story does account for the fact that sorting is rising. Indeed, the minimum wage effects just described would imply *decreases* in sorting. That is the motivation for building a framework where supply and demand factors affect wages through not only worker productivity but also firm wage premiums and assortative matching.

4 The task-based production function

Task-based models of comparative advantage are increasingly used to model wage inequality. Acemoglu and Autor (2011) show that 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.

In this section, I demonstrate an additional advantage of the task-based approach: It allows for intuitive, tractable, and parsimonious modeling of firm heterogeneity in both competitive and imperfectly competitive labor markets. All proofs appear in Appendix A.

4.1 Setup, definitions, and the assignment problem

Workers are characterized by their labor type $h \in \{1, \ldots, H\}$ and the amount of labor efficiency units they can supply, $\varepsilon \in \mathbb{R}_{>0}$. They use their labor to produce tasks that are indexed by complexity $x \in \mathbb{R}_{>0}$.[7] Although all labor types are perfect substitutes in the production of any particular task, 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 an example with two workers. Alice, characterized by h, ε, can produce $\varepsilon e_h(x)$ tasks of complexity x if she uses all of her time working on those tasks. Bob (h', ε'), who belongs to a lower type ($h' < h$), can still produce more of those tasks than Alice, provided his quantity of efficiency units is sufficiently high (i.e., if $\varepsilon' > \varepsilon e_h(x)/e_{h'}(x)$). But Alice has a comparative advantage: Moving toward more complex tasks increases her productivity relative to Bob's.

Each good, indexed by $g = 1, \ldots, G$, 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).*

Consider a firm trying to produce good g after hiring $\boldsymbol{l}$ efficiency units of labor in the labor market. I assume that tasks cannot be traded; they must be produced internally. Firms can assign its employees to work on specific tasks by choosing assignment functions $m_h : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ (assumed to be right-continuous). They can be interpreted in the following way: given any arbitrary range of tasks $[\underline{x}, \bar{x})$, the total quantity of efficiency units of labor type h used in the production of those tasks is $\int_{\underline{x}}^{\bar{x}} m_h(x)dx$.

Given this structure, the production function is defined as the maximum quantity of goods that the firm can produce by assigning workers to tasks optimally, given that it has access to a fixed quantity of efficiency units of labor of each type $\boldsymbol{l} = \{l_1, \ldots, l_H\}$:

[7] In the quantitative exercises, worker skill is mapped to educational achievement, meaning more complex tasks should be interpreted as those better performed by formally educated workers. The assumption of a single complexity dimension is maintained throughout. Quantitative models using multi-dimensional skills and tasks include Lindenlaub (2017) and Lise and Postel-Vinay (2020).

Definition 3. *The **task-based production function** is given by*

$$
\begin{aligned}
f(\boldsymbol{l};b_g) = \max_{q\in\mathbb{R}_{\geq 0},\{m_h(\cdot)\}_{h=1}^{H}} \quad & q \\
s.t. \quad & qb_g(x) = \sum_h m_h(x)e_h(x) \quad \forall x \in \mathbb{R}_{>0} \\
& l_h \geq \int_0^\infty m_h(x)dx \quad \forall \in \{1,\dots,H\}
\end{aligned}
$$

and is defined for all $\boldsymbol{l} \in \mathbb{R}_{\geq 0}^{H-1} \times \mathbb{R}_{>0}$ and blueprints b_g.

This definition assumes a positive amount of labor of type H, which is not restrictive for my applications. See Appendix B.1 for a brief discussion.

The next subsections characterize the properties of this production function under different labor market structures. Before arriving there, I present a general result on the optimal assignment of workers to tasks:

Lemma 1 (Optimal allocation is assortative)**.** *For every combination of inputs $\boldsymbol{l}, b_g$, there exists a unique set of $H-1$ complexity thresholds $\bar{x}_1 < \cdots < \bar{x}_{H-1}$ that defines the range of tasks performed by each worker type in an optimal allocation: $m_h(x) > 0 \Longleftrightarrow x \in [\bar{x}_{h-1}, \bar{x}_h)$, with $\bar{x}_0 = 0$ and $\bar{x}_H = \infty$. Thresholds satisfy:*

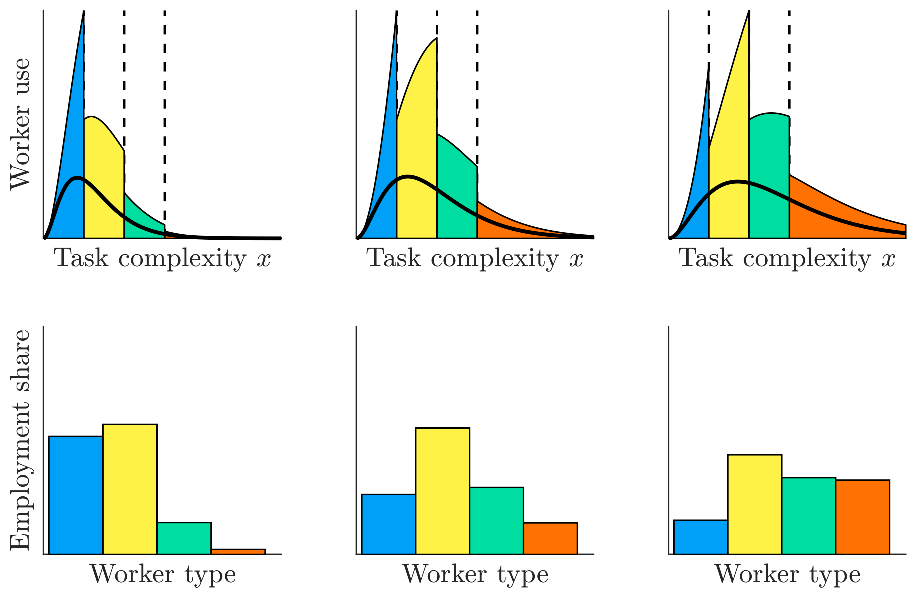
$$\frac{e_{h+1}(\bar{x}_h)}{e_h(\bar{x}_h)} = \frac{f_{h+1}}{f_h} \quad h \in \{1,\dots,H-1\} \tag{2}$$

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

Lower types specialize in low-complexity tasks and vice-versa. Equation (2) means that the shadow cost of using neighboring worker types is equalized at the task that separates them. This result is the starting point for obtaining compensated labor demands, as I describe in the following subsection.[8]

[8] In general, the task-based production function and its derivatives do not have simple closed-form representations. To evaluate output and marginal productivities as a function of labor inputs, one must first solve the system of H compensated labor demand equations (3) on q and the $H-1$ thresholds. Next, use equation (2) 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



4.2 Compensated labor demand in competitive labor markets

To study the implications of task-based production for labor demand, I start with a partial equilibrium analysis. Consider an individual firm, which produces good g, attempting to minimize labor costs given a production target q. The labor market is competitive, such that unit costs per efficiency unit of each labor type are constants $\boldsymbol{w} = \{w_1, \ldots, w_H\}$.

Optimality requires that marginal product ratios equal wage ratios. Then, from Equation (2):

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

Because the left-hand side is strictly increasing in $\bar{x}_h$, this expression pins all task thresholds as functions of wage ratios and comparative advantage functions. That is, thresholds are strictly increasing functions $\bar{x}_h(w_{h+1}/w_h)$. This renders the compensated labor demand as follows:

$$l_h(q, b_g, \boldsymbol{w}) = q \int_{\bar{x}_{h-1}(w_h/w_{h-1})}^{\bar{x}_h(w_{h+1}/w_h)} \frac{b_g(x)}{e_h(x)} dx \tag{3}$$

Now suppose that different firms produce different goods in this partial equilibrium, competitive environment. Because neither efficiency functions nor labor costs are firm-specific, all firms choose the same task thresholds.

Figure 1 illustrates how blueprints determine demand for skills. The graphs at the top show the discussed compensated labor demand integral for three different blueprints. The heavy, continuous line is the blueprint. Thus, as we move from left to right, the blueprints become more intensive in high-complexity tasks. The vertical dashed lines are the thresholds defining the ranges of tasks assigned to each worker type. The colored areas represent the labor demand integrals from Equation 3. The bottom panels show corresponding factor intensities as histograms.

Due to the infinite-dimensional blueprints and efficiency functions, the task-based structure might appear exceedingly flexible at first glance. Proposition 1 extends the results of Teulings (2005) and shows that, on the contrary, there are strong constraints on substitution patterns.[9]

Proposition 1 (Curvature)**.** *The task-based production function is concave, features constant returns to scale, and is twice continuously differentiable with strictly positive first derivatives. Appendix A provides formulas for elasticities of complementarity and substitution.*

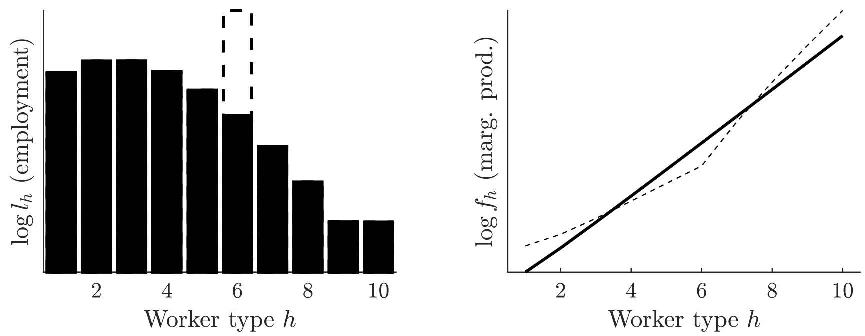
Corollary 1 (Distance-dependent complementarity)**.** *For a fixed h, the partial elasticity of complementarity between that type and another type h' is strictly increasing in h' for $h' \geq h$ and strictly decreasing in h' for $h' \leq h$.*

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

Now consider a firm that employs workers of many types. If that firm hires an additional worker, how does that affect the marginal productivity of existing employees? The answer depends on the elasticities of complementarity, the objects of interest in the corollary. They are thus of central importance in this paper since one of its goals is to evaluate how workforce composition changes affect wage distribution. For example, if the number of college-

[9] 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. In an application of assignment models to optimal taxation, Ales, Kurnaz and Sleet (2015) derive elasticities of substitution in a model of production where output is CES in tasks, instead of Leontief .

Figure 2: Distance-dependent complementarity



educated people in the economy grows, firms will employ more college workers. That will change marginal products of labor of all worker types, with corresponding effects on wages.

The distance-dependent substitution property means that, in the task-based production function, worker pairs that are "close" in terms of their group h are substitutes. At the same time, those with very different h are complements. Figure 2 illustrates that pattern. The black bars on the left panel show the initial quantities of labor at a particular firm. At those employment levels, marginal products correspond to the solid line on the right panel. Suppose there is an exogenous increase in employment of workers of type $h = 6$ at that firm. In addition to an increase in output, that will cause changes in marginal products of labor as the firm reassigns workers to tasks. The new marginal products are shown as the dashed line in the right panel. They fall for workers of types four through seven (that is, they are substitutes with the new type six workers) and rise for low and high types further away (complements).

Teulings (2000) shows that distance-dependent complementarity is also useful for modeling minimum wage spillovers, that is, changes in the distribution of wages at quantiles where the minimum wage does not bind. If a minimum wage causes disemployment of low-skilled workers, then the logic of Figure 2 implies that marginal products—and hence wages—should increase for workers close to the minimum. The core contribution of Teulings (2000) is to show that, differing from a "canonical" CES approach, a task-based model with many worker types can explain realistic levels of spillovers even when the disemployment effects are small.[10] My framework differs from Teulings (2000) in that I allow for firm heterogeneity and imperfect competition, which I start discussing in the next subsection.

[10] Teulings and van Rens (2008) derives a sufficient statistic that can be used to compare the degree of substitution across worker types in different models. For some combinations of shocks and outcomes of interest, task-based models and the canonical model can produce very similar predictions. But this is typically not true for minimum wage shocks.

4.3 Labor demand in a monopsonistic labor market

Suppose that firms have wage-setting power. Each firm j posts a prices per efficiency unit w_{hj} for each type h. At that posted wage, it is able to attract a quantity of labor equal to $l_{hj} = l_h(w_{hj}) = L_h \cdot \left(\frac{w_{hj}}{\omega_h}\right)^{\beta}$.[11] The core implication of upward-sloping supply curves to the firm is that the more intensely a factor is used, the higher its marginal cost. Thus, if firms differ in their skill intensity because they use different blueprints, their marginal product ratios differ. Equation (2) implies that their optimal assignments will also differ:

Lemma 2 (Differences in skill intensity, monopsony, and task assignment)**.** *Consider a partial equilibrium environment where firms have wage-setting power as described above. Suppose that the optimal labor choices of two firms indexed by $j \in \{1,2\}$ satisfy $\frac{l_{h+1,2}}{l_{h,2}} > \frac{l_{h+1,1}}{l_{h,1}}$ for some h. Then, $\bar{x}_{h,2} > \bar{x}_{h,1}$ (where $\bar{x}_{h,j}$ denotes the task threshold $\bar{x}_h$ at firm j).*

When a worker moves from one firm to another that is more skill-intensive, they will be assigned to more complex tasks. I test that prediction in Subsection 6.1. Lemma 2 also shows that wage-setting power may generate productive mismatch, similarly to how search frictions introduce mismatch in Teulings and Gautier (2004).[12]

Another implication of wage-setting power and task-based production is that an aggregate shock may produce different responses at different firms:

Proposition 2 (Complementarity patterns may differ between firms)**.** *Consider a partial equilibrium model with three worker types ($H = 3$), two goods with positive prices, and wage-setting power as described above. Good $g = 1$ has a degenerate blueprint requiring a unit measure of low-complexity tasks, $x = 0$. Good $g = 2$ has a regular blueprint. Then:*

1. *Firms producing either good employ workers of all types h.*
2. *If there is an increase in L_1 but all other supply parameters remain unchanged, posted wages do not change for firms producing good $g = 1$. But for firms producing good $g = 2$, all posted wage gaps $w_{h+1,j}/w_{h,j}$ become larger.*

The first part of this proposition exemplifies the production mismatch mentioned above. In a competitive market, firms that only need tasks $x = 0$ would not hire workers of high types.

[11] This expression is consistent with the general equilibrium model described in Section 5, in a special case with no minimum wage. β is the firm-level elasticity of labor supply, L_h is the aggregate supply of labor of type h, and ω_h is a sufficient statistic for labor demand by other firms in the market.

[12] Specifically, a planner that maximizes aggregate output given any vector of prices for goods will choose a different assignment of workers to tasks, compared to the monopsonistic allocation.

However, given isoelastic firm-level supply curves, it is sensible to hire at least a few such workers because, at sufficiently low employment levels, they become very cheap. More generally, there is less employment specialization under monopsony, although we should still expect firms demanding more complex tasks to be more skill-intensive.

The second part of Proposition 2 highlights a key feature of my framework: Firms differ in terms of not only demand for skill but also substitution patterns. For firms using the regular blueprint, $g = 2$, an increase in the aggregate supply of labor type $h = 1$ widens all within-firm skill wage differentials. This reflects distance-dependent complementarity. For firms producing the low-complexity good $g = 1$, posted wages do not change: The shock increases employment of workers of type $h = 1$ but has no other impact .

The degenerate blueprint used in the Proposition is very stylized, but it serves to illustrate a more general pattern. Suppose that the blueprints are those shown in Figure 1. Then, the intuition from Proposition 2 still applies: We should expect wage responses to be more muted for firms using the blueprint in the left panel because workers are closer substitutes in those firms. In Subsection 5.6, I show that this property has implications for the equilibrium effects of minimum wages.

4.4 Exponential-Gamma parameterization

In the quantitative exercises, I employ a parameterization with exponential efficiency functions and blueprints shaped like the density of a Gamma distribution:

$$e_h(x) = \exp(\alpha_h x) \qquad b_g(x) = \frac{x^{\kappa-1}}{\Gamma(\kappa)\theta_g{}^{\kappa}} \exp\left(-\frac{x}{\theta_g}\right)$$

The coefficients α_h are increasing and determine the degree of comparative advantage of a labor type. The parameter θ_g relates to average task complexity. All else being equal, goods with higher θ_g require more complex tasks and thus have a higher demand for skills. Given that the shape parameter κ is assumed to be common across firms, goods with higher θ_g also have more diffuse task requirements, meaning that workers are more likely to be complements at those firms.

Appendix C presents the mapping between marginal productivity gaps and task thresholds for a generalized version of this parametrization, as well as formulas for compensated labor demand integrals in terms of incomplete gamma functions or power series. These formulas do not require numerical integration, making them computationally efficient.

4.5 Discussion

In this section, I introduced the task-based production function, explained its properties in competitive and imperfectly competitive labor markets, and showed that it admits a parsimonious and computationally efficient parameterization. Before proceeding to the general equilibrium model, I briefly discuss why this formulation is appropriate for this paper.

As a first comparison point, consider models where workers are perfect substitutes within firms (e.g., Bagger and Lentz, 2019; Engbom and Moser, 2022; Lamadon, Mogstad and Setzler, 2022). These models can tractably deal with two-sided heterogeneity but have two limitations that are consequential for the research questions explored in this paper. First, marginal products of labor within the firm are exogenous, ruling out decreasing marginal returns to specific types of labor following an increase in their aggregate supply. Second, these models predict that, without search frictions or idiosyncratic worker preferences for firms, every firm would hire workers of a single work type. Since one of my goals is to rationalize changes in sorting patterns over time, it is helpful to use a framework where imperfect assortative matching can arise from the division of labor within firms, instead of solely from labor market imperfections.

Another alternative approach would be to use a CES production function (e.g., Berger, Herkenhoff and Mongey, 2024). While these models feature imperfect substitution within firms, the elasticities of complementarity are the same for all pairs of worker types, in stark contrast to the distance-dependent substitution property. One can introduce richer substitution patterns by adding a nested structure at the cost of increasing the number of estimated parameters. Nevertheless, even in the nested case, there is a fundamental difference between that approach and the task-based production function; in the former, substitutability is intrinsic to the worker types, while in the latter, it depends on endogenous assignment to tasks.[13]

For a concrete example of why cross-firm differences in substitution patterns could be relevant, consider a trained engineer choosing between a technical job at a manufacturing firm

[13] In principle, one could estimate nested CES models where elasticities vary by firm type level. However, the number of parameters to be estimate can grow very quickly, leading to identification challenges. Eeckhout and Pinheiro (2014) and Trottner (2019) also model large firms with multiple jobs but with common elasticities of substitution across all pairs of worker types. Herkenhoff et al. (2018) allows for search frictions and within-firm complementarities, but firms may only employ up to two workers. Models of hierarchical firms in the tradition of Garicano (2000), Garicano and Rossi-Hansberg (2006), and Antràs, Garicano and Rossi-Hansberg (2006) imply a within-firm division of labor. However, the modeling of costly information transmission within firms reduces their tractability. My production structure can be viewed as a hierarchical firm model without that cost and without the restriction that hierarchies need to be pyramidal.

or a low-level managerial position at a local grocery store. The wage is significantly higher at the manufacturing job, but that particular engineer may be indifferent between the two options because the commute to the grocery store is much shorter. At the manufacturing job, the engineer is assigned to high-complexity tasks that very few high school workers could execute. That makes high school workers poor substitutes for the engineer at that firm. In contrast, if employed at the grocery store, the engineer can be easily replaced by workers without a college education, because the marginal task that separates the two types at that firm—perhaps managing worker schedules—has much lower complexity (Lemma 2).

In Section 6, I provide reduced-form evidence of differences in the assignment of workers to tasks consistent with Lemma 2. I also find that a structural model employing the task-based production function can successfully rationalize the transformations observed in the Brazilian economy despite its parsimony.

5 Markets and wages

This section builds a general equilibrium 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 presents an important property of the model: Goods encapsulate firm heterogeneity in skill intensity and wages. The third subsection describes firm wage differentials. The remaining subsections discuss comparative statics with respect to supply, demand, and minimum wage shocks.

Although the model is static, Appendix C.3 discusses a simple dynamic extension that can be used to simulate moments that require a panel dimension. Unless otherwise noted, all parameters are assumed to be strictly positive.

5.1 Factors, goods, technology, and preferences

There are two factors of production. The first is labor. The total number of workers of type h is denoted by N_h, and the distribution of efficiency units ε within group h is continuous with density $r_h(\cdot)$ and support over the positive real line. The second factor is an entry input used to create firms. The total stock of the entry input is normalized to one, and it is fully owned by a representative entrepreneur.

The economy features G firm-produced goods. Firms can only produce one of the goods, and the decision of which good the firm produces is made when the firm is created. The entry cost per firm, F_g, depends on the chosen good. The entrepreneur's action is to choose the number of firms J_g, conditional on the entry input constraint $\sum_g F_g J_g \leq 1$.

Firm-produced goods are sold in competitive markets at prices p_g. Consumers (workers or the representative entrepreneur) combine them into the final consumption good using a constant elasticity of substitution (CES) aggregator:

$$c = z\left[\sum_{g=1}^{G} \gamma_g Q_g^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{1-\sigma}}$$

where z is a productivity parameter and γ_g is a taste shifter. The elasticity of substitution σ may depend on the interpretation of goods in the model: lower for different sectors, higher for different varieties within sectors, or close to infinity for different production technologies used to produce the same good. A large σ can also be an approximation for a small open economy where all goods are tradable.[14] I use the corresponding price index as the numeraire in this economy: $P \equiv \left[\sum_{g=1}^{G} \gamma_g^{\sigma} p_g^{1-\sigma}\right]^{\frac{1}{1-\sigma}} = 1$.

Alternatively, workers that choose not to work for any firm can produce the final good via home production. A worker of type (h, ε) can produce $c = \varepsilon z_{0,h}$ units for its own consumption. The productivity parameters $z_{0,h}$ are intended to capture the value of outside options such as informal employment, self-employment, and government transfers to unemployed adults. The quantitative section allows those parameters to vary flexibly at the region, time, and education levels.

The entrepreneur's preferences are monotonic in the final good. Worker preferences depend on not only consumption but also where they are employed:

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

where i denotes worker identity, c is its final good consumption, and j denotes the employment choice. Home production is denoted by $j = 0$. Employment in any of the firms is denoted by $j = 1, \ldots, J$ where $J = \sum_g J_g$. The η_{ij} parameters denote idiosyncratic preferences of workers towards their employment options. The importance of those components

[14] In the empirical exercise, I do not map goods to industries because the within-industry dimension is important. In many contexts, changes in inequality happen within industries (see Card, Heining and Kline, 2013; Song et al., 2018). The validation exercise in Subsection 6.1 suggests substantial task heterogeneity within finely defined sectors.

relative to consumption is regulated by λ.

The idiosyncratic preference components capture match-specific features, such as distance to the workplace, personal relationships with the manager or other coworkers, and how much they like staying at home for $j = 0$. The full vector of idiosyncratic preferences for a worker is drawn from the following cumulative distribution function:

$$CDF\left(\{\eta_{ij}\}_{j=0}^{J}\right) = \exp\left\{-\exp\left(-\eta_{i0}\right) - \left[\sum_{j=1}^{J} \exp\left(-\eta_{ij} \cdot \frac{\beta}{\lambda}\right)\right]^{\frac{\lambda}{\beta}}\right\}$$

This is a nested logit, with all firms included in one nest and home production in another. The parameter $\beta \geq \lambda$ denotes the correlation in preferences between firms. In the following section, I demonstrate that λ pins down the macro elasticity of labor supply to all firms, while β determines the firm-level elasticity of labor supply.

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

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 efficiency units of 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. Because the model has no variation in hours worked, earnings and hourly wages are interchangeable. And because workers with low ε might have a marginal product of labor below $\underline{y}$ at some firms, I allow firms to reject workers with productivity below some minimum value $\underline{\varepsilon}_{hj}$.

5.2.1 Firm-level labor supply and labor costs

The timing of the labor market is as follows. First, all firms post rejection cutoffs $\underline{\varepsilon}_{hj}$ and earnings schedules $y_{hj}(\varepsilon) : [\underline{\varepsilon}_{hj}, \infty) \to [\underline{y}, \infty)$. Second, workers observe all $\underline{\varepsilon}_{hj}$ and $y_{hj}(\varepsilon)$ and choose their employment option j. Third, firms observe (h, ε) of workers who applied to them (but not idiosyncratic preference shifters η_{ij}) and hire those with $\varepsilon \geq \underline{\varepsilon}_{hj}$. Finally, production occurs and hired workers are paid. 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 option j. It can be written as:

$$
\begin{aligned}
V_{ih}(\varepsilon,j) =& \exp\left(\lambda\log\left(\varepsilon z_{0,h}\right)+\eta_{ij}\right)^{\frac{1}{\lambda}} && \text{if } j=0\\
V_{ih}(\varepsilon,j) =& \mathbf{1}\left\{\varepsilon\geq\underline{\varepsilon}_{hj}\right\}\exp\left(\lambda\log y_{hj}(\varepsilon)+\eta_{ij}\right)^{\frac{1}{\lambda}} && \text{if } j\geq 1
\end{aligned}
$$

Given the distribution of η_{ij}, the probability of a worker (h,ε) choosing a particular option j is given by:

$$
\begin{aligned}
\Pr\left(0=\underset{j'\in\{0,1,\ldots,J\}}{\arg\max}\;V_{ih}(\varepsilon,j')\right) &= \frac{(\varepsilon z_{0,h})^{\lambda}}{(\varepsilon z_{0,h})^{\lambda}+\omega_{\varepsilon,h}^{\lambda}}\\
\Pr\left(j=\underset{j'\in\{0,1,\ldots,J\}}{\arg\max}\;V_{ih}(\varepsilon,j')\right) &= \frac{\omega_{\varepsilon,h}^{\lambda}}{(\varepsilon z_{0,h})^{\lambda}+\omega_{\varepsilon,h}^{\lambda}}\times\left(\frac{\mathbf{1}\{\varepsilon\geq\underline{\varepsilon}_{hj}\}y_{hj}(\varepsilon)}{\omega_{\varepsilon,h}}\right)^{\beta} \qquad \text{for } j\geq 1\\
\text{where } \omega_{\varepsilon,h} &= \left(\sum_{j=1}^{J}\mathbf{1}\{\varepsilon\geq\underline{\varepsilon}_{hj}\}y_{hj}(\varepsilon)^{\beta}\right)^{\frac{1}{\beta}}
\end{aligned}
$$

The "inclusive value" $\omega_h(\varepsilon)$ is a measure of demand for skills coming from firms. The employment rate for workers with productivity (h,ε) is given by a logit formula comparing that value against those workers' efficacy at home production. The macro elasticity of labor supply with respect to $\omega_h(\varepsilon)$ is given by λ multiplied by the share of those workers in home production.

As in Card et al. (2018), I assume that firms ignore their own contribution to $\omega_h(\varepsilon)$, an approximation that is adequate when firms are small relative to the size of the labor market. Under that assumption, each firm's labor supply elasticity for workers of a particular type (h,ε) is given by β as long as earnings are above the minimum wage.

The number of workers choosing a particular firm, the resulting supply of efficiency units of labor, and total labor costs are increasing in posted earnings and decreasing in rejection cutoffs:

$$
n_h(y_{hj},\underline{\varepsilon}_{hj}) = N_h\int_{\underline{\varepsilon}_{hj}}^{\infty}\frac{\omega_{\varepsilon,h}^{\lambda}}{(\varepsilon z_{0,h})^{\lambda}+\omega_{\varepsilon,h}^{\lambda}}\left(\frac{y_{hj}(\varepsilon)}{\omega_h(\varepsilon)}\right)^{\beta} r_h(\varepsilon)d\varepsilon \tag{4}
$$

$$
l_h(y_{hj},\underline{\varepsilon}_{hj}) = N_h\int_{\underline{\varepsilon}_{hj}}^{\infty}\frac{\omega_{\varepsilon,h}^{\lambda}}{(\varepsilon z_{0,h})^{\lambda}+\omega_{\varepsilon,h}^{\lambda}}\left(\frac{y_{hj}(\varepsilon)}{\omega_h(\varepsilon)}\right)^{\beta} \varepsilon r_h(\varepsilon)d\varepsilon \tag{5}
$$

$$C_h(y_{hj}, \underline{\varepsilon}_{hj}) = N_h \int_{\underline{\varepsilon}_{hj}}^{\infty} \frac{\omega_{\varepsilon,h}^{\lambda}}{(\varepsilon z_{0,h})^{\lambda} + \omega_{\varepsilon,h}^{\lambda}} \frac{y_{hj}(\varepsilon)^{\beta+1}}{\omega_h(\varepsilon)^{\beta}} r_h(\varepsilon) d\varepsilon \tag{6}$$

5.2.2 Problem of the firm

Firms maximize profit by choosing posted earnings schedules and rejection cutoffs:

$$\pi_j = \max_{\boldsymbol{y}_j, \underline{\boldsymbol{\epsilon}}_j} p_g f\left(\boldsymbol{l}(\boldsymbol{y}_j, \underline{\boldsymbol{\epsilon}}_j), b_g\right) - \sum_{h=1}^{H} C_h(y_{hj}, \underline{\varepsilon}_{hj})$$

The following Lemma shows that this problem has intuitive solutions:

Lemma 3. *Firms producing the same good g choose the same earnings schedules and rejection criteria, denoted by y_{hg} and $\underline{\varepsilon}_{hg}$. Optimal earnings schedules have the form $y_{hg}(\varepsilon) = \max\{w_{hg}\varepsilon, \underline{y}\}$. The following first-order conditions define prices per efficiency unit w_{hg} and hiring thresholds:*

$$p_g f_h\left(\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g), b_g\right) \frac{\beta}{\beta+1} = w_{hg} \qquad h = 1, \ldots, H \tag{7}$$

$$p_g f_h\left(\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g), b_g\right) \underline{\varepsilon}_{hg} = \underline{y} \qquad h = 1, \ldots, H \tag{8}$$

Equation 7 defines optimal prices per efficiency unit $w_{h,g}$ as constant markdowns of their marginal revenue products, a common result in monopsony models with a constant elasticity of labor supply to the firm. Equation 8 is the first-order condition on the rejection cutoffs. A lower cutoff brings in additional workers with $\varepsilon = \underline{\varepsilon}_{hj}$, each of which increases revenues by $p_g f_h \underline{\varepsilon}_{hj}$. When firms choose thresholds optimally, that additional revenue equals the minimum wage $\underline{y}$, which is the cost of labor at that margin.

5.2.3 Firm creation and equilibrium

A finite σ engenders positive firm creation for all goods for two reasons. First, with the CES functional form for the consumption aggregator, marginal utilities for each good are unbounded as consumption moves to zero, enabling arbitrarily high equilibrium prices even if entry and marginal costs are large. Second, firms are guaranteed to record positive profits due to the constant markdowns of log wages.[15]

[15] I assume that the number of firms in every market is sufficiently large that we can ignore the integer constraint in optimal firm creation. Accordingly, I treat J_g as a continuous variable.

An equilibrium of this model is defined by vectors of aggregate consumption $\{Q_g\}_{g=1}^G$, firm entry $\{J_g\}_{g=1}^G$, choices by representative firms $\{\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g\}_{g=1}^G$, and prices $\{p_g\}_{g=1}^G$ such that:

1. Markets for firm-produced goods clear:

$$Q_g = \gamma_g^\sigma p_g^{-\sigma} I = J_g f\left(\boldsymbol{l}(\boldsymbol{y}_g, \underline{\boldsymbol{\epsilon}}_g), b_g\right) \quad \forall g \tag{9}$$
$$\text{where } I = \sum_{g=1}^G J_g \left[\pi_g + \sum_{h=1}^H C_h(w_{hg}, \underline{\varepsilon}_{hg})\right] = \sum_{g=1}^G J_g p_g f\left(\boldsymbol{l}(\boldsymbol{y}_g, \underline{\boldsymbol{\epsilon}}_g), b_g\right)$$

2. For all g, firm choices solve the first-order conditions (7) and (8).

3. Firm creation is optimal and feasible:

$$\frac{\pi_g}{F_g} = \frac{\pi_{g'}}{F_{g'}} \; \forall (g, g') \text{ and } \sum_g J_g F_g = 1 \tag{10}$$

Labor market clearing is embedded in the firm-level labor supply curve. Appendix C presents an efficient numerical algorithm to solve for equilibrium given a set of parameters.

5.3 Firm wage premiums

The following proposition describes how equilibrium wages vary between firms:

Proposition 3. *1. If $b_g(x) = b(x)/z_g$ for scalars $z_1, \ldots, z_G$ and F_g is the same for all firm-produced goods, then there are no firm wage premiums:*

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

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

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

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

3. *If there is no minimum wage and there are firm types g, g' and worker types h' h such*

that $\ell_{h'g'}/\ell_{hg'} > \ell_{h'g}/\ell_{hg}$ (that is, good g' is relatively more intensive in h'), then:

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

The first part of Proposition 3 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 entry costs. Notably, differences in physical productivity across goods (z_g) or in taste shifters (γ_g) are insufficient to generate wage differentials between firms. This is because if entry costs are the same, differences in physical productivity or tastes lead to additional entry and reduced marginal utility of consumption of the good with greater productivity, up to the point where the marginal revenue product of labor is equalized across firms.

The second part highlights the role of entry costs in generating wage differences across firms. Optimal firm creation implies that all else being equal, firms producing goods with higher entry costs need to operate at a larger scale. To hire more workers, these firms must post higher wages. At equilibrium, prices for those goods will also be higher, such that worker earnings are proportional to the marginal revenue product of labor.

The third part of Proposition 3 shows how skill intensity heterogeneity generates differential wage gaps across firms. Firms using some factors more intensively than others must pay a relative premium for that factor. This model's inability to simultaneously generate log-additive wages and assortative matching echoes some results in the literature on labor market sorting, such as those in Eeckhout and Kircher (2011). However, it is possible for skill-intensive firms to pay all workers a positive wage premium if those firms have high entry costs, such that the model can still include "high-wage firms" as a meaningful concept.

Appendix B.2 adds vertical differentiation of non-wage amenities to the model. Those extra parameters can be used to match firm sizes without affecting the rest of the theory.

5.4 Supply shocks

It is possible that labor supply, labor demand, and the minimum wage evolve in concert, making the economy more productive while leaving wage distribution unchanged (see Proposition 7 in Appendix B.3). However, if there are imbalances in this race, relative prices for goods and labor may change.

I start with supply shocks. To focus on what general equilibrium and firm entry add to the model, the following proposition abstracts from within-firm complementarities by assuming that each good only requires one task (i.e., workers are perfect substitutes within firms):

Proposition 4 (Supply shock and reallocation)**.** *Consider an economy with two comparative advantage types, two goods, full employment ($z_{0,h} = 0$), and no minimum wage. Assume both goods $g = 1,2$ have degenerate blueprints such that each unit of output requires a unit measure of tasks of complexity x_g, with $x_2 > x_1$. Then:*

$$\frac{d\left(\frac{s_{2,1}\log w_{2,1}+s_{2,2}\log w_{2,2}}{s_{1,1}\log w_{1,1}+s_{1,2}\log w_{1,2}}\right)}{d\log\left(L_2/L_1\right)} = \frac{d\log\left(\frac{p_2}{p_1}\right)}{d\log\left(\frac{L_2}{L_1}\right)}\left[\left(s_{2,2}-s_{1,2}\right)+\left(\beta+1-\sigma\right)\left(s_{2,1}s_{2,2}\log\frac{w_{2,2}}{w_{2,1}}-s_{1,1}s_{1,2}\log\frac{w_{1,2}}{w_{1,1}}\right)\right]$$

where $s_{h,g}$ denotes the share of efficiency units of labor of type h employed by firms producing good g, and $\frac{d\log(p_2/p_1)}{d\log(L_2/L_1)} < 0$.

Corollary 2. *For any set of parameters satisfying the conditions of Proposition 4, there exists a number $\bar{\beta}$, such that by changing β to $\beta' > \bar{\beta}$ and F_g to $F_g' = F_g\frac{\beta+1}{\beta'+1}$, the effect of rising supply on the mean log wage gap is negative.*

The effect of increased supply of skills on the aggregate skill wage premium has two components. The first is the direct effect of the supply shock on marginal products of labor via prices. That component is always negative because positive supply shocks reduce p_2/p_1 and $s_{2,2} > s_{1,2}$. The second component is the reallocation of labor across firms paying different wage premiums. If the reallocation effect is positive and sufficiently large, the aggregate skill premium can widen in response to the supply shock.

The strength of the reallocation effect depends on the magnitude of firm wage premiums, initial sorting patterns, and the elasticities β and σ. Those elasticities also determine the direction of net reallocation flows. As mentioned, the supply shock reduces p_2/p_1. Because that price change passes on to wages, individual firms producing $g = 1$ can attract more workers, with elasticity β. However, the reduction in p_2/p_1 also shifts consumption toward the second good, increasing relative firm entry J_2/J_1. If $\sigma > \beta + 1$, the second effect wins, and there is net reallocation to firms producing $g = 2$.

Corollary 2 emphasizes how imperfect competition is essential to the result that positive supply shocks may widen the aggregate skill premium. By moving the parameters close to the competitive limit ($\beta \to \infty$, $F_g \to 0$), supply shocks are guaranteed to compress the skill wage premium. This result exemplifies how Proposition 4 differs fundamentally from the directed technical change channel emphasized by Acemoglu (1998, 2007).

In a more general environment with non-degenerate blueprints, the expression for the change in the aggregate skill wage premium would include additional terms deriving from imperfect substitution within firms. The total impact of supply shocks on the aggregate skill premium may be positive even in these cases, as the quantitative analysis demonstrates.

5.5 Demand shocks

There are three ways to model skill-biased demand shocks in this economy. The first is by changing blueprints in a way that increases the demand for complex tasks. Analogously to the monotone comparative statics used by Costinot and Vogel (2010), this should increase all wage gaps w_{h+1}/w_h in a competitive economy with a single good.

The second form of skill-biased shock is an increase in demand for skill-intensive goods, which may represent improvements in the quality of those goods or trade shocks affecting demand for goods that are more skill intensive.[16]

Proposition 5 (Demand for goods and returns to skill)**.** *Consider a competitive version of this economy ($\beta \to \infty$, $F_g = 0$) with full employment ($z_{0,h} = 0$), two goods, and no minimum wage. Assume $b_2(x)/b_1(x)$ is increasing in x (good $g = 2$ is more intensive in high-complexity tasks). Then, an increase in γ_2/γ_1 increases all wage gaps w_{h+1}/w_h.*

Proposition 5 has a more general implication: If other shocks change aggregate consumption patterns in the direction of more or less complex tasks, there may be secondary effects on skill wage premiums. I return to this point in the discussion of general equilibrium effects of minimum wage policies.

The third type of skill-biased demand shock is a reduction in relative entry costs F_2/F_1 when good 2 is more skill intensive. It reallocates labor towards more complex tasks by reducing relative prices p_2/p_1 and increasing relative entry J_2/J_1. As Proposition 3 describes, that shock also reduces the magnitude of firm wage premiums when skill-intensive firms are also high-wage. The net effect on inequality measures is ambiguous.

[16] In the latter interpretation, Proposition 5 is in the same spirit as the classic result of Stolper and Samuelson (1941). At the limit $\sigma \to \infty$, the model is equivalent to a small open economy with prices $p_2/p_1 = \gamma_2/\gamma_1$.

In the empirical exercise, I allow for regional and time differences in these three dimensions of labor demand.

5.6 Minimum wage

In this section, I explain how minimum wage affects the model economy. This discussion serves two purposes. First, it includes some novel insights that may be of value to economists who study minimum wages, including potential pitfalls to avoid in reduced-form empirical studies. Second, it clarifies what channels are accounted for in the simulations presented later in the paper. Appendix B.4 discusses causal pathways not included in this framework and explains why their omission may not be consequential in the Brazilian context.

5.6.1 Channel 1: "monopsony" (mechanical wage increases, disemployment, positive employment effects, and reallocation)

Suppose that, starting from an initial equilibrium, the minimum wage increases to $\underline{y}' > \underline{y}$. I update earnings schedules from $y_{h,g}(\varepsilon)$ to $y'_{h,g}(\varepsilon) = \max\left\{y_{h,g}(\varepsilon), \underline{y}'\right\}$. I also update the minimum hiring thresholds to account for the fact that, keeping marginal products of labor constant, some low-skilled workers become unprofitable under the new minimum. Then, I allow workers to change their employment options based on the new earnings schedules and hiring thresholds. All other equilibrium variables remain unchanged.

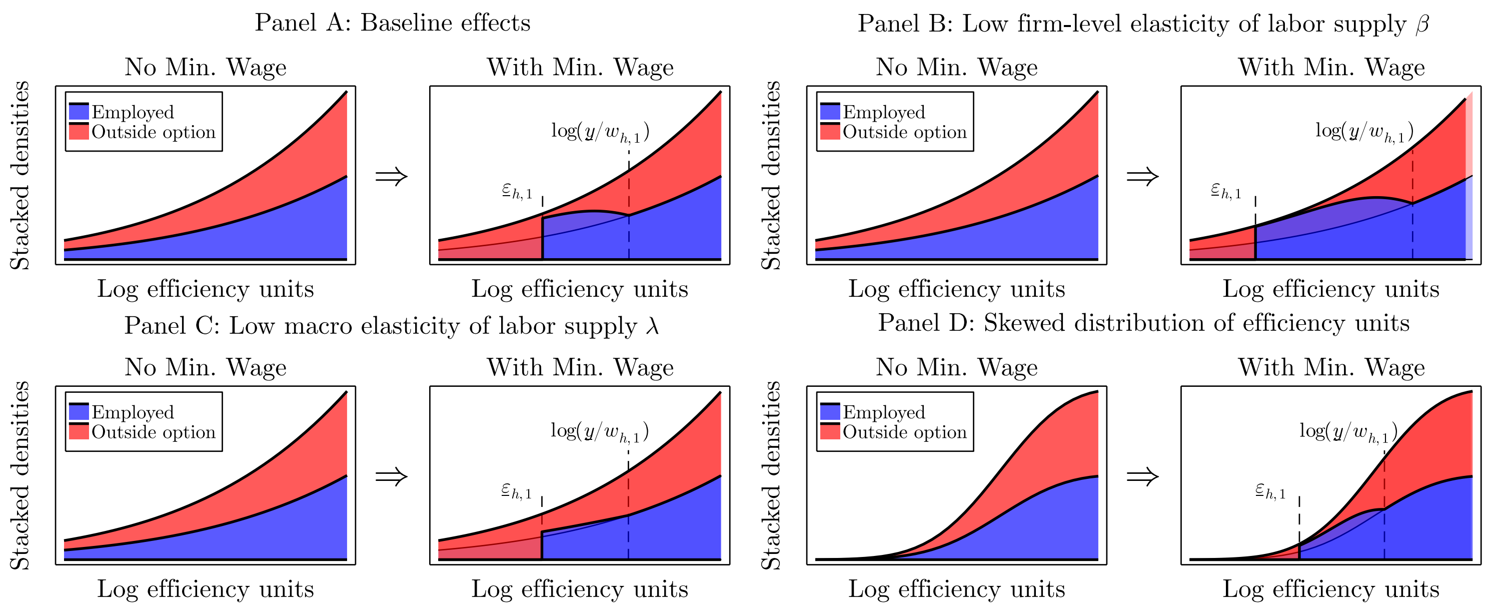
Figure 3 illustrates the counterfactual employment choices in a model with a single good. The graphs show the mass of workers by employment choice and worker productivity, providing a close-up view of the left tail of the productivity distribution.

Consider the baseline scenario in Panel A. For workers with $\varepsilon > \log(\underline{y}/w_{h,1})$, employment options remain unchanged, as do their optimal choices. Because workers with $\varepsilon < \underline{\varepsilon}_{hj}$ are no longer employable at formal firms, all of them move to their outside options. Finally, workers with $\varepsilon \in \left[\underline{\varepsilon}_{h,1}, \log(\underline{y}/w_{h,1})\right]$ are the ones receiving a mechanical "wage boost" at formal firms. If they choose to work there, they earn exactly the minimum wage. Thus, the blue mass of workers in that interval corresponds to the minimum wage spike.

Positive employment effects of minimum wage arise from workers in that middle interval. One important takeaway is that, even if the *total* change in employment is non-negative, the minimum wage may still cause disemployment for very low-productivity workers.[17]

[17] The inability of minimum wages to correct monopsony-induced underemployment for all worker types

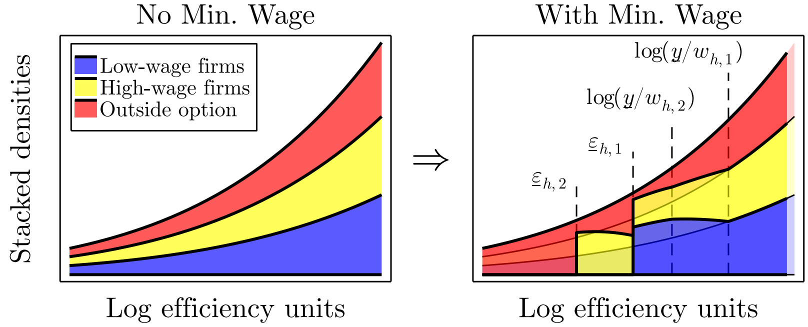
Figure 3: Minimum wage effects with a single firm type



Notes: This figure shows the "monopsony" effects of minimum wages on worker employment options in a context with a single firm type. The graph shows the lower tail of the distribution of efficiency units ε for a particular worker group h. Each vertical slice of the graphs shows the stacked densities of employed and non-employed workers for each ε, such that the total height corresponds to the total density $N_h r_h(\varepsilon)$. Workers with $\varepsilon < \underline{\varepsilon}_{h,1}$ lose their formal jobs. Workers with $\varepsilon \in [\underline{\varepsilon}_{h,1}, \log(\underline{y}/w_{h,1})]$ earn exactly the minimum wage. These are the workers for whom formal wages are increasing mechanically and for whom we should see positive employment effects.

Panel B shows that a lower β increases the range of productivity levels where the minimum wage may cause positive employment effects. **Panel C** shows that if λ is small, positive employment effects are also likely to be small. **Panel D** shows that estimates of minimum wage effects depend crucially on assumptions about the shape of the underlying productivity distribution.

Figure 4: Minimum wage effects with two firm types



Notes: This figure shows the impact of minimum wage on worker employment options when there are two firm-produced goods (equivalently, two firm types). The "high-wage firms", $g = 2$, have higher revenue productivity and can afford to hire workers with lower ε after the introduction of the minimum wage. This generates reallocation from low- to high-wage firms for workers with $\varepsilon \in \left[\underline{\varepsilon}_{h,2}, \underline{\varepsilon}_{h,1}\right]$. The neighborhood around $\log(\underline{y}/w_{h,2})$ may feature the opposite type of reallocation (from high-wage to low-wage firms).

Panels B, C, and D in Figure 3 illustrate how minimum wage effects depend on the firm-level elasticity of labor supply, the aggregate elasticity of labor supply, and the shape of the underlying productivity distribution. In the quantitative section, I calibrate one of those elasticities, estimate the other, and allow for flexible distributions of worker ability within educational groups.

Figure 4 resembles Figure 3 except that it shows a scenario with two goods. The initial equilibrium has workers evenly split between low-wage firms ($g = 1$), high-wage firms ($g = 2$), and home production. The high-wage firms have higher revenue productivity and can afford to hire workers with lower ε after the introduction of the minimum wage. This generates reallocation from low- to high-wage firms for workers with $\varepsilon \in \left[\underline{\varepsilon}_{h,2}, \underline{\varepsilon}_{h,1}\right]$, a pattern that is the model analog of the empirical results in Dustmann et al. (2021).

The model also predicts some reallocation from high- to low-wage firms, especially for workers with $\varepsilon \approx \log(\underline{y}/w_{h,2})$. This is because the minimum wage does not affect their wage at high-wage firms but makes low-wage firms more attractive. This result has implications for empirical studies of minimum wages that compare workers based on their initial wage. Even if there are no strategic wage-posting responses and no general equilibrium effects, workers earning more than the new minimum may still be affected by the minimum wage, precluding them from being a valid control group.

simultaneously was first noted by Stigler (1946).

5.6.2 Channel 2: Wage-posting responses and within-firm returns to skill

To quantify the role of this channel, I calculate a partial equilibrium where prices p_g and firm creation J_g are kept constant following the increase in the minimum wage. Firms can reoptimize earnings schedules $y_{h,g}(\varepsilon)$ and hiring thresholds $\underline{\varepsilon}_{hj}$. Then, I compare the simulated outcomes of this partial equilibrium to the baseline equilibrium and subtract the contribution of the "Monopsony" channel described in the previous subsection.

Why would firms choose different posted earnings following the introduction of a minimum wage? Holding earnings schedules constant, disemployment and reallocation effects imply changes in factor shares within firms. Because the production function is concave, marginal products of labor also change. Then, firms need to adjust w_{hg} to ensure that they are proportional to the marginal revenue products of labor.

The combination of monopsony power, firm heterogeneity, and task-based production generates novel predictions regarding minimum wage effects compared to both competitive task-based models (Teulings, 2000) and monopsonistic models of minimum wage (Engbom and Moser, 2022). Suppose that there are two firms with blueprints that are equally skill-intensive but with $F_2 \gg F_1$. A newly introduced minimum wage may bind for low-h workers at firms producing good $g = 1$ but not good $g = 2$. This may generate reallocation of low-h labor. Within-firm skill premiums could fall at firms producing $g = 1$ and widen at firms producing $g = 2$.

Perhaps a more typical scenario is one where low-wage firms are also low-skill. Suppose that good $g = 1$ has a blueprint fully concentrated in tasks of complexity $x = 0$, as described in Proposition 2. Then, internal skill premiums at firms producing that good will not respond to the minimum wage. Reallocation will still widen skill premiums at firms producing $g = 2$. Combining those effects, it is possible that wage changes induced by the minimum wage are ultimately less progressive, especially for middle-skill workers. The quantitative section shows that this channel is responsible for the negative wage effects of minimum wages for workers in the middle of Brazil's productivity distribution.[18]

This theoretical prediction also has implications for empirical minimum wage designs. Some papers compare firms in the same region based on the proportion of their workers that earns

[18]This channel may not always cause reductions in real wages for low-wage workers at high-wage firms. As an example, in a scenario where minimum wage causes strong mechanical increases in wages at low-wage firms but not much disemployment, the resulting increase in $\omega_{h,\varepsilon}$ can lead to positive wage effects at high-wage firms.

below the new value of the minimum wage. The preceding discussion demonstrates that the minimum wage may also affect high-wage firms, albeit in a fundamentally different way. This means that those high wage firms may not constitute an appropriate control group. Note that because this mechanism does not depend on changes in prices and firm entry, arguing that there are no general equilibrium effects is insufficient to validate the "fraction affected" design.

5.6.3 Channel 3: General equilibrium

Finally, I account for minimum wage-induced changes in prices p_g and firm creation J_g. The strength of those equilibrium effects depends crucially on the elasticity of substitution in consumption σ. To make the analysis concrete, consider a scenario with two goods in which skill-intensive firms are also high-wage.

Start with the Leontief case, $\sigma = 0$. Minimum wages reduce profits at low-wage firms by compressing their markdowns. In general equilibrium, falling profits at low-wage firms induce an increase in J_2/J_1. In the Leontief world, $Q_2/Q_1 = (J_2/J_1) \cdot (q_2/q_1)$ is constant, so q_2/q_1 must fall. That change in relative scale can only be achieved by compressing firm wage premiums because minimum-wage-induced reallocation tends to increase q_2/q_1. Consequently, the cost ratio falls, as does the price ratio p_2/p_1.

Now consider the other extreme with perfect substitution: $\sigma \to \infty$. Relative prices are now invariant, $p_2/p_1 = \gamma_2/\gamma_1$, and changes in relative profits induce changes in firm entry. There is more reallocation of labor from low- to high-wage firms because there is no need for offsetting entry with scale responses to keep quantities constant.

Comparing both scenarios, we should expect minimum wages to be less progressive if σ is large. With a low σ, low- and medium-skilled workers benefit from the increase in the relative price for low-skill goods even if the minimum wage does not mechanically increases their wages. An increase in p_1 also attenuates disemployment effects. With a large σ, firm-creation responses increase aggregate demand for complex tasks, benefiting skilled workers.

6 Quantitative exercises

I now apply the framework to the data. The first subsection uses reduced-form regressions to test basic implications of the theory. The second subsection structurally estimates a parametric model of the Brazilian economy. The third subsection contains counterfactual exercises.

Table 3: Validation of the task-based production function.

	Non-routine cognitive task content of occupation				Log wage
	Estab. average	Worker level			
	(1)	(2)	(3)	(4)	(5)
Mean schooling in establishment	0.07921 (0.00049)				
Own schooling		0.06304 (0.00159)			
Mean schooling of coworkers in establishment			0.00663 (0.00077)	0.00343 (0.00086)	
Own $\times$ mean schooling of coworkers in estab.					0.00162 (0.00045)
Sample	Estabs.	All workers	Movers	Movers	All workers
Years used	1997	1997	1997, ’99	1997, ’99	1997, ’99
Microregion-time fixed effects	✓		✓	✓	✓
Establishment fixed effects		✓			✓
Sector fixed effects				✓	
Worker fixed effects			✓	✓	✓
r2	0.26216	0.40172	0.84463	0.85033	0.95789
N	93,606	11,551,108	2,673,660	2,673,659	14,996,848

Notes: RAIS data, largest connected set in each of the 151 selected microregions. Standard errors (in parenthesis) are robust in Column (1), clustered at the establishment level in Column (2), and two-way clustered at the worker and establishment levels in the other columns. The standard deviation of the task content variable is approximately one.

6.1 Firm heterogeneity, task assignment, and wage premiums

In this subsection, I test four implications of the model: (i) skill-intensive firms have more demand for complex tasks (Figure 1); (ii) within firms, more skilled workers are assigned to more complex tasks (Lemma 1); (iii) with monopsony power, workers moving to more skill-intensive firms are reallocated to more complex tasks (Lemma 2); and (iv) wage gaps between high- and low-skill firms should be larger for skilled workers (Proposition 3).

To test these predictions, I need proxies for worker skill and task complexity. Skill is measured by years of schooling. Appendix Table D2 reports results for an alternative measure. For task complexity, I use the non-routine analytical task content of Brazilian occupations created by de Sousa (2020). That measure reflects whether O*NET survey respondents believe that their occupation requires mathematical reasoning and was created following the methodology in Deming (2017).[19]

Columns (1) and (2) in Table 3 test the first two predictions using data for 1997. Column (1) reports a firm-level regression of the establishment's average task complexity on the average years of schooling of that establishment's employees. Consistent with the theory, I find a positive relationship. Column (2) is a worker-level regression of the task content of the worker's occupation on that worker's schooling, controlling for firm fixed effects. The positive coefficient confirms the prediction for within-firm assignment.

Next, I use worker transitions between establishments to test the third prediction. Specifically, I regress the analytical task content of the worker's occupation on mean schooling of other workers in the same establishment, controlling for worker fixed effects. That regression uses data from 1997 and 1999 and only includes movers. Column (3) demonstrates that the estimate is positive and significant, although the correlation is weaker than in Column (2). Workers moving to firms with more educated colleagues tend to be assigned to more analytical occupations, consistent with differences in optimal assignment across firms in imperfectly competitive environments.

I also investigate whether changes in assignment are driven by workers moving between sectors. Column (4) shows results for a specification similar to Column (3) but with sector fixed effects.[20] I find that the coefficient falls by about half but remains highly significant. This suggests sizable within-sector variation in skill intensity and task content of occupations, consistent with the interpretation that goods in the model might represent differentiated varieties or technologies within industries.

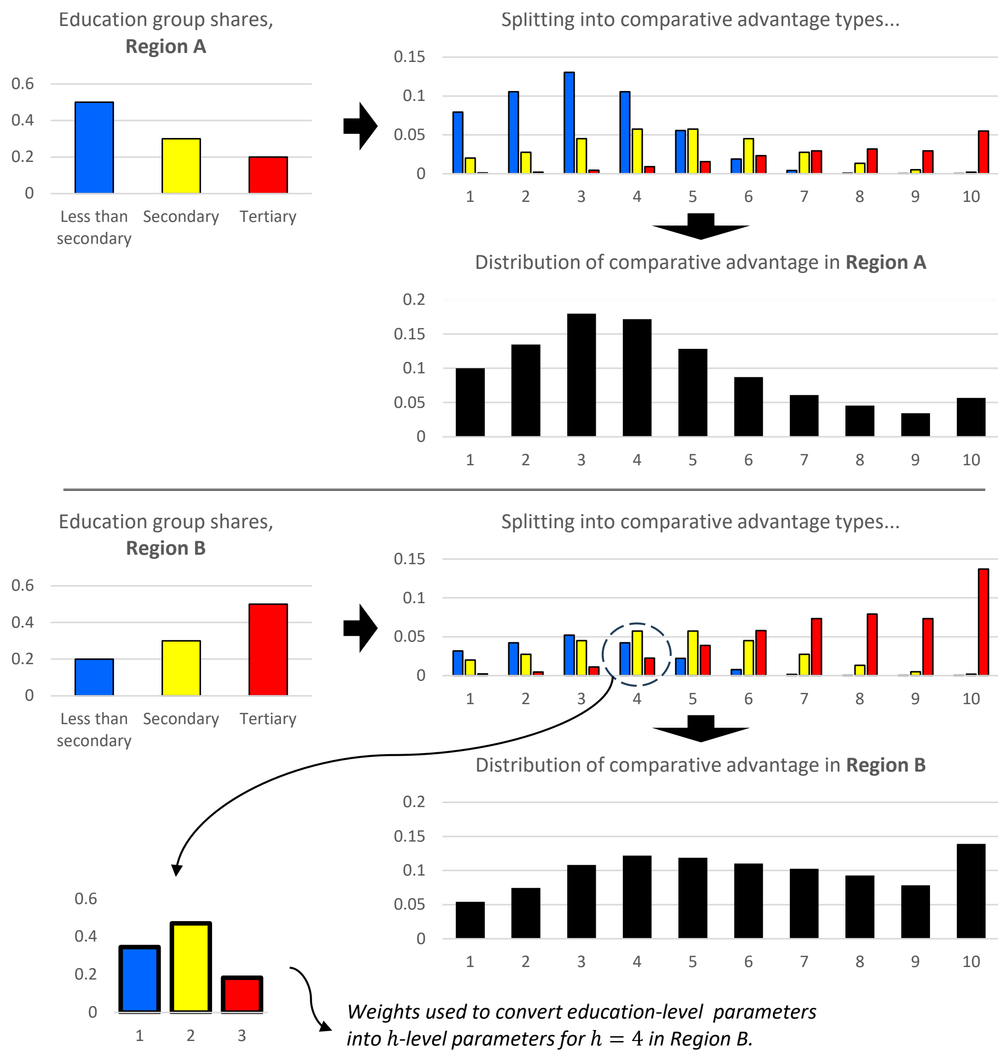
Finally, Column (5) tests the fourth prediction, again using panel data. It reports a regression of log wage on worker fixed effects, firm fixed effects, and the interaction between a worker's years of schooling and the average schooling of coworkers in their workplace. I find a positive, statistically significant estimate, consistent with the theory. In Appendix Table D2, I demonstrate that this result is not a mechanical consequence of the minimum wage.

---

[19]The O*NET survey asks workers in the US about their jobs, including skill requirements and the degree of automation in the occupation. Deming (2017) describes how that survey is collected and processed to produce data that describe each occupation as a combination of tasks of varying intensities. de Sousa (2020) links SOC occupation codes with occupation codes in the RAIS data before calculating the task content of occupations using O*NET data and the procedures in Deming (2017).

[20]There are 560 "CNAE10" sectors in the regression sample. 507 include at least 100 movers.

Figure 5: Parameterization of worker types in the structural model



6.2 Structural estimation

6.2.1 Parameterization

Each microregion-time combination is treated as an isolated economy, indexed by (r,t). For each, the general equilibrium model specifies a mapping from the estimated parameters to simulated endogenous outcomes. The estimation procedure minimizes deviations between the observed endogenous outcomes and their simulated values. In this subsection, I describe the parameterization of the model. In the following subsection, I formalize the data-generating process and discuss identification, estimation, and inference.

Elasticities of labor supply: I set $\beta = 4$, based on recent studies that find that find elasticities of labor supply to the firm between three and five.[21] The λ parameter, which determines the macro elasticity of labor supply, is assumed to be common across regions and periods and is estimated jointly with the other parameters in the model.

Worker types and outside options: I continue to sort workers into three educational groups, as in the descriptive section of the paper, to measure some of the endogenous outcomes such as the returns to college. But in model, I assume that there are ten latent worker types, $H = 10$. Thus, productivity differences between workers in the same educational group can come from both dispersion in efficiency units and differences in comparative advantage. The relative importance of one vis-à-vis the other is an important determinant of how the variance of log wages within each educational group responds to exogenous shocks.

Figure 5 illustrates how I convert observed education shares for groups $\hat{h} \in \{1,2,3\}$ into distributions of workers along the 10 latent types $h \in \{1,\ldots,10\}$ for two hypothetical regions A and B. I assume that the distribution of h-types within education group, $\Pr\left(h|\hat{h}\right)$, is the same in all regions and periods. Using those shares, which are determined by four estimated parameters, I first calculate the quantity of workers of each h coming from educational group $\hat{h}$ (this step corresponds to the horizontal black arrow). Then, I sum the mass of workers in each h coming from all educational groups (the vertical black arrow).

When discussing the effects of minimum wages in the model, I emphasized how they depend crucially on the parametric distribution of efficiency units. I assume that, conditional on (h,r,t), ε has a Skew Normal distribution with mean zero and dispersion parameter $S_{h,r,t}$. The skewness parameter χ captures the possibility that the lower tail of productivity has leaner or fatter tails than what would be implied by a Normal distribution.

Finally, I specify how the 10 h-types in each region-time differ in the dispersion parameters $S_{h,r,t}$ and outside options $z_{0,h,r,t}$. Appendix D.5 details how they are determined by their education-level counterparts $\hat{S}_{\hat{h}}$ and $\hat{z}_{0,\hat{h},r,t}$. The $\hat{z}_{0,\hat{h},r,t}$ are, in turn, the product of region-time, region-education, and education-time parameters to richly capture other factors that determine heterogeneity in formal employment rates, such as intensity of enforcement of labor regulations. To go from $\hat{h}$-level parameters to h-level parameters in each region, I use the probabilities $\Pr\left(\hat{h}|h,r,t\right)$, as illustrated in the bottom part of Figure 5.

[21] I discuss why I chose to calibrate this parameter, rather than estimating it, in Appendix D.4. Card et al. (2018) use the same value in their calibration, which implies a markdown of 20%. Lamadon, Mogstad and Setzler (2022) estimate firm-level elasticities of labor supply between 6.02 and 6.52, corresponding to markdowns around 14%. Kroft et al. (2023) find elasticities between 3.5 and 4.1. Berger, Herkenhoff and Mongey (2022) find average firm-level markdowns of 22% or 11%, depending on whether the average is weighted by payroll.

Labor demand: There are $G = 2$ goods in each region.[22] Blueprints follow the Exponential-gamma parameterization presented in Subsection 4.4. There are five demand-side parameters that vary at the region-time level. The first is the productivity parameter $z_{r,t}$, which is unrestricted. The others are blueprint complexities $\theta_{g=1,r,t}$ and $\theta_{g=2,r,t}$, relative entry costs $F_{2,r,t}/F_{1,r,t}$, and relative consumer preference $\gamma_{2,r,t}/\gamma_{1,r,t}$.[23] They are determined by region-time-specific covariates as follows:

$$\begin{aligned} D^d_{r,t} = &\delta^{d,t}_0 + \delta^{d,t}_1 ShareHighSchool_{r,1998} + \delta^{d,t}_2 ShareCollege_{r,1998} \\ &+ \delta^{d,t}_3 ShareAgriculture_{r,1998} + \delta^{d,t}_4 ShareManufacturing_{r,1998} \\ &+ \delta^{d,t}_5 \left(log(min.wage) - meanLogWage\right)_{r,t} \end{aligned} \tag{11}$$

for $d \in \{\theta, F, \gamma\}$, where:

$$D^\theta_{r,t} = \log \theta_{2,r,t} \qquad D^F_{r,t} = \log\left(\frac{F_{2,r,t}}{F_{1,r,t}}\right) \qquad D^\gamma_{r,t} = \log\left(\frac{\gamma_{2,r,t}}{1-\gamma_{2,r,t}}\right)$$

and:

$$\theta_{1,r,t} = \theta_{2,r,t}\tilde{\theta}_t$$

That is: blueprint complexities, entry costs, and consumer tastes vary between regions depending on a set of covariates. The gap in blueprint complexity between the two goods, $\tilde{\theta}_t$, is assumed to be the same in all regions within each period. This formulation yields a total of 38 demand-side parameters to be estimated: 36 of the $\delta^{d,t}_i$ form plus the two $\tilde{\theta}_t$.

This formulation is essential for unbiased estimation of the effects of supply, demand, and minimum wage shocks on the wage distribution. To see why, suppose that the college share and college wage premium are strongly correlated in the data. That relationship could reflect a positive causal link from education to the skill premium. Alternatively, it could be that firms in more educated parts of the country use more skill-intensive technologies for reasons outside of the model. The latter possibility would be captured by $\delta^{\theta,t}_2 > 0$. Without this term in Equation (11), the model would rule out that alternative explanation by assumption, such that the causal effects of education would be inferred from a simple comparison of

[22] I used two goods to keep the model as simple as possible. There is no technical impediment to using a larger number of goods. The estimator proposed by Bonhomme, Lamadon and Manresa (2019) may be helpful in higher-dimensional applications.

[23] Entry costs only matter in relative terms because I do not target average firm sizes. For computational purposes, I set $F_{2,r,t} = 1$. Consumer preferences also only matter in relative terms, given that outside option parameters are fully flexible. As such, I normalize $\gamma_{1,r,t} + \gamma_{2,r,t} = 1$.

mean outcomes between high- and low-educated regions. Similar arguments can be made regarding identifying the wage distribution effects of the minimum wage.[24]

Because of regional convergence, allowing the $\delta_i^{d,t}$ to be time-specific is essential in the Brazilian context. Regions that are less educated as of 1998 may "catch up" in terms of technology, such that the relationship between initial education and current demand parameters may become weaker over time. In this flexible formulation, the variation used to identify the wage effects of labor supply shocks is the *change* in educational shares, and the relevant comparison group is regions that had similar levels of education in 1998.

Initial shares of the workforce engaged in agriculture and manufacturing are used as additional predictors of labor demand shocks. This approach is analogous to the "shift-share designs" used to evaluate the consequences of labor demand shocks on employment and wages, where the "shift" component is effectively a dummy for $t = 2012$.

Summing up, there are 52 estimated parameters common across regions: eight defining worker types; two outside option shifters at the education-time level; one determining whether regional shocks to outside options have weaker effects on college workers (see Appendix D.5); 38 determinants of local demand; blueprint shape κ; and the elasticities σ and λ. In addition, there are six region-specific parameters: two time-specific TFP's determining minimum wage bindingness and four formal employment shifters (three region-education base levels plus one region-time shock).

6.2.2 The data-generating process and identification

The data-generating process is:

$$\boldsymbol{Y}_r = a(\boldsymbol{Z}_r, \boldsymbol{\theta}_0^G, \boldsymbol{\theta}_r^R) + \boldsymbol{u}_r \qquad r \in \{1, \ldots, R\}$$

where $\boldsymbol{Y}_r$ is a vector of 26 endogenous outcomes (13 for each of the two periods, 1998 and 2012). It includes inequality measures within and between groups, variance components from the AKM decomposition, formal employment rates, and minimum wage bindingness

[24] In the model, the bindingness of the minimum wage is mostly determined by TFP parameter $z_{r,t}$. One may wonder why Equation 11 is expressed in terms of the endogenous "effective minimum wage", instead of the exogenous $z_{r,t}$. The reason is computational tractability. As the next section demonstrates, the estimation procedure requires "inverting" region-specific parameters, including $z_{r,t}$, from the observed moments. If the expressions for the demand parameters were written in terms of $z_{r,t}$, that inversion procedure would be impossible. The chosen formulation solves the endogeneity concern—possible correlation between TFP and other demand parameters—while keeping the model tractable.

measures. The full list corresponds to the non-italicized moments in Table 5. The vector $\boldsymbol{Z}_r$ includes all region-specific covariates. The 51 general parameters are represented by the $\boldsymbol{\theta}_0^G$ vector (where the subscript denotes the true value). Finally, $\boldsymbol{\theta}_r^R$ represents the six region-specific parameters. The function $a(\cdot)$ simulates the endogenous outcomes using the model parameters implied by $(\boldsymbol{Z}_r, \boldsymbol{\theta}_0^G, \boldsymbol{\theta}_r^R)$. The residuals $\boldsymbol{u}_r$ represent both mean-zero shocks at the regional level and sampling variation in the measurement of endogenous variables.[25]

Let $PB(\boldsymbol{Y})$ be a function that selects the following six moments from $\boldsymbol{Y}$: formal employment rates for each of the educational groups in $t = 1998$, the formal employment rate for high school workers in $t = 2012$, and minimum wage bindingness in both years (defined as log minimum wage minus mean log wage). These endogenous outcomes are used to "invert" the region-specific parameters given a guess of the other parameters, as formalized in the following identification assumptions:

Assumption 1 (Exogeneity)**.** $E[\boldsymbol{u}_r | Z_r, \boldsymbol{\theta}_r^R] = \mathbf{0}_{26\times 1}$.

Assumption 2 (Independence between microregions)**.** *If* $r \neq r'$, *then* $E[\boldsymbol{u}_r \boldsymbol{u}_{r'}'] = \mathbf{0}_{26\times 26}$.

Assumption 3 (Correct specification of employment and bindingness)**.** $PB(\boldsymbol{u}_r) = \mathbf{0}_{6\times 1}\ \forall r$.

Assumption 4 (Invertibility of outside options and TFP)**.** *For all r and all allowable $\boldsymbol{\theta}^G$, there is a function $\hat{\boldsymbol{\theta}}^R\left(\cdot | \boldsymbol{Z}_r, \boldsymbol{\theta}^G\right)$ such that:* $\boldsymbol{Y} = a\left(\boldsymbol{Z}_r, \boldsymbol{\theta}^G, \boldsymbol{\theta}^R\right) \Leftrightarrow \boldsymbol{\theta}^R = \hat{\boldsymbol{\theta}}^R\left(PB(\boldsymbol{Y}) | \boldsymbol{Z}_r, \boldsymbol{\theta}^G\right)$.

Assumption 5 (Rank condition)**.** *Define:*

$$\tilde{a}\left([\boldsymbol{Z}_r', PB(\mathbf{Y}_r)']', \boldsymbol{\theta}^G\right) = a\left(\boldsymbol{Z}_r, \boldsymbol{\theta}^G, \hat{\boldsymbol{\theta}}_r^R\left(PB(\mathbf{Y}_r) | \boldsymbol{Z}_r, \boldsymbol{\theta}^G\right)\right)$$

Denote the 51×1 *gradient of the o-eth endogenous outcome of the* $\tilde{a}(\cdot)$ *function, with respect*

[25] One component of the error terms $\boldsymbol{u}_r$ are measurement errors in the left-hand side variables arising from the fact that the sample used to calculate the AKM decomposition moments differs from that used for the other moments. That is because the former includes more years and restricts attention to the leave-one-out connected set. If those errors are not mean zero, or if they correlate with the covariates or region-level parameters, then they may introduce model misspecification. To address this issue, I renormalize the AKM decomposition outcomes such that the total variance of log wages in the leave-one-out set is the same as in the primary sample. Engbom and Moser (2022) addresses the same concern differently: they only use the connected set to calculate all statistics, leaving workers employed at small, disconnected firms out of the sample. The best solution would be to formally model selection into the leave-one-out set, which would add significant complexity to the paper and is thus left to future work.

to $\boldsymbol{\theta}^G$, in region r, by $J_{r,o}(\boldsymbol{\theta}^G)$. Then, the following matrix exists and is nonsingular:

$$\boldsymbol{A}_0 = \underset{R\to\infty}{plim} \frac{1}{R} \sum_{r=1}^{R} \sum_{o=1}^{26} J_{r,o}(\boldsymbol{\theta}_0^G) J_{r,o}(\boldsymbol{\theta}_0^G)'$$

Assumption 6 (Limited dispersion of structural residuals)**.** *The following matrix exists and is positive definite:*

$$\boldsymbol{B}_0 = \underset{R\to\infty}{plim} \frac{1}{R} \sum_{r=1}^{R} \sum_{o=1}^{26} \sum_{o'=1}^{26} J_{r,o}(\boldsymbol{\theta}_0^G) J_{r,o'}(\boldsymbol{\theta}_0^G)' u_{r,o} u_{r,o'}$$

These assumptions allow for the identification of model parameters:

Proposition 6 (Identification, estimation, and inference)**.** *Under Assumptions 1 through 6, the following nonlinear least squares estimator*

$$\hat{\boldsymbol{\theta}}^G = \arg\min_{\boldsymbol{\theta}^G} \sum_{r=1}^{R} \left[\boldsymbol{Y}_r - \tilde{a}\left([\boldsymbol{Z}_r', PB(\boldsymbol{y}_r)']', \boldsymbol{\theta}^G\right)\right]' \left[\boldsymbol{Y}_r - \tilde{a}\left([\boldsymbol{Z}_r', PB(\boldsymbol{y}_r)']', \boldsymbol{\theta}^G\right)\right]$$

has the following asymptotic distribution:

$$\sqrt{R}(\hat{\boldsymbol{\theta}}^G - \boldsymbol{\theta}_0^G) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{A}_0^{-1} \boldsymbol{B}_0 \boldsymbol{A}_0^{-1}\right)$$

Appendix D.6 contains a thorough discussion of identification. First, it demonstrates how the invertibility assumption allows for addressing unobserved heterogeneity at the regional level without causing incidental parameter bias. Next, it provides an intuitive description of how each parameter is identified. It also proposes a parallel between my estimator and a nonlinear instrumental variables design, with standard errors clustered at the region level. Finally, it discusses the identification assumptions in the Brazilian context and considers threats such as regional differences in schooling quality.

The empirical model is over-identified. To see that, note that there are 52 general parameters to be estimated, but Assumptions 1 and 3 imply 260 moments of the form $E[u_{r,o}X_{r,k}] = 0$ where o indexes outcomes not used in the inversion procedure and $X_{r,k}$ is either a constant, a region-level covariate used to construct supply or demand parameters, or a region-specific parameter (see Appendix D.6 for a discussion of the theoretical restrictions leading to over-identification). Thus, different from a standard regression model, my structural model is not guaranteed to match average levels of target outcomes, and cross-sectional fit may in principle be poor. In the following, I show that, on the contrary, the model fits the data rather well given its relative parsimony.

6.2.3 Estimated parameters

I estimate the model using the Levenberg-Marquardt method with region and equation weights. Region weights are identical to those used in Section 3: total formal employment in the region (adding up both years). Equations were weighted by the inverse mean squared error from the "Simple" regressions described in Appendix D.7.4. In essence, the procedure down-weights moments that have more residual variation after eliminating the linear contributions of time effects, educational composition, and minimum wage bindingness. The one exception to this rule is the minimum wage spike. There, the equation weight is half of what those residuals imply. The model tends to overestimate the spike, possibly because it does not include factors such as fairness/relative wage concerns within the firm.[26] Assigning half weight to the spike allows the model to match that moment fairly well without worsening the fit quality in other dimensions. See Appendix D.4 for an extended discussion.

Estimation is computationally costly because, for each region, one must invert the regional parameters based on the subset of endogenous variables, find the equilibrium, and then simulate all moments. Each optimization step requires performing that procedure 16,006 times: 151 regions $\times$ 2 time periods $\times$ (1 base value + 52 Jacobian columns). Furthermore, because the loss function may not be globally concave, several starting points must be used. Appendix D.7 details the implementation, describing, for example, how the inversion and equilibrium finding procedures can be performed simultaneously.

Table 4 shows a subset of the estimated parameters. The others—labor demand determinants $\delta_i^{d,t}$—appear in Appendix Table D3. Before discussing the values I find for the key elasticities in the model, I note that one parameter, the dispersion in comparative advantage for workers with less than secondary education, was estimated at the boundary of the parametric space: $\hat{S}_{\hat{h}=1} = 0$. The interpretation is that, within the group of workers with little formal education, productivity differences exclusively reflect heterogeneity in efficiency units of labor

[26]Suppose that after a minimum wage hike, if wages are adjusted to match the new minimum, entry-level wages become as high as wages for more senior workers engaged in the same type of work. In that case, the firm may introduce raises for the latter such that they remain above the spike to preserve the ordering of earnings within the firm.

Table 4: Selected parameter estimates

Parameter	Estimate	Std. Error
Panel A: Worker types		
$\mu_{\hat{h}=2}$ (modal comp. adv. type, secondary)	3.67	(0.10)
$\rho_{\hat{h}=1}$ (dispersion in comp. adv., less than secondary)	0	-
$\rho_{\hat{h}=2}$ (, secondary)	2.52	(0.07)
$\rho_{\hat{h}=3}$ (, tertiary)	3.05	(0.28)
$\hat{S}_{\hat{h}=1}$ (dispersion in abs. adv., less than secondary)	0.81	(0.06)
$\hat{S}_{\hat{h}=2}$ (, secondary)	0.19	(0.04)
$\hat{S}_{\hat{h}=3}$ (, tertiary)	0.40	(0.15)
χ (skewness of abs. adv. distribution)	-0.99	(0.29)
Panel B: Worker preferences		
σ (elast. of substition between goods)	8.36	(3.33)
λ (aggregate labor supply parameter)	1.51	(0.16)
$\log \hat{z}^{HT}_{0,\hat{h}=1,t=2}$ (participation shock, less than secondary)	-0.04	(0.02)
$\log \hat{z}^{HT}_{0,\hat{h}=3,t=2}$ (participation shock, tertiary)	-0.66	(0.19)
Λ (rel. effect of regional part. shocks on tertiary)	0.27	(0.11)
Panel C: Labor demand		
κ (blueprint shape)	6.80	(1.27)
$\theta_{1,1998}/\theta_{2,1998}$ (rel. complexity of low-skill good, 1998)	0.32	(0.04)
$\theta_{1,1998}/\theta_{2,2012}$ (, 2012)	0.11	(0.01)

Notes: Standard errors are cluster-robust at the region level. They are calculated using the asymptotic formula in Proposition 6, using sample analogs for the populational matrices A_0 and B_0.

instead of comparative advantage.[27]

The estimated elasticity of substitution between goods is $\sigma = 8.36$. That estimate is somewhat imprecise, but the hypothesis $\sigma \leq 1$ (goods are *not* net substitutes) is rejected at the 95% confidence level. Thus, we should expect significant reallocation effects in the long run. Since the point estimate is significantly higher than $\beta + 1 = 5$, an increase in the supply of skilled workers should also increase the share of those workers employed at skill-intensive firms, following Proposition 4.

One may wonder whether the model predicts unplausible changes in the composition of firm types in the economy in the long run, given the large estimate for σ. To investigate that possibility, I calculate shares of employed workers at firms producing good $g = 2$ in each region and period based on the estimated model parameters. The mean change in that

[27] Table 4 does not report standard errors for that parameter because the asymptotic formula is not valid at the boundary. To conduct inference on other parameters, I assume that $\hat{S}_{\hat{h}=1}$ is known to be equal to zero, such that it does not belong to the vector Θ^G.

share is -0.116, with a standard deviation of 0.070. The highest increase is from 0.219 to 0.366, while the most negative change is from 0.625 to 0.252. That means the production possibilities frontier is "concave enough" to prevent unrealistic reallocation responses and corner solutions, even though the shocks affecting the Brazilian economy are substantial.

The estimated λ implies aggregate labor supply elasticities to the formal sector of around 0.6-0.7 for college workers. These values are in the upper range of steady-state elasticities inferred from microdata in the US but also below the values between 1 and 2 typically used in macroeconomic models (Keane and Rogerson, 2012). Elasticities are larger for less skilled workers, reaching 1.1 for those with less than high school in 1998. This difference aligns well with informality being an important outside option for those workers. For example, Dix-Carneiro and Kovak (2017) find evidence of formal-informal transitions in microregions more affected by trade liberalization.

6.2.4 Quality of fit

Columns (1)—(4) in Table 5 show that the model closely tracks averages in the data, successfully capturing the overall decline in inequality (especially within groups) and the increase in sorting. The most significant deviation is in the mean return to college (tertiary/secondary), which increases in the data but falls in the estimated model. That moment has the lowest estimation weight. Although the model fails to capture the average increase in the college premium, there are 33 regions in the estimated model where the college premium rises, compared to 47 such regions in the data.

A more comprehensive measure of fit is the R2 statistic for each individual moment, reported in Column (5). The statistics are positive for all targeted moments, even for the college premium. But it is difficult to make sense of that metric without context. A low R2 may come from either a failure of the model to fit the data or a lack of sufficient explanatory power in the covariates used by the model. To distinguish between these two possibilities, I estimate benchmark predictive models based on Ordinary Least Squares (OLS) regressions. The "Simple" model is constructed to have the same number of parameters as the structural model. It includes the minimum wage bindingness measure, educational shares for secondary and tertiary, and time dummies as regressors. The "Large" model includes several other variables, such as initial sectoral shares and a quadratic component for minimum wage bindingness. It features a total of 112 parameters, more than twice as many as in the structural model. Those models are guaranteed to match time-specific averages for all moments.

Table 5: Quality of fit and comparison to benchmark predictive models

	Data		Model		R2	Benckmark R2	
	1998	2012	1998	2012	Model	Simple	Large
Moments	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Wage inequality measures							
Secondary / less than secondary	0.498	0.168	0.478	0.168	0.755	0.78	0.812
Tertiary / secondary	0.965	1.038	0.981	0.954	0.127	0.169	0.407
Within less than secondary	0.41	0.241	0.401	0.233	0.607	0.706	0.792
Within secondary	0.684	0.355	0.67	0.331	0.848	0.761	0.86
Within tertiary	0.702	0.624	0.701	0.637	0.139	0.255	0.379
Total variance of log wages	*0.745*	*0.553*	*0.733*	*0.514*	*0.758*		
Two-way fixed effects decomposition							
Variance establishment effects	0.126	0.054	0.126	0.04	0.586	0.634	0.667
Covariance worker, estab. effects	0.052	0.046	0.056	0.059	0.374	0.354	0.485
Variance worker effects	*0.454*	*0.368*	*0.44*	*0.316*	*0.439*		
Correlation worker, estab. effects	*0.224*	*0.315*	*0.234*	*0.539*	*-1.395*		
Formal employment rates							
Less than secondary	0.266	0.337	0.266	0.335	0.953	0.956	0.979
Secondary	0.435	0.508	0.435	0.508	1.0	1.0	1.0
Tertiary	0.539	0.629	0.539	0.63	0.89	0.93	0.95
Minimum wage bindingness							
Log min. wage - mean log wage	-1.418	-0.922	-1.418	-0.922	1.0	1.0	1.0
Share < log min. wage + 0.05	0.031	0.053	0.046	0.084	0.528	0.576	0.785
Share < log min. wage + 0.30	0.086	0.212	0.107	0.201	0.873	0.738	0.904

Notes: Moments targeted by the estimation procedure appear as plain text. Untargeted moments are *italicized*. Columns (1) through (4) report national averages of the corresponding moments for each year, calculated using region weights based on total formal employment. Column (5) reports the usual R2 metric $r_e^2 = 1 - \left[\sum_{t=1}^{2}\sum_{r=1}^{151} s_r \left(Y_{e,r,t} - \hat{Y}_{e,r,t}\right)^2\right] / \left[\sum_{t=1}^{2}\sum_{r=1}^{151} s_r \left(Y_{e,r,t} - \bar{Y}_e\right)^2\right]$, where e indexes the specific target moment, $\hat{Y}_{e,r,t}$ is the model prediction, and $\bar{Y}_e$ is the sample average using the region weights s_r. Columns (6) and (7) report analogous R2 metrics for benchmark OLS models for comparison purposes (see Appendix D.7.4).

See Appendix D.7.4 for details.

My model fits the data approximately as well as the Simple OLS benchmark. It is worse for inequality measures and participation rates among college workers but better for AKM moments and bindingness measures. Although the Large OLS model has a better R2 for all moments, for many of them, the difference is not substantial. These results support the view that the functional form assumptions and theoretical restrictions imposed by the model are reasonable, since they do not prevent the model from fitting the data as well as standard regression models.

To further validate the model, I verify the quality of fit for outcomes not directly targeted by the estimation procedure. Table 5 shows that the model has predictive power for the overall variance of log wages and the variance of worker effects. The correlation between worker and establishment fixed effects is significantly higher in the model than in the data for 2012, due to a combination of the covariance component being overestimated and both variance components being underestimated. Still, the qualitative pattern of increasing sorting of high-wage workers to high-wage firms within regions is replicated.[28] Appendix D.7.5 shows a series of additional measures of fit, including histograms of log wages and measures of minimum wage bindingness by educational group. It also shows that good quality of fit is not an artifact of using region weights. These exercises reinforce the conclusion that the model provides a good approximation for Brazilian labor markets.

6.3 Counterfactual exercises

This subsection presents the counterfactual analyses that I use to understand how supply, demand, and minimum wage shocks affected Brazilian labor markets between 1998 and 2012. The supply shock is the change in the educational composition of the adult population, and the minimum wage shock is an increase of 66.1 log points in the minimum wage relative to the price index in all regions. The demand shock combines all other time-varying factors in the model: TFP, task requirements, relative entry costs, relative consumer taste, and outside option parameters. In Appendix D.8.2, I discuss why outside options are included in the demand shock and separately show comparative statics for different demand parameters.

6.3.1 Supply, demand, minimum wage, and their interactions

Table 6 shows the impact of those shocks on mean log wages, the variance of log wages, and variance components from the AKM decomposition. Columns (1) and (2) show base levels and total changes for each outcome, averaged over regions. Columns (3), (4), and (5) explore counterfactuals where only one factor changes. Columns (6), (7), and (8) show pairwise interactions. Specifically, I simulate the combined effect of two shocks and then subtract the corresponding individual effects. Finally, Column (9) shows the triple interaction, that is, the difference between Column (2) and the sum of Columns (3)–(8). Appendix D.8.1 shows similar decompositions for other outcomes of interest.

[28] As mentioned in the descriptive section, the KSS estimate of the correlation between worker and establishment effects may be biased, such that part of the low quality of fit for this untargeted moment may be due to measurement issues.

Table 6: Effects of supply, demand, minimum wage, and their interactions

Outcome	Base value (1)	All Changes (2)	Individual effects: S (3)	D (4)	M (5)	Interactions S+D (6)	S+M (7)	D+M (8)	Triple (9)
Mean log real wage	1.42	0.17	0.24	-0.06	0.15	-0.08	-0.04	-0.04	0.01
Variance of log wages	0.73	-0.22	0.04	-0.18	-0.13	0.00	0.02	0.03	0.01
Var. worker effects	0.44	-0.12	-0.01	-0.06	-0.06	0.01	0.01	-0.03	0.01
Var. estab. effects	0.13	-0.09	0.00	-0.09	0.00	0.00	-0.01	0.00	0.01
2×Cov. worker, estab	0.11	0.01	0.04	-0.02	-0.07	-0.01	0.01	0.06	-0.01
Var. residuals	0.06	-0.02	0.00	-0.01	-0.01	0.00	0.00	0.00	0.00
Corr. worker, estab effects	0.23	0.30	0.09	0.19	-0.14	0.01	0.04	0.15	-0.03

Notes: Each row shows within-region effects averages across all 151 regions using total formal employment summed over 1998 and 2012 as weights. "S" is for supply (the rise in educational achievement of the adult population), "D" is for the combined changes in demand-side parameters, and "M" is for the real minimum wage increase of 66.1 log points. See the text for an explanation of each column.

The combination of demand shocks and the minimum wage had strong inequality-reducing effects in Brazil, while the supply shock had a weak—but positive—effect. The model also reveals significant interactions that would not be detectable without a unified approach. For example, if minimum wages were the only change happening between 1998 and 2012, the variance of log wages would have fallen by 0.13. However, another meaningful counterfactual involves considering what would have happened if supply and demand changed, but the minimum wage stayed at the 1998 level. In that case, inequality in 2012 would be higher, but only by 0.07. In other words, the inequality-reducing effects of the minimum wage were significantly dampened by other transformations affecting the Brazilian economy during the same period.

The increase in assortative matching of high-wage workers to high-wage firms comes mainly from the shocks in demand. The increase in educational achievement also contributes to that phenomenon. If minimum wages were the only transformation affecting the Brazilian economy between 1998 and 2012, sorting would have fallen significantly. However, when accompanied by the supply and demand changes described in the previous paragraph, its marginal effect on the covariance (or correlation) between worker and establishment effects is close to zero.[29]

[29] One may wonder whether the heterogeneity in the causal effects of the minimum wage would be captured by models that assume minimum wage effects are heterogeneous depending on how binding it is. The answer is no. In both exercises, the initial and final levels of average minimum wage bindingness are about the same regardless of which bindingness metric is used. If anything, the increase in the log minimum wage relative to the mean log wage is higher in the second exercise, where the minimum wage impacts are weakest. See

The following two sections discuss the effects of supply and minimum wage in more detail. Before getting there, I briefly discuss the nature and impact of labor demand shocks in the Brazilian context. They are more challenging to interpret for two reasons. First, in my methodology, these shocks are not observed. Instead, they inferred from observed changes in inequality and sorting after netting out the contribution of supply and minimum wage, somewhat analogously to how skill-biased technical change is inferred from a residual trend in Katz and Murphy (1992). Second, several types of labor demand shocks could have plausibly affected Brazilian labor markets in the period I study: trade liberalization, the commodities boom, import competition from China in the same period in the manufacturing sector, skill-biased technical change from computers and other forms of capital equipment, routine-biased technical change, the expansion of broadband internet, and increased adoption of computer numerical control machines and industrial robots.

According to the model, the net effect of those transformations was a reduction in most inequality measures and an increase in assortative matching. They come from changes in structural parameters that, on average, make the two firm types in the model closer to each other regarding entry costs but further apart in task complexity, such that segregation by skill increases. Interestingly, the significant reduction in the variance of establishment fixed effects in the Brazilian context, first documented by Alvarez et al. (2018), is explained in my model by those demand-side transformations. That interpretation is consistent with the results in Engbom and Moser (2022). Using their estimated structural model, they find that the Brazilian federal minimum wage increase barely affects the variance of establishment effects (see Appendix E.2 in their paper).

6.3.2 Supply effects: composition, returns to skill, or reallocation?

Supply shocks may affect wage distribution via a purely compositional effect: with more skilled workers, average wages should increase. The variance in log wages should also increase because there is more within-group productivity dispersion among more educated adults. The compositional change may also have a statistical effect on measured sorting.

The model also specifies two types of endogenous responses to the supply shock. The first derives from firms reoptimizing their wage-posting decisions. To isolate this effect, I calculate a partial equilibrium in which firm creation and prices for goods remain at their initial levels. This channel is powered by the concavity of the task-based production function and

Appendix Table D7, Panel C for details.

Table 7: Decomposition of the impact of supply shocks

Outcome	Total supply effect (1)	Compositional effect (2)	Firm choices (3)	Entry and prices (4)
Mean log real wage	0.24	0.17	0.01	0.06
Variance of log wage	0.04	0.10	-0.03	-0.03
Corr. worker, estab. effects	0.09	0.10	-0.02	0.02

Notes: Column (1) repeats Column (3) from Table 6. Column (2) measures changes induced by a re-weighting of worker types, keeping log wages and employment shares unchanged. Column (3) measures changes from that scenario to a "partial equilibrium" where firms and workers allowed to reoptimize while but firm entry and log prices are kept constant. Column (4) corresponds to the change from the partial equilibrium to the new general equilibrium, accounting for entry and price responses.

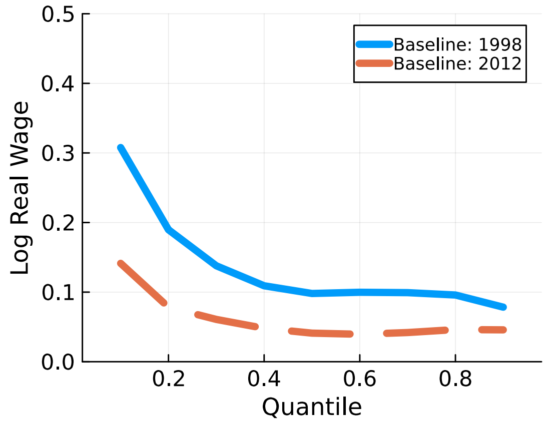
represents the central component of competitive models that focus on between-group inequality.

The other endogenous response derives from the changes in firm entry and prices emphasized in Proposition 4. Given that the estimated σ is large, we should expect net reallocation of labor toward high-wage, skill-intensive firms.

Table 7 reports the magnitudes of each of those channels. Although I find that the compositional effect is the most important, equilibrium effects cannot be ignored. Wage posting responses cut the inequality and sorting effects associated with compositonal changes by 20-30%. Meanwhile, entry responses boost the effect of supply on the mean log real wage by 33%.

In the discussion of Proposition 4, I argued that positive supply shocks may widen the aggregate skill wage premium, instead of compressing it as in the canonical competitive model. To verify that possibility in the Brazilian context, I simulate the effects of small increases in the share of workers with complete college, with a corresponding reduction in the share with less than high school. If the baseline models are the 1998 equilibria, the mean log wage gap between those educational groups falls in all but six regions. However, if the baseline models are the 2012 equilibria, mean log wage gaps widen in 109 out of 151 regions. This result reinforces the importance of accounting for firm wage premiums, sorting, and endogenous firm entry when calculating the long-run effects of educational shocks. It also illustrates how reallocation effects depend on not only structural elasticities but also the characteristics of the initial equilibrium, such as the starting level of segregation by skill across firms.

Figure 6: Minimum wage spillovers



Notes: This figure shows minimum wage impacts on quantiles of within-region log wage distributions, averaged over all regions. The blue line corresponds to a 66.1 log point increase in the minimum wage starting from the 1998 equilibria. The orange dashed line corresponds to a similar-sized reduction starting from the 2012 equilibria.

6.3.3 The impacts of the rising minimum wage

A simple, but useful way of visualizing the impact of a rising minimum wage on the wage distribution is plotting by the causal change in log wage quantiles. Figure 6 shows average within-region spillovers implied by the estimated model. Real log wages increase for all deciles, especially the lowest. The minimum wage substantially reduces inequality in the lower tail of the wage distribution but has essentially no inequality-reducing effects in the upper tail. The difference between the two curves reflects the significant interactions I have described in Subsection 6.3.1: the wage distribution effects of a minimum wage increase starting from 1998 are about twice as strong as similar-sized reduction based on the model as of 2012.

Next, I evaluate how the minimum wage affects average outcomes of groups defined based on productivity. That exercise differs from Figure 6 because the same quantile of the wage distribution may correspond to workers of different productivity levels before and after the introduction of the minimum wage. If the minimum wage causes disemployment of low-skilled workers, changes in quantiles may be mechanical consequences of truncation, as explained by Lee (1999). And even if net employment effects are zero, the minimum wage may introduce compositional changes in the lower tail which could affect observed log wage quantiles, as illustrated in Figure 3.

Table 8 reports the results of that analysis, showing the effects of a ceteris paribus increase in

Table 8: Wage and employment effects of the minimum wage

Prod. decile (1)	Pop. share (2)	Base wage (3)	Mean wage changes: Monops. (4)	Ret. sk. (5)	Gen. eq. (6)	Base emp. (7)	Emp. elasticities w.r.t.: Min. (8)	Mean (9)	·, monops. (10)
1	0.15	1.24	0.75	-0.01	-0.01	0.21	-0.61	-0.97	-0.92
2	0.12	1.78	0.49	-0.02	-0.02	0.27	-0.16	-0.61	-0.58
3	0.11	2.35	0.02	-0.01	0.02	0.28	0.01		
4	0.11	2.97	-0.00	-0.02	0.04	0.29	-0.01		
5	0.10	3.75	-0.00	-0.02	0.05	0.31	-0.01		
6	0.10	4.76	0.00	-0.02	0.06	0.33	-0.01		
7	0.09	6.11	-0.00	-0.01	0.07	0.37	-0.01		
8	0.08	8.13	-0.00	0.00	0.07	0.40	-0.00		
9	0.07	11.91	0.00	0.04	0.08	0.45	-0.00		
10	0.06	25.04	-0.00	0.11	0.09	0.50	0.00		

Notes: Each row shows causal effects of an increase of 66.1 log points in the minimum wage in all regions for a subset of adults. Adults are grouped based on productivity at the skill-intensive firms, such that each row corresponds to 10% of the employed population (i.e., the product of Columns (2) and (7) is constant across rows). Wage effects are decomposed as described in Subsection 5.6: monopsony, returns to skill, and general equilibrium. Columns (8) and (9) report elasticities of employment with respect to the log real minimum wage and the mean wage for the group. Column (10) resembles Column (9) but only considers the monopsony channel.

the minimum wage starting from the 1998 equilibria. The workers are grouped at the national level based on their productivity if they were employed at skill-intensive firms based on their region. Column (3) shows the mean wage for the subset of adults employed at the initial equilibrium. Columns (4), (5), and (6) show how the mean wage for employed workers changes within that group of adults. Each column isolates one of the channels described in Subsection 5.6: "monopsony" (which combines disemployment, positive employment effects, mechanical wage increases, and cross-firm reallocation), "returns to skill" (the partial equilibrium analysis with firm creation and prices fixed at their initial levels), and "general equilibrium" (corresponding firm creation and price responses).

I find that only workers in the bottom two deciles of productivity see large changes in average wages; for all others, there is a small increase of about one percent once all channels are accounted for. The strong wage effects in the lower tail of productivity come mostly from the "monopsony" channel. The small increases for all other workers, in turn, come from the other two channels.

The returns to skill channel lead to small reductions in average wages for low- and middle-skill workers, coupled with wage increases at the top two deciles. Those effects originate from the reallocation of low-skilled low- to high-wage firms, which increases the returns to skill at those firms.

General equilibrium effects generate modest wage increases for almost all worker types, coming from increased entry of high-wage firms. But for very low productivity workers, those effects are negative. The reason is that as the number of low-wage, low-skill firms decrease, relative labor demand for low-skill workers falls.

The small wage effects for workers in productivity deciles three and above may, at first glance, seem difficult to reconcile with the positive effects on all wage quantiles reported in Figure 6. The difference emerges from disemployment effects in the lower two deciles. Columns (8) and (9) of Table 8 show that the implied employment elasticities for the lowest group are in the lower range of estimates for the US (Harasztosi and Lindner, 2019; Neumark and Shirley, 2021).[30] In Appendix D.8.4, I discuss possible reasons why the predictions of my model regarding disemployment effects in the lower tail differ from recent reduced-form regression results reported by Engbom and Moser (2022).

I end this section by examining why my estimates of minimum wage effects along the wage distribution differ from predictions from the structural model estimated by Engbom and Moser (2022). Specifically, their simulations imply that the wage effects of the Brazilian federal minimum wage extend much farther up the wage distribution.

There are two main reasons for that difference. Engbom and Moser (2022) uses a wage-posting model with perfect substitution between workers at the firm level and without non-wage amenities. A consequence of those assumptions is that an increase in the minimum wage shock boosts posted wages at all other firms (with stronger effects at low-wage firms). In my model, wage-posting responses at high-wage firms can instead be negative for low-skill workers, due to the returns to skill channel mentioned above.

The second main reason why my results differ from those in Engbom and Moser (2022) is my use of a local labor markets approach. In their model, disemployment effects for very low-skilled workers are dampened by reallocation to firms in the top 5% of the productivity distribution (see their Appendix Figure E.3). Many of these firms may be in the wealthiest parts of the country, while the displaced workers may be in the poorest. My model does not

[30] In Column (10), I report the elasticity of employment with respect to the minimum wage ignoring the returns to skill and general equilibrium channels. The resulting numbers smaller but close to the ones in Column (9), suggesting that ignoring those channels when evaluating the impact of the minimum wage for very low skill workers may not lead to dramatic biases.

allow for geographical mobility, limiting the extent of minimum wage-induced reallocation. This approach is consistent with Dix-Carneiro and Kovak (2017), who document that the Brazilian microregions most affected by tariff reductions in the 1990s saw declines in formal employment but no systematic out-migration responses.

7 Conclusion

The unified framework proposed in this paper combines two labor economics perspectives on the determinants of wage distribution: supply/demand models focusing on endogenous productivity gaps between workers and imperfectly competitive labor market models focusing on firm wage differentials and sorting. I have demonstrated that these two traditions interact. One such interaction is how including firm wage premiums in a supply/demand framework may qualitatively change the effects of education on wage distribution. After estimating a structural model to understand the fall of wage inequality Brazilian labor markets, I find that a significant part of the positive effects of education on average wages comes not from increased worker productivity or changes in returns to skill but from increased entry of firms that pay high wage premiums to its workers. Moreover, because those high-wage firms are also more skill-intensive, the effect of an increase in the supply of skilled workers on the average skill premium can be positive in some cases.

I also show that combining task-based production, firm heterogeneity, and monopsony power generates new channels through which the minimum wage affects employment and wages. For example, following a minimum wage increase, high-wage firms may decrease wages offered to low-skill workers. This effect arises from the reallocation of low-skill workers to high-wage firms, which changes the assignment of workers to tasks at those firms in a way that reduces the marginal product of low-skilled labor. The unified approach also allows for measuring how minimum wages interact with changes in supply and demand factors. In Brazil, the rise of the federal minimum wage was accompanied by growing educational achievement and a series of labor demand shocks. The structural model predicts that, while the minimum wage contributed to falling inequality, it would have had a much stronger impact if it were the only shock affecting the economy at that time.

As a technical contribution, the paper introduces the task-based production function and shows that it is a convenient tool for studying labor markets with rich worker and firm heterogeneity. It offers a tractable, intuitive, and parsimonious means of modeling cross-firm differences in labor demand patterns. It also enables the modeling of different forms of technical change. One avenue for further research using that approach is understanding the effects of *routine-biased technical change* (Autor, Levy and Murnane, 2003; Acemoglu and Autor, 2011) in a context with firm heterogeneity and imperfect competition.

References

Aaronson, Daniel, Eric French, Isaac Sorkin, and Ted To. 2018. "INDUSTRY DYNAMICS AND THE MINIMUM WAGE: A PUTTY-CLAY APPROACH." *International Economic Review*, 59(1): 51–84.

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

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *The Quarterly Journal of Economics*, 113(4): 1055–1089.

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. 2007. "Equilibrium Bias of Technology." *Econometrica*, 75(5): 1371–1409.

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.

Ales, Laurence, Musab Kurnaz, and Christopher Sleet. 2015. "Technical Change, Wage Inequality, and Taxes." *American Economic Review*, 105(10): 3061–3101.

Almeida, Rita, and Pedro Carneiro. 2012. "Enforcement of Labor Regulation and Informality." *American Economic Journal: Applied Economics*, 4(3): 64–89.

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.

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.

Azar, José, Emiliano Huet-Vaughn, Ioana Marinescu, Bledi Taska, and Till von Wachter. 2019. "Minimum Wage Employment Effects and Labor Market Concentration." Working Paper 26101.

Bagger, Jesper, and Rasmus Lentz. 2019. "An Empirical Model of Wage Dispersion with Sorting." *The Review of Economic Studies*, 86(1): 153–190.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *American Economic Review*, 112(4): 1147–93.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2024. "Minimum Wages, Efficiency and Welfare." *Unpublished manuscript.*

Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah. 2017. "Julia: A fresh approach to numerical computing." *SIAM review*, 59(1): 65–98.

Blundell, Richard, David A Green, and Wenchao Jin. 2021. "The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium." *The Review of Economic Studies*, 89(1): 142–180.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data." *Econometrica*, 87(3): 699–739.

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.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review*, 39(2): 257–273.

Butschek, Sebastian. 2022. "Raising the Bar: Minimum Wages and Employers' Hiring Standards." *American Economic Journal: Economic Policy*, 14(2): 91–124.

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 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.

Carneiro, Pedro, Kai Liu, and Kjell G Salvanes. 2022. "The Supply of Skill and Endogenous Technical Change: Evidence from a College Expansion Reform." *Journal of the European Economic Association*, 21(1): 48–92.

Costa, Francisco, Jason Garred, and João Paulo Pessoa. 2016. "Winners and losers from a commodities-for-manufactures trade boom." *Journal of International Economics*, 102: 50–69.

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

Coviello, Decio, Erika Deserranno, and Nicola Persico. 2022. "Minimum Wage and Individual Worker Productivity: Evidence from a Large US Retailer." *Journal of Political Economy*, 130(9): 2315–2360.

Deming, David, and Lisa B. Kahn. 2018. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics*, 36(S1): S337–S369.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics*, 132(4): 1593–1649.

Derenoncourt, Ellora, François Gérard, Lorenzo Lagos, and Claire Montialoux. 2021. "Racial Inequality, Minimum Wage Spillovers, and the Informal Sector." Unpublished manuscript.

de Sousa, Gustavo. 2020. "The Labor Market Consequences of (In)Appropriate Technology." Unpublished manuscript.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2017. "Trade Liberalization and Regional Dynamics." *American Economic Review*, 107(10): 2908–46.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp vom Berge. 2021. "Reallocation Effects of the Minimum Wage*." *The Quarterly Journal of Economics*, 137(1): 267–328.

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

Eeckhout, Jan, and Roberto Pinheiro. 2014. "Diverse Organizations and the Competition for Talent." *International Economic Review*, 55(3): 625–664.

Engbom, Niklas, and Christian Moser. 2022. "Earnings Inequality and the Minimum Wage: Evidence from Brazil." *American Economic Review*, 112(12): 3803–47.

Engerman, S.L., and K.L. Sokoloff. 2012. *Economic Development in the Americas Since 1500: Endowments and Institutions. NBER Series on Long-Term Facto*, Cambridge University Press.

Firpo, Sergio, and Alysson Portella. 2019. "Decline in Wage Inequality in Brazil: A Survey." World Bank Policy Research Working Paper 9096.

Flinn, Christopher J. 2006. "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates." *Econometrica*, 74(4): 1013–1062.

Fortin, Nicole M., Thomas Lemieux, and Neil Lloyd. 2021. "Labor Market Institutions and the Distribution of Wages: The Role of Spillover Effects." *Journal of Labor Economics*, 39(S2): S369–S412.

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.

Haanwinckel, Daniel. 2024. "Does regional variation in wage levels identify the effects of a national minimum wage?" arXiv:2307.01284.

Hahn, Jinyong, and Guido Kuersteiner. 2002. "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large." *Econometrica*, 70(4): 1639–1657.

Hahn, Jinyong, and Whitney Newey. 2004. "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models." *Econometrica*, 72(4): 1295–1319.

Harasztosi, Peter, and Attila Lindner. 2019. "Who Pays for the Minimum Wage?" *American Economic Review*, 109(8): 2693–2727.

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.

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M Phillips. 2018. "Production and Learning in Teams." National Bureau of Economic Research Working Paper 25179.

Hurst, Erik, Patrick Kehoe, Elena Pastorino, and Thomas Winberry. 2022. "The Distributional Impact of the Minimum Wage in the Short and Long Run." *Unpublished manuscript.*

"IBGE", Instituto Brasileiro de Geografia e Estatística. 2003. *Metodologia do Censo Demográfico 2000.* IBGE.

Jäger, Simon, and Jörg Heining. 2022. "How Substitutable Are Workers? Evidence from Worker Deaths." *Unpublished manuscript*.

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.

Keane, Michael, and Richard Rogerson. 2012. "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom." *Journal of Economic Literature*, 50(2): 464–476.

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.

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler. 2023. "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry." NBER Working Paper 27325.

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.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2022. "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market." *American Economic Review*, 112(1): 169–212.

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.

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

Lise, Jeremy, and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *American Economic Review*, 110(8): 2328–76.

Neumark, David, and Peter Shirley. 2021. "Myth or Measurement: What Does the New Minimum Wage Research Say about Minimum Wages and Job Loss in the United States?" Working Paper 28388.

Neumark, David, and William Wascher. 1992. "Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws." *ILR Review*, 46(1): 55–81.

Ponczek, Vladimir, and Gabriel Ulyssea. 2021. "Enforcement of Labour Regulation and the Labour Market Effects of Trade: Evidence from Brazil." *The Economic Journal*, 132(641): 361–390.

Rohlin, Shawn M. 2011. "State minimum wages and business location: Evidence from a refined border approach." *Journal of Urban Economics*, 69(1): 103–117.

Sattinger, Michael. 1975. "Comparative Advantage and the Distributions of Earnings and Abilities." *Econometrica*, 43(3): 455–468.

SEDLAC. 2022. "Socio-Economic Database for Latin America and the Caribbean." CEDLAS and The World Bank.

Shephard, Andrew, and Modibo Sidibe. 2019. "Schooling Investment, Mismatch, and Wage Inequality." PIER Working Paper 19-013.

Smith, Alexander A. 2021. "The minimum wage and teen educational attainment." *Labour Economics*, 73: 102061.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2018. "Firming Up Inequality*." *The Quarterly Journal of Economics*, 134(1): 1–50.

Stigler, George J. 1946. "The Economics of Minimum Wage Legislation." *The American Economic Review*, 36(3): 358–365.

Stolper, Wolfgang F., and Paul A. Samuelson. 1941. "Protection and Real Wages." *The Review of Economic Studies*, 9(1): 58–73.

Teulings, Coen, and Thijs van Rens. 2008. "Education, Growth, and Income Inequality." *The Review of Economics and Statistics*, 90(1): 89–104.

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.

Teulings, Coen N, and Pieter A Gautier. 2004. "The Right Man for the Job." *The Review of Economic Studies*, 71(2): 553–580.

Trottner, Fabian. 2019. "Who Gains From Scale? Trade and Wage Inequality Within and Between Firms." Unpublished manuscript.

Verhoogen, Eric A. 2008. "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector*." *The Quarterly Journal of Economics*, 123(2): 489–530.

Online Appendix
Supply, Demand, Institutions, and Firms

Daniel Haanwinckel
UCLA and NBER

June 13, 2024

Table of Contents

A Proofs

Section 4: Task-based production function

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

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

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

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

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

$$
\begin{aligned}
m'_{h_1}(x) &= m_{h_1}(x) - \tau, && x \in [x_2, x_2+\delta) \\
m'_{h_2}(x) &= m_{h_2}(x) + \tau\frac{e_{h_1}(x)}{e_{h_2}(x)}, && x \in [x_2, x_2+\delta) \\
m'_{h_2}(x) &= m_{h_2}(x) - \tau\frac{e_{h_1}(x+\Delta)}{e_{h_2}(x+\Delta)}, && x \in [x_1, x_1+\delta) \\
m'_{h_1}(x) &= m_{h_1}(x) + \tau\frac{e_{h_1}(x+\Delta)}{e_{h_2}(x+\Delta)}\frac{e_{h_2}(x)}{e_{h_1}(x)}, && x \in [x_1, x_1+\delta)
\end{aligned}
$$

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

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

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

Proof. Consider two cases:

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

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

Remember that the domain of f imposes $l_H > 0$. Because there is no slack of labor H, there must be some $\underline{x}$ with $m_H(\underline{x}) > 0$. Pick an arbitrarily small $\tau > 0$. By right continuity of m_H, there is a small enough $\delta > 0$ such that $m_H(x) > \tau\ \forall x \in [\underline{x}, \underline{x} + \delta)$. Let $\tilde{u}_h = \int_{\underline{x}}^{\underline{x}+\delta} e_H(x)/e_h(x)dx < \infty$ 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 $m'_h(x) = m_h(x) + g e_H(x)/e_h(x)$ in the same interval. The new candidate solution satisfies all constraints, has right continuous and non-negative assignment functions, and has slack of labor of type H. $\square$

Proof of Lemma 1, except non-arbitrage condition. From Lemma 5, we know that any optimal solution must not have any slack. The same Lemma implies that any candidate solution satisfying the conditions in Lemma 4 is also not optimal. So any optimal solution must be such that for any two tasks $x_1 < x_2$ and two types $h_1 < h_2$, $m_{h_2}(x_1) > 0 \Rightarrow m_{h_1}(x_2) = 0$ and $m_{h_1}(x_2) > 0 \Rightarrow m_{h_2}(x_1) = 0$. This property can be re-stated as: for any pair of types

$h_1 < h_2$, there exists at least one number ${}_{h_1}\bar{x}_{h_2}$ such that $m_{h_2}(x) = 0 \; \forall x < {}_{h_1}\bar{x}_{h_2}$ and $m_{h_1}(x) = 0 \; \forall x > {}_{h_1}\bar{x}_{h_2}$. By combining all such requirements together, there must be $H-1$ numbers $\bar{x}_1, \ldots, \bar{x}_{H-1}$ such that, for any type h, $m_h(x) = 0 \; \forall x \notin [\bar{x}_{h-1}, \bar{x}_h]$ (where $\bar{x}_0 = 0$ and $\bar{x}_H = \infty$ are introduced to simplify notation).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The determinant is:

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

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

□

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

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

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

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

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

Proof of Corollary 1: Distance-dependent complementarity

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

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

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

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

Which has first order conditions:

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

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

$$\frac{e_{h+1}\left(\bar{x}_{h,j}\right)}{e_h\left(\bar{x}_{h,j}\right)} = \frac{\omega_{h+1}}{\omega_h}\left(\frac{l_{h+1,j}}{l_{h,j}}\right)^{\frac{1}{\beta}}\left(\frac{L_{h+1,j}}{L_{h,j}}\right)^{-\frac{1}{\beta}} \qquad h \in \{1, \ldots, H-1\} \tag{12}$$

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

Proof of Proposition 2: Complementarity patterns may differ between firms

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

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

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

For firms producing $g = 2$, it is sufficient to show that all task thresholds move to the right following an increase in L_1. To see that, plug the labor supply expression into Equation 12 to find a monotonic link between posted wages and task thresholds:

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

Rewrite Equation 12 with task thresholds as the only endogenous variables (note that when the labor choices are divided, the choice of quantity cancels out):

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

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

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

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

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

$$\beta\left[\frac{e_1\left(\bar{x}_{1,j}\right)}{e_2\left(\bar{x}_{1,j}\right)}\right]\frac{d}{d\bar{x}_{1,j}}\left[\frac{e_2\left(\bar{x}_{1,j}\right)}{e_1\left(\bar{x}_{1,j}\right)}\right]+\frac{b_g(\bar{x}_{1,j})}{l_2e_2(\bar{x}_{1,j})}+\frac{b_g(\bar{x}_{1,j})}{l_1e_1(\bar{x}_{1,j})}>\frac{\frac{b_g(\bar{x}_{1,j})}{l_2e_2(\bar{x}_{1,j})}\frac{b_g(\bar{x}_{2,j})}{l_2e_2(\bar{x}_{2,j})}}{\beta\left[\frac{e_2\left(\bar{x}_{2,j}\right)}{e_3\left(\bar{x}_{2,j}\right)}\right]\frac{d}{d\bar{x}_{2,j}}\left[\frac{e_3\left(\bar{x}_{2,j}\right)}{e_2\left(\bar{x}_{2,j}\right)}\right]+\frac{b_g(\bar{x}_{2,j})}{l_3e_3(\bar{x}_{2,j})}+\frac{b_g(\bar{x}_{2,j})}{l_2e_2(\bar{x}_{2,j})}}$$

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

$$\begin{aligned}\frac{\frac{b_g(\bar{x}_{1,j})}{l_2e_2(\bar{x}_{1,j})}\frac{b_g(\bar{x}_{2,j})}{l_2e_2(\bar{x}_{2,j})}}{\beta\left[\frac{e_2\left(\bar{x}_{2,j}\right)}{e_3\left(\bar{x}_{2,j}\right)}\right]\frac{d}{d\bar{x}_{2,j}}\left[\frac{e_3\left(\bar{x}_{2,j}\right)}{e_2\left(\bar{x}_{2,j}\right)}\right]+\frac{b_g(\bar{x}_{2,j})}{l_3e_3(\bar{x}_{2,j})}+\frac{b_g(\bar{x}_{2,j})}{l_2e_2(\bar{x}_{2,j})}} &< \frac{\frac{b_g(\bar{x}_{1,j})}{l_2e_2(\bar{x}_{1,j})}\frac{b_g(\bar{x}_{2,j})}{l_2e_2(\bar{x}_{2,j})}}{\frac{b_g(\bar{x}_{2,j})}{l_2e_2(\bar{x}_{2,j})}} \\ &= \frac{b_g(\bar{x}_{1,j})}{l_2e_2(\bar{x}_{1,j})}\end{aligned}$$

Section 5: Markets and wages

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

Proof of Lemma 3: Firm problem and representative firms

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

A positive amount of l_H ensures that all other types are employed as well. Consider a particular type $h < H$ and whether it is optimal to set $l_h = 0$, fixing employment of all other types. Because $l_H > 0$, $\bar{x}_{H-1}$ is finite, and thus threshold $\bar{x}_h$ (the highest task performed by h) is guaranteed to be finite as well. Then, from Equation 2, the marginal product of type

h is bound below by $\underline{f}_H e_h(\bar{x}_{H-1})/e_H(\bar{x}_{H-1})$. A similar reasoning as above establishes that employing small quantities of labor h is more profitable than setting $l_h = 0$.

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

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

Proof of Proposition 3: Wage differentials across firms

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

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

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

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

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

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

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

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

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

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

$$l_h(w_{hg}, \cdot, \bar{a}_g) = \bar{a}_g \left(\frac{w_{hg}}{\omega_h} \right)^{\beta}$$

$$C_h(w_{hg}, \cdot, \bar{a}_g) = w_{hg} l_h(w_{hg}, \cdot, \bar{a}_g)$$

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

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

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

Also, rewrite the zero profit condition as:

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

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

The corresponding prices that lead to zero profits are:

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

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

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

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

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

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

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

□

Proof of Proposition 4: Supply shocks

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

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

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

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

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

Next, optimal consumption implies:

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

Combining both expressions:

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

Now we are ready to derive expressions for employment shares:

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

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

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

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

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

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

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

After careful manipulations, this expression can be rewritten as:

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

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

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

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

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

ing. Then:

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

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

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

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

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

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

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

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

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

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

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

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

It is enough to show that the expression is positive, ensuring that $\bar{x}'_h \geq \bar{x}_h$. Consider two cases. *If* $\bar{x}'_{h-1} \leq \bar{x}_h$, then use the first expression. The induction assumption guarantees positivity of the first term, and the integrand of the second term is positive because $\bar{x}_h < z^*$. *If instead* $\bar{x}'_{h-1} > \bar{x}_h$, the second expression is more convenient. There, all integrands are positive and the integration upper bounds are greater than the lower bounds.

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

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

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

Section 6: Wage inequality and sorting in Brazil

Proof of Proposition 6: Identification, estimation, and inference

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

The non-standard part of the proposed identification strategy is the inversion of region-specific parameters using a subset of the endogenous variables. Assumptions 3 and 4 imply that this condition is satisfied. See Appendix D.6 for a discussion of why invertibility is feasible in the theoretical model. Then, the model to be estimated is the one described in Assumption 5:

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

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

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

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

B Appendix to the theory

B.1 Definition of the task-based production function

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

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

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

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

B.2 Firm sizes and non-wage amenities

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

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

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

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

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

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

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

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

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

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

Proposition 3a.

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

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

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

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

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

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

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

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

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

B.3 Tinbergen's race

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

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

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

If these conditions are satisfied, the equilibrium under the new parameter set is identical to the initial equilibrium, except that prices for goods are uniformly lower: $p_g' = p_g/(1+\Delta_2)$ and $P' = P/(1+\Delta_2)$.[31]

Proof. The proof is simple once one notes that the difference between the two economies is a linear change of variables in the task space $x' = (1+\Delta_1)x$, coupled with a reduction in task demand by a factor of $(1+\Delta_2)$. Let $\bar{x}_h^g$ 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.

[31] Using the exponential-gamma parametrization, changes in comparative advantage functions and blueprints are equivalent to $\alpha_h' = \alpha_h/(1+\Delta_1)$, $\theta_g' = (1+\Delta_1)\theta_g$, $\kappa_g' = \kappa_g$, and $z_g' = (1+\Delta_2)z_g$.

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. $\square$

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

B.4 Discussion: missing minimum wage channels

In this appendix, I briefly discuss three minimum wage channels that are not present in this paper. The first is interactions of minimum wage with labor market concentration. By using a "monopsonistic competition" assumption and assuming that the β parameter is common across regions and skill levels, my model rules out the possibility that labor market power varies significantly across regions, as suggested by the empirical work of Azar et al. (2019). My assumptions also rule out the possibility that, by reallocating labor from smaller to larger firms, the minimum wage increases the labor market power of the latter—a channel that is present in the theoretical model of Berger, Herkenhoff and Mongey (2024).

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

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

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

The third channel not included in the paper are endogenous increases in worker efficiency in response to the minimum wage. Such "efficiency wage" effects may arise either because of reciprocity/fairness concerns, or because workers would choose to put in more effort at some utility cost to avoid being disemployed following a minimum wage hike. The second effect is the most important for the analysis of employment and wage effects. Coviello, Deserranno and Persico (2022) find support for that hypothesis in the US, but only for workers who are monitored more intensely.

The omission of these worker effort effects is likely insignificant because, to the extent that this channel exists, it should reflect in the minimum wage spike. That is because workers would put the necessary effort to be above the recruitment bar, but they do not need to put in so much effort that it overcomes the wage mark-down. Suppose I estimated an augmented model where a quantitatively significant number of workers bunch at the minimum wage due to endogenous effort. That would make the predicted spike bigger. However, in the structural estimation part of the paper, I find that the predicted spike is larger than the real one. Thus, adding that additional channel would decrease, rather than improve, the fit quality.

The final channel not included in the paper is endogenous responses in educational attainment caused by changes in the national minimum wage. Using data from the US, Smith (2021) documents that a ten percent increase in minimum wage lowers the probability of dropping out of high school by between four and ten percent, but only for teenagers in the low socio-economic status (SES) group (corresponding to 20% of teenagers).

Suppose such an effect is present in Brazil as well. In that case, the part of the effects of the minimum wage coming from this education channel will be assigned to education, instead of minimum wage, in the main counterfactual decomposition (Table 6). It is not obvious, however, that the results from the US are applicable in my context. For low-SES teenagers in Brazil, formality rates are meager, such that changes in the minimum wage may not substantially affect the opportunity cost of schooling. Even if it does, the magnitude of those effects is likely to be small compared to all other reasons why educational achievement has risen in Brazil, which I list in Appendix D.6.4. That said, one should interpret the minimum wage effects reported in Section 6.3 as not including this potential channel.

C Numerical implementation

C.1 Task-based production function

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

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

$$
\begin{aligned}
e_h(x) &= \exp(\alpha_h x) & \alpha_1 < \alpha_2 < \cdots < \alpha_{H-1} < \alpha_H \\
b_g(x) &= \frac{x^{\kappa_g - 1}}{z^g \Gamma(\kappa_g) \theta_g^{\kappa_g}} \exp\left(-\frac{x}{\theta_g}\right) & (z_g, \theta_g, \kappa_g) \in \mathbb{R}^3_{>0}
\end{aligned}
$$

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

$$
\bar{x}_h\left(\frac{f_{h+1}}{f_h}\right) = \frac{\log f_{h+1}/f_h}{\alpha_{h+1} - \alpha_h} \tag{13}
$$

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

$$
= \begin{cases} \sum_{m=0}^{\infty} \frac{\bar{x}_h^{\kappa_g} \exp\left(-\Upsilon_{hg}\bar{x}_h\right)\left(\Upsilon_{hg}\bar{x}_h\right)^m - \bar{x}_{h-1}^{\kappa_g} \exp\left(-\Upsilon_{hg}\bar{x}_{h-1}\right)\left(\Upsilon_{hg}\bar{x}_{h-1}\right)^m}{z_g \theta_g^{\kappa_g} \Gamma(\kappa_g + m + 1)} & \text{if } \Upsilon_{hg} \neq 0 \\ \frac{1}{z_g \kappa_g \Gamma(\kappa_g)} \left[(\bar{x}_h/\theta_g)^{\kappa_g} - (\bar{x}_{h-1}/\theta_g)^{\kappa_g}\right] & \text{otherwise} \end{cases} \tag{15}
$$

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

Expression 14 is simple to code and fast to run in software packages such as Matlab, where optimized implementations of the incomplete Gamma function are available.[32] When $\Upsilon_{hg} < 0$, that expression requires calculating complex numbers as intermediate steps. This is not a problem in Matlab.

If using complex numbers is not convenient or reduces computational efficiency, then the power series representation in 15 should be used. In my Julia implementation, I only use real (floating point) numbers. I use formulation 14 when $\Upsilon_{hg} \geq 0$, and 15 when $\Upsilon_{hg} < 0$. Another option, not used in this paper, is to change the normalization of α_h such that they are all non-negative.

Calculating the production function and its derivatives — that is, solving for output and marginal productivities given labor inputs — is not needed in the equilibrium computation nor in estimation. However, it might be useful for other purposes. Those numbers are obtained from a system of H equations implied by requiring that labor demand equals 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.2 Equilibrium

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

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

into a vector of $(H+1)\times G$ "residuals" which must be zero in an equilibrium:

1. Start with values for mean output $\bar{q}_g$ and task thresholds $\bar{\boldsymbol{x}}_g = \{\bar{x}_{1g}, \ldots, \bar{x}_{Hg}\}$ for the representative firms of each type, along with prices for goods p_g.
2. Use the compensated labor demand integral for the task-based production function to find average labor demands $\bar{l}_{hg}$ (Equation 3 in the text, or Equation 14 in Appendix C if using the exponential-Gamma parametrization).
3. Find marginal products of labor f_{hg} via the non-arbitrage conditions (2) and the constant returns to scale relationship $\sum_h f_{hg}\bar{l}_{hg} = \bar{q}_g$.
4. Employ the first order conditions of the firm (7) and (8) to find wages w_{hg} and rejection cutoffs $\underline{\varepsilon}_{hg}$, respectively.
5. Calculate relative consumption $Q_g/Q_1 = (p_g/p_1)^{-\sigma}$ and relative firm entry $J_g/J_1 = (Q_g/Q_1)/(\bar{q}_g/\bar{q}_1)$.
6. Pin down entry of firm type 1 (and thus all others) with entrepreneurial talent clearing: $J_1 = T/(\sum_g F_g J_g/J_1)$.
7. Calculate the real minimum wage as the sum of the minimum wage parameter and the price index implied by the guess of prices for goods.
8. For each $h \in \{1,\ldots,H\}$, integrate over ε to find labor supply and labor costs for each firm:
    (a) Choose minimum and maximum values $\varepsilon_{h,lowest}$ and $\varepsilon_{h,highest}$ for numerical integration, based on quantiles of the r_h distribution. In my application I use 0.001 and 0.999 as quantiles.

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

    (c) For each of those segments:

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

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

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

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

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

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

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

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

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

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

C.3 Simulating measures of wage inequality

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

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

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

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

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

Given these assumptions, the results of an AKM decomposition of log wages using simulated panel data are identical to running a two-way fixed effects model based on simulated data from one period, using a "worker id" indicator for each combination of (h, ε) and a "firm id" indicator for each good. Each observation is a (h, ε, g) cell. The regression is weighted by the share of the employed population in the corresponding cell. Finally, the estimated worker fixed effects are shrinked by the factor A, since they correspond only to the portable portion of productivity. The persistence parameter A is calibrated such that the R^2 of the simulated AKM regression is 0.9, about the same as the empirical regressions.[33]

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

D Appendix to the quantitative exercises

D.1 Sample sizes

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

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

The estimation of variance components follows the methodology proposed in Kline, Saggio and Sølvsten (2018), henceforth KSS. For each period (1998 and 2012), I use a three-year panel centered around the base year. The sample used for estimation is the largest leave-one-out connected set. This concept differs from the usual connected set in matched employer-employee datasets because it requires that firms need to be connected by at least two movers,

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

Table D1: Sample sizes for the 151 selected microregions

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

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

such that removing any worker from the sample does not disconnect this set. Table D1 presents the size of that largest connected set in each period.

The variance of log wages in the leave-one-out connected set is typically a bit smaller than the overall variance of log wages using the whole sample. To keep all measures in each region-time consistent with one another, I rescale the KSS variance components. Specifically, I multiply those components by the ratio of the overall variance of log wages in a region-time to the same variance in the leave-out connected set.

I implement the variance decomposition using the Julia code provided by KSS.[34] There are some implementation choices required in this estimation, stated below:

- Dealing with controls (year fixed effects): "Partialled out" prior to estimation.
- Maximum number of interactions: 300
- Sample selection: includes both movers and stayers. The leave-out procedure leaves a whole match out, not simply a worker-time observation.
- Number of simulations for JLA algorithm: 200

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

Table D2 shows additional versions of the validation exercises from Table 3. Panel A repeats the results from that table for quick referencing. Panels B and C show sample restrictions

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

where regions where the minimum wage binds more strongly are eliminated. That exercise tests whether the log-wage complementarities shown in Column (5) are mechanical consequences of minimum wages. That could be a concern since minimum wages censor the bottom of the wage distribution, and thus reduce the possibility of cross-firm wage differentials for unskilled workers.

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

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

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

D.4 Discussion: estimating the labor supply elasticity β

In this Appendix, I discuss the decision to calibrate the firm-level elasticity of supply parameter β instead of estimating it. I also discuss assigning half weight to the minimum wage "spike" in the estimation procedure since both decisions are related.

In principle, it is possible to estimate the firm-level elasticity of labor supply based on the size of the spike in log wage distributions, given the structure of the model. Comparing Panels A and B in Figure 3, one can note how β determines the range of workers who earn exactly the minimum wage conditional on posted wages and the distribution of abilities.

I did not pursue this strategy in the paper because the spike may be strongly affected by

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

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

Notes: See notes from Table 3.

economic factors not included in the model, such that this approach to identifying β is not as credible as state-of-the-art methods exploited in recent literature. One example of a theoretical channel not included in the model but with potentially significant consequences for the spike is the role of fairness considerations and relative earnings within the firm (see Footnote 26). Another is the possibility that, when deciding whether to work in the formal sector, workers have imperfect information about the level of earnings they will receive. In this case, a higher minimum wage may induce search efforts by workers who think there is a significant probability that they will earn exactly the minimum wage. That's because, for those workers, the expected earnings rise with the minimum. However, some may receive employment offers with wages a bit above the minimum. Because workers in my model have perfect information about their potential earnings in all firms in the economy, the positive employment effects of the minimum wage concentrate on the spike.

Indeed, when I estimate the model with a level of β similar to values estimated in recent papers and assign half weight to the spike target, I find that it over-estimates the size of the spike (see Table 5). Assigning full weight to the spike leads to minor improvements in that moment but significantly decreases the quality of fit in other dimensions. That's because the residual variance in the spike moment is relatively small compared to the wage inequality moments.

An alternative approach would be to estimate the β parameter by targetting the spike and assigning full weight to that target. I pursued this strategy in a previous version of this paper (available upon request). In that version, the predicted spike is closer to the measured one, though still larger. The estimated β was 10.2, consistent with the mechanism from Figure 3 being used to match the smaller spike. The conclusions from the main counterfactual exercises were very similar concerning the role of each factor in explaining changes in wage inequality in Brazil. As expected, the disemployment effects of the minimum wage are larger with a higher elasticity β.[35]

As a final note, the spike size may be different in the Brazilian context if one uses alternative data sets in the estimation procedure, which provides another potential reason to downweight that moment. For example, one can find more significant estimates of the minimum wage spike using the *Pesquisa Nacional por Amostragem de Domicílios* (PNAD) household survey. See Figure 2, Panel D in Derenoncourt et al. (2021) for a specific example. The discrepancies between the spike size in my paper and that in Derenoncourt et al. (2021) are primarily due to sample selection: my data excludes the smallest microregions in the country, where the minimum wage binds more strongly. I conjecture that they may also reflect rounding bias in the PNAD data compared to the RAIS. That is because the RAIS is an administrative data set managed by the government, and thus, firms are incentivized not to over-report wages since some mandatory contributions are proportional to the reported wage. The spike size in my paper is similar to that in Engbom and Moser (2022). That said, if it is true that the actual minimum wage spike is larger than what is implied by the RAIS data I use, then that would make the moments in the data closer to the values predicted by the fitted structural model.

[35] In the previous version of the paper, the minimum wage caused small, negative wage spillovers for workers in the middle of the productivity distribution. In the current version, the spillovers are also small but positive. The difference is not due to the β parameter. Rather, the previous model version imposed that low-wage firms had degenerate demand for low-complexity tasks. In contrast, those firms may demand high-complexity tasks in the current version. Thus, differences in returns to skill between firms were magnified in the previous version, strengthening the returns to skill channel of minimum wage effects.

D.5 Details on the parameterization of worker types

I set $H = 10$. The comparative advantage functions for these ten groups are fixed:

$$\begin{aligned} e_h(x) = & \quad \exp(\alpha_h x) \\ \alpha_h = & \quad -1 + \left(\sum_{h'=1}^{h-1} \frac{1}{h'}\right) / \left(\sum_{h'=1}^{H-1} \frac{1}{h'}\right) \end{aligned}$$

This formulation implies that the highest type the same productivity in all tasks, while the lowest type has $e_1(x) = \exp(-x)$. The values for intermediate types are such that if task thresholds are equally spaced for a firm g, then ratios of marginal products of labor between neighboring worker types are identical for all types. Although not essential, this property helps make skill premiums between groups reasonably uniform.

The exogenous number of workers N_h is determined by the observed shares of the adult population in each educational group $\hat{h} \in \{1,2,3\}$ (less than high school, high school, and college or more) according to the following probabilities:

$$\begin{aligned} \Pr(h=1|\hat{h}) &= \Phi\left(\frac{1.5-\mu_{\hat{h}}}{\rho_{\hat{h}}}\right) \\ \Pr(h|\hat{h}) &= \Phi\left(\frac{h+0.5-\mu_{\hat{h}}}{\rho_{\hat{h}}}\right) - \Phi\left(\frac{h-0.5-\mu_{\hat{h}}}{\rho_{\hat{h}}}\right) \qquad h \in \{2,\dots,9\} \\ \Pr(h=10|\hat{h}) &= 1-\Phi\left(\frac{9.5-\mu_{\hat{h}}}{\rho_{\hat{h}}}\right) \end{aligned}$$

where Φ is the cumulative distribution function of a standard Normal. Those probabilities resemble an "ordered Probit" model with thresholds $1.5, 2.5, \dots, 9.5$. I normalize $\mu_{\hat{h}=1} = 3$ and $\mu_{\hat{h}=3} = 8$. That is, the median worker with less than high school corresponds to $h = 3$, and the median college worker has $h = 8$. The comparative advantage of the median high-school worker is given by the estimated parameter $\mu_{\hat{h}=2}$. The model allows for dispersion in comparative advantage within an educational group, depending on the magnitude of $\rho_{\hat{h}}$.

The distribution of efficiency units ε within latent group h is a mean-zero Skew Normal:

$$\begin{aligned} r_{h,r,t}(\varepsilon) &= \frac{2}{S_{h,r,t}} \phi(\tilde{\varepsilon}) \Phi(\chi \tilde{\varepsilon}) \\ \tilde{\varepsilon} &= \frac{\varepsilon}{S_{h,r,t}} - \chi \sqrt{\frac{2}{\pi(1+\chi^2)}} \end{aligned}$$

$$S_{h,r,t} = \sum_{\hat{h}=1}^{3} \Pr(\hat{h}|h,r,t)\hat{S}_{\hat{h}}$$

where ϕ is the density of a standard Normal. The skewness is determined by χ. This degree of freedom helps the model fit the left tail of the wage distribution, which is essential for the effects of minimum wages. The parameters $\hat{S}_{\hat{h}}$ determine the dispersion of the efficiency units associated with each educational group $\hat{h}$.

The value of outside options is determined by:

$$\begin{aligned}
z_{0,h,r,t} &= \sum_{\hat{h}=1}^{3} \Pr(\hat{h}|h)\hat{z}_{0,\hat{h},r,t} \\
\text{where } \hat{z}_{0,\hat{h},r,t} &= \hat{z}^{HT}_{0,\hat{h},t} \cdot \hat{z}^{RH}_{0,r,\hat{h}} \cdot \hat{z}^{RT}_{0,r,t}{}^{\left(1+\Lambda \mathbf{1}\{\hat{h}=3\}\right)} \\
\text{and normalizing: } \hat{z}^{HT}_{0,\hat{h},t} &= 1 \quad \text{ if } t = 1998 \text{ or } \hat{h} = 2 \\
\text{and } \hat{z}^{RH}_{0,r,\hat{h}} &= 1 \quad \text{ if } \hat{h} = 2
\end{aligned}$$

The easiest way to understand that formulation is to focus on $\hat{z}_{0,\hat{h},r,t}$, the average value for educational group $\hat{h}$. It is determined by flexible education-time (HT), region-education (RH), and region-time (RT) components, which absorb confounders determining formal employment such as regional differences in the enforcement of labor regulation. The region-time shocks are allowed to have stronger or weaker effects on college workers ($\hat{h} = 3$) depending on the Λ parameter.

Once the outside options for the three educational groups are known, they can be transformed into outside options for latent worker groups, $z_{0,h,r,t}$, using the conditional probabilities $\Pr(\hat{h}|h,r,t)$ (similarly to the approach for the dispersion of efficiency units).

D.6 Discussion: identification of the structural model

D.6.1 Overidentification

In this section, I provide examples of how the economic model imposes strong constraints on the data, leading to an over-identified empirical model. First, note that I target seven inequality measures at the region-time level, but only allow three demand-side parameters to vary flexibly between regions. The model does not have enough degrees of freedom to simultaneously match, for each of those inequality measures, their time-specific means, variances, and how they correlate with the covariates we use.

Intuitively, the three demand parameters allowed to vary systematically between regions—blueprint complexity, relative entry costs, and relative taste for skill-intensive goods—are closely linked to the returns to college, the variance of establishment effects, and the covariance of worker and establishment effects, respectively. One could thus identify the 36 parameters $\delta^{d,t}$ by estimating period-specific regressions of those endogenous outcomes on the covariates in Equation (11), and then using the estimated regression coefficients as targets. After this, one could then consider using only a few selected moments of the other four inequality measures as targets to estimate the remaining parameters. Just trying to match their period-specific averages and variances would correspond to 16 moments, the same number of parameters left to be estimated after the $\delta^{d,t}$ terms are recovered. But there is important economic content not only in averages and variances of these inequality measures but also in how they correlate with each other. Below, in Section D.6, I discuss how the correlation between within-group variances of log wages and the covariance between worker and firm effects is important for identifying the elasticity σ. By trying to match all of the moments at the region-time level, the least-squares estimator uses that variation for identification.

Another example comes from the minimum wage bindingness measures. There is no guarantee that the model can replicate period-specific shares of workers at the minimum wage spike or up to 30 log points of the minimum wage, as they all depend fundamentally on a single parameter—the skewness of efficiency units of labor χ. A failure to match these bindingness measures would suggest misspecification of the distribution of skills or the economic mechanism that generates the minimum wage spike. The β parameter could, in principle, be estimated by targeting the size of the minimum wage spike (I thank an anonymous referee for this suggestion). In a previous discussion (Appendix D.4), I explained why this approach wass not pursued in the paper.

D.6.2 Avoiding incidental parameter bias

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

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

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

This is the motivation for specifying the regression-style models for the biased demand parameters, and using a subset of the endogenous outcomes to invert the flexible region-specific parameters. Three region-specific outside option parameters are recovered from formal employment rates in 1998, capturing heterogeneity in outside options at the microregion-education group level. The formal employment rate for high school workers in 2012 recovers the common region-specific shock to outside options for all groups. That could reflect, for instance, location-specific changes in the enforcement of labor regulations, which affects informality rates (Almeida and Carneiro, 2012).[36] Local TFP in each period is inferred from the minimum wage bindingness level. In effect, those endogenous outcomes are used as covariates, somewhat analogously to how empirical strategies such as Lee (1999) use measures of minimum wage bindingness as independent variables in regressions. An important difference is that the inversion procedure explicitly takes into account that observed bindingness depends on several other characteristics at the local level in addition to TFP, such as the educational distribution and labor demand characteristics.

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

As mentioned in the main text, the residuals $\boldsymbol{u}_r$ include misspecification in functional forms,

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

omitted variables, and sampling error. Functional form issues are not an issue, since the model can always match observed formal employment rates and levels of minimum wage bindingness by shifting the flexible productivity and outside option parameters. As for omitted variables, Assumption 3 can be viewed as a normalization: the "z" parameters to be inverted should be interpreted as encompassing all factors that drive formal employment and bindingness other than the wage index.

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

D.6.3 Identifying variation and instrumental variables analogy

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

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

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

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

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

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

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

The second piece of information is revealed by how the changes in the variance of log wages correlate with changes in skill premiums at the microregion.[37]

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

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

D.6.4 Identifying variation in the Brazilian context

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

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

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

More systematic efforts to expand schooling followed in the 1990's and 2000's. In 1996, a new law ("Lei de Diretrizes e Bases da Educação Nacional") established guidelines and attributed formal responsabilities to federal, state, and municipal agents in promoting the universalization of schooling. In 1995, the federal government created an effective system to collect school quality data at the national level ("Saeb"). Another system for evaluating secondary education followed in 1998 ("Enem"). In 2001, the federal government implemented a national cash transfer program conditional on school enrollment ("Bolsa-Escola", later incorporated into the "Bolsa Família" program). And starting in 2005, the "ProUni" program subsidizes low-income students who wished to attend private colleges and universities (public universities are tuition-free in Brazil, but few low-income students are able to pass the entry exams). This list of reforms and policies, which is not exhaustive, shows that that the rise in schooling achievement in Brazil was not an accident, nor should be viewed as "automatic" consequence of economic growth.[38]

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

Why does schooling rise faster in some regions, compared to others? It could be due to differences in policies implemented before 1998, or due to the fact that some national poli-

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

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

D.6.5 Threats to identification

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

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

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

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

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

D.7 Estimation

D.7.1 Numerical implementation of the loss function

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

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

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

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

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

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

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

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

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

D.7.2 Estimator and starting points

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

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

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

D.7.3 Estimates of demand parameters

Table D3 shows estimates of the $\delta_i^{d,t}$ demand-side parameters. The coefficients are reported for demeaned variables within each period, such that the constants capture the year-specific averages of the parameter transformations. Those averages point to an overall demand shock that combines three elements. First, task complexity requirements at the skill-intensive firms are increasing. Second, the relative entry cost ratio falls, such that it becomes relatively easier (from the point of view of entry inputs) to create skill-intensive firms. And third, there is a reduction in the relative taste for the skill-intensive good (corresponding to an exogenous average increase in the price for the low-skill good, since $\sigma \to \infty$ in the estimated model).

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

Table D3: Estimates of demand parameters

	$\log \theta_{2,r,t}$		$\log\left(\frac{F_{2,r,t}}{F_{1,r,t}}\right)$		$\log\left(\frac{\gamma_{2,r,t}}{1-\gamma_{2,r,t}}\right)$	
	1998	2012	1998	2012	1998	2012
Constant	-0.46	0.13	4.36	2.75	1.51	1.41
	(0.19)	(0.24)	(0.06)	(0.06)	(0.07)	(0.07)
Initial share high school	-0.04	1.24	-0.79	2.09	0.03	-0.61
	(0.49)	(1.10)	(1.90)	(1.88)	(0.63)	(0.86)
Initial share college	3.45	-4.75	3.05	6.74	-0.05	-0.50
	(0.76)	(1.43)	(2.86)	(3.77)	(0.85)	(0.88)
Initial share agriculture	0.14	1.18	0.39	-2.24	-0.23	-1.19
	(0.27)	(0.28)	(1.07)	(0.88)	(0.22)	(0.31)
Initial share manufacturing	-0.45	-2.27	-3.61	-3.05	-1.23	-1.04
	(0.27)	(0.44)	(0.94)	(0.69)	(0.28)	(0.23)
Current log min. wage minus mean log wage	0.41	0.32	-0.14	-1.03	0.02	-0.36
	(0.10)	(0.22)	(0.27)	(0.66)	(0.09)	(0.18)

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

of allowing for those correlations in the empirical model.

D.7.4 Benchmark regression models for quality of fit

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

Simple OLS: I run separate regressions for each moment. For all outcomes except the formal employment rates, the regressions include both time periods (302 observations in each). The regressors are time effects, share of adults with high school, share of adults with college, and the difference between the minimum wage and the mean log wage. I run two additional regressions, one for formal employment rates of adults with less than secondary, and the same outcome for adults with college education. Each uses data only for 2012 (151 observations each). The regressors are a constant, the lagged employment rate (i.e., for the same group in 1998), and the current formal employment rate for high school workers. That makes the employment rate regression comparable to the structural model, as it features region-education and region-time effects estimated by matching lagged participation values and the employment rates for high school workers. The model has a total of 51 parameters ($9 \times 5 + 2 \times 3$). This is the exact number of estimated parameters in the structural model, if I do not count the dispersion parameter that is found at the boundary of the parametric space ($\rho_{\hat{h}=1} = 0$).

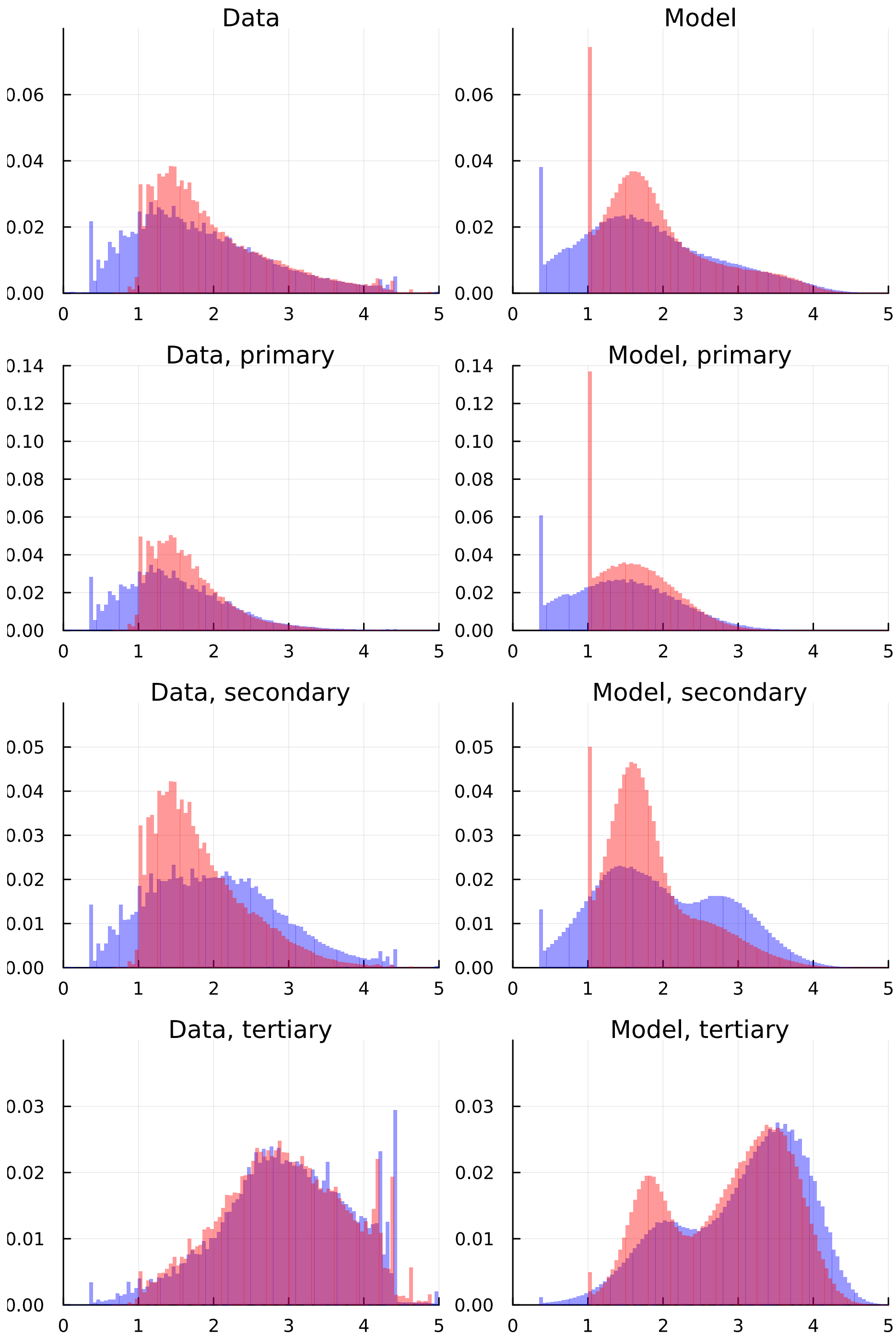
Large OLS: That model is an augmented version of the Simple OLS with more regressors and allowing for nonlinearities in the effect of the effective minimum wage. For outcomes other than employment rates, the regressors are time effects, current share of adults with high school, initial share of adults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

D.7.5 Additional measures of fit

In this section, I show additional measures of the quality of fit. I start with a comparison of the national histogram of log wages to that predicted by the model. The top panels in Figure D1 shows that the model closely approximates the real histogram, highlighting the quality of fit in both the inequality and relative formal employment across worker groups and regions. The other panels shows separate histograms for each educational group. Again, the model fits the data very well. The worst fit is for college workers. That is consistent with the lower quality of fit shown in Table 5 for the returns to college and the variance of log wages for college-educated workers. This lower quality of fit comes from the fact that thos moments have more residual variance in the data, and thus receive lower weight in the estimation procedure.

Next, I investigate whether the model is able to explain the cross-sectional variation within years. Table D4 shows that, for almost all target moments, the R2 metrics are positive. The only exception is the variance of log wages for college workers, which is the moment with the worst fit in the aggregate. Table D4 also shows the corresponding measures of fit for the two benchmark OLS models described in Appendix D.7.4. Similar to the discussion of the overall quality of fit, the Simple OLS model is comparable to the structural model. The Large

Figure D1: Distribution of log wages, data and model



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

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

	Model		Simple OLS		Large OLS	
	1998	2012	1998	2012	1998	2012
Moments	(1)	(2)	(3)	(4)	(5)	(6)
Wage inequality measures						
Secondary / less than secondary	-0.148	0.269	0.034	0.286	0.19	0.373
Tertiary / secondary	0.062	0.277	0.143	0.263	0.274	0.626
Within less than secondary	0.391	0.175	0.47	0.503	0.671	0.573
Within secondary	0.285	0.605	-0.055	0.319	0.311	0.661
Within tertiary	0.237	-0.247	0.302	0.021	0.381	0.279
Two-way fixed effects decomposition						
Variance establishment effects	0.168	-0.838	0.112	0.248	0.188	0.338
Covariance worker, estab. effects	0.289	0.501	0.241	0.531	0.349	0.705
Formal employment rates						
Less than secondary	1.0	0.908	1.0	0.915	1.0	0.959
Secondary	1.0	1.0	1.0	1.0	1.0	1.0
Tertiary	1.0	0.161	1.0	0.471	1.0	0.619
Minimum wage bindingness						
Log min. wage - mean log wage	1.0	1.0	1.0	1.0	1.0	1.0
Share < log min. wage + 0.05	0.673	0.407	0.634	0.509	0.848	0.731
Share < log min. wage + 0.30	0.832	0.828	0.697	0.617	0.864	0.878

Notes: This table displays the within-year quality of fit of the model, as measured by the R2 metric. The R2 can be negative if the model fits the data more poorly than a constant equal to the weighted mean of the target moment. The table also shows the quality of fit of the two benchmark OLS models described in Appendix D.7.4.

OLS model fits the data better in most dimensions, but again, the differences are not large with respect to the minimum wage bindingness measures, two-way fixed effects moments, and employment rate for workers with less than secondary.

The following exercise verifies the quality of fit regarding the spike and the share close to the minimum wage, separately by education. Those measures are not targeted by the estimation procedure, and thus serve as a test of whether the distributional assumptions on worker productivity seem warranted. In addition, if β varies strongly by skill, instead of being common as assumed in the model, then the data and the model would likely disagree regarding the relative size of the spike for different educational groups.

Table D5 shows that this is not the case. The overall pattern of a good fit for the spike in 1998, and an over-estimate in 2012, holds for all worker types. The fit of share close to the minimum wage is excellent for workers with secondary or less. For college workers, the R2

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

	Data		Model		R2
	1998	2012	1998	2012	Model
Moments	(1)	(2)	(3)	(4)	(5)
Less than sec., up to 5 log points	0.041	0.077	0.066	0.144	0.156
Secondary, up to 5 log points	0.022	0.05	0.02	0.062	0.616
Tertiary, up to 5 log points	0.005	0.009	0.002	0.007	0.434
Less than sec., up to 30 log points	0.117	0.287	0.148	0.298	0.809
Secondary, up to 30 log points	0.054	0.22	0.056	0.181	0.859
Tertiary, up to 30 log points	0.01	0.032	0.006	0.027	0.681

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

metric is close to zero, but the shares are very low to begin with. Thus, the lack of excellent quality of fit there is likely not very consequential for counterfactual analysis.

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

D.8 Counterfactuals

D.8.1 Additional decomposition outcomes

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

D.8.2 Demand shocks

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

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

	Data		Model		R2	Benckmark R2	
	1998	2012	1998	2012	Model	Simple	Large
Moments	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Wage inequality measures							
Secondary / less than secondary	0.478	0.131	0.473	0.122	0.726	0.723	0.764
Tertiary / secondary	0.978	0.953	0.992	0.929	0.043	-0.099	0.068
Within less than secondary	0.362	0.212	0.37	0.221	0.528	0.606	0.715
Within secondary	0.681	0.307	0.666	0.293	0.825	0.724	0.827
Within tertiary	0.755	0.612	0.744	0.639	0.274	0.364	0.423
Total variance of log wages	*0.687*	*0.465*	*0.686*	*0.46*	*0.702*		
Two-way fixed effects decomposition							
Variance establishment effects	0.117	0.048	0.119	0.028	0.413	0.481	0.524
Covariance worker, estab. effects	0.041	0.033	0.043	0.049	0.066	0.112	0.219
Variance worker effects	*0.428*	*0.311*	*0.428*	*0.299*	*0.539*		
Correlation worker, estab. effects	*0.193*	*0.256*	*0.188*	*0.545*	*-2.095*		
Formal employment rates							
Less than secondary	0.256	0.336	0.256	0.335	0.934	0.942	0.967
Secondary	0.425	0.509	0.425	0.509	1.0	1.0	1.0
Tertiary	0.534	0.632	0.533	0.637	0.836	0.917	0.936
Minimum wage bindingness							
Log min. wage - mean log wage	-1.237	-0.831	-1.237	-0.831	1.0	1.0	1.0
Share < log min. wage + 0.05	0.046	0.062	0.065	0.104	0.277	0.486	0.688
Share < log min. wage + 0.30	0.121	0.235	0.148	0.24	0.834	0.671	0.857

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

component in isolation.

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

Changes in the technologies used by formal firms may not be the only reason why the $z_{0,h}$ parameters changed. Another example, previously mentioned in the paper, would be changes

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

Outcome	Base value (1)	All Changes (2)	Individual effects: S (3)	D (4)	M (5)	Interactions S+D (6)	S+M (7)	D+M (8)	Triple (9)
Panel A: Inequality between and within groups									
Between groups: 2/1	0.48	-0.31	-0.02	-0.20	-0.06	0.02	0.01	-0.05	0.00
Between groups: 3/2	0.98	-0.03	-0.04	-0.00	-0.07	0.03	0.01	0.04	-0.01
Within group: 1	0.40	-0.17	-0.00	-0.04	-0.13	-0.02	0.02	0.02	-0.00
Within group: 2	0.67	-0.34	-0.04	-0.29	-0.09	0.02	0.01	0.06	-0.00
Within group: 3	0.70	-0.06	-0.09	-0.01	-0.05	0.04	0.01	0.03	-0.01
Panel B: Formal employment rates									
All workers	0.32	0.11	0.04	0.10	-0.02	-0.00	0.01	-0.01	0.00
Group 1	0.27	0.07	0.01	0.10	-0.03	-0.01	0.01	-0.02	-0.00
Group 2	0.44	0.07	-0.01	0.09	-0.02	-0.00	0.00	0.01	-0.00
Group 3	0.54	0.09	-0.01	0.10	-0.00	0.01	-0.00	-0.00	0.00
Panel C: Minimum wage bindingness									
Log min. wage - mean log wage	-1.42	0.50	-0.24	0.06	0.51	0.08	0.04	0.04	-0.01
Share < log min. wage + 0.30	0.11	0.09	-0.04	-0.04	0.19	0.01	-0.03	-0.01	0.01

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

in the enforcement of labor regulations that make the formal sector more or less appealing to some workers. Whether such a shock is on the supply or demand side is a matter of interpretation—in this paper, I classify them as demand shocks.

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

Another way of framing this issue is that, to identify the independent effect of specific de-

Table D8: Decomposition of demand shock

Outcome	All demand shocks (1)	Task demand (2)	Consumer taste (3)	Entry cost (4)	TFP and outside opt. (5)
Panel A: Inequality and sorting					
Mean log real wage	-0.06	-0.18	0.05	0.11	-0.04
Variance of log wages	-0.18	-0.28	-0.01	0.10	0.01
Corr. worker, estab effects	0.19	0.03	-0.06	0.21	0.00
Panel B: Inequality between and within groups					
Between groups: 2/1	-0.20	-0.31	0.01	0.10	0.01
Between groups: 3/2	-0.00	-0.15	-0.09	0.21	0.03
Within group: 1	-0.04	-0.06	0.02	0.00	-0.00
Within group: 2	-0.29	-0.38	-0.02	0.08	0.02
Within group: 3	-0.01	0.13	-0.16	0.03	-0.01
Panel C: Two-way fixed effects decomposition					
Variance of log wages	-0.18	-0.28	-0.01	0.10	0.01
Var. worker effects	-0.06	-0.15	0.01	0.08	0.01
Var. estab. effects	-0.09	-0.06	-0.01	-0.02	-0.00
2×Cov. worker, estab	-0.02	-0.04	-0.01	0.03	0.00
Var. residuals	-0.01	-0.02	-0.00	0.01	0.00
Panel D: Formal employment rates					
All workers	0.10	0.07	0.03	-0.12	0.12
Group 1	0.10	0.08	0.03	-0.13	0.11
Group 2	0.09	0.06	0.03	-0.13	0.14
Group 3	0.10	-0.04	0.02	-0.07	0.19

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

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

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

Table D9: Effects of a small increase in the minimum wage

Prod. decile	Pop. share	Base wage	Mean wage changes: Monops.	Ret. sk.	Gen. eq.	Base emp.	Emp. elasticities w.r.t.: Min.	Mean	·, monops.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.15	1.24	0.07	-0.00	0.00	0.21	-0.49	-0.98	-0.94
2	0.12	1.78	-0.00	-0.00	0.00	0.27	-0.01		
3	0.11	2.35	-0.00	-0.00	0.00	0.28	-0.01		
4	0.11	2.97	0.00	-0.00	0.00	0.29	-0.01		
5	0.10	3.75	-0.00	-0.00	0.00	0.31	-0.01		
6	0.10	4.76	0.00	-0.00	0.00	0.33	-0.01		
7	0.09	6.11	-0.00	-0.00	0.00	0.37	-0.00		
8	0.08	8.13	-0.00	0.00	0.00	0.40	-0.00		
9	0.07	11.91	0.00	0.00	0.01	0.45	0.00		
10	0.06	25.04	0.00	0.01	0.01	0.50	0.00		

Notes: This table is similar to Table 8 in the main text, except that it describes the effects of a 10 log point increase in the minimum wage.

D.8.3 Effects of a small increase in the minimum wage

The increase in the Brazilian minimum wage between 1998 and 2012 is very large. One may be interested in the predicted effects of a small increase in the minimum wage. To that end, I generate a copy of Table 8 based on a change in the log minimum wage of 10 log points, instead of 66.1 log points as in the main exercise.

Table D9 shows that, with a small change in the minimum wages, its effects are fully concentrated on the lowest decile of worker productivity. For that group, wage effects come from the truncation, censoring, and reallocation effects which compose the "monopsony" channel. In other words, the minimum wage shock is too small to introduce quantitatively significantly changes in returns to skill, firm entry, and prices. Notably, the elasticities of employment with respect to either the minimum wage increase or the change in the mean wage are similar to those in Table 8.

D.8.4 Why do regressions find no employment effects of minimum wages in Brazil?

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

Engbom and Moser (2022) exploit variation in the "effective minimum wage," that is, the log of the national minimum wage minus the median log wage in each state-time combination, which they refer to as the Kaitz-50 index.[39] They run regressions of formal employment on the effective minimum wage, its square, and controls. This approach has a long tradition in labor economics, going back at least as far as Neumark and Wascher (1992, who used the minimum wage relative to the mean in the state-year instead of the median). In the specification they report in the paper, Engbom and Moser (2022) use state fixed effects and state-specific time trends as controls.

In Haanwinckel (2024), I discuss the necessary identification assumptions required for this design to work well in contexts with no regional variation in minimum wage laws, such as Brazil. I find that, in general, this approach is subject to large biases if the minimum wage causes changes in the median wage, even if these effects are small at the individual level, and even if these effects have mean zero. In addition, I show that the design relies crucially on unobserved shocks that move the entire log wage distribution to the left or to the right, while being orthogonal to changes in the shape of the distribution or in employment measures. Even small correlations between these unobserved shocks and other determinants of the log wage distribution can introduce large biases. In other words, if the econometrician cannot pinpoint what is the quasi-experimental source of variation that makes the minimum wage bind more in some regions than in others, then it is possible that the effective minimum wage design may be subject to substantial misspecification biases.

In the remainder of this appendix, I discuss specific factors in the Brazilian context that complicate the identification of minimum wage effects in reduced form.

One challenge with the effective minimum wage design is that the median wage, used to construct the effective minimum wage, is an endogenous object. As emphasized in this paper, wages are determined at the local labor market level by a combination of region-specific supply and demand parameters. They correlate with each other, and also with local TFP. That introduces correlations between those factors and the Kaitz-50 index. On the supply side, I find that microregion-level changes in educational achievement are positively correlated with the change in the Kaitz-50 index, which is somewhat surprising. In addition, Table D3

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

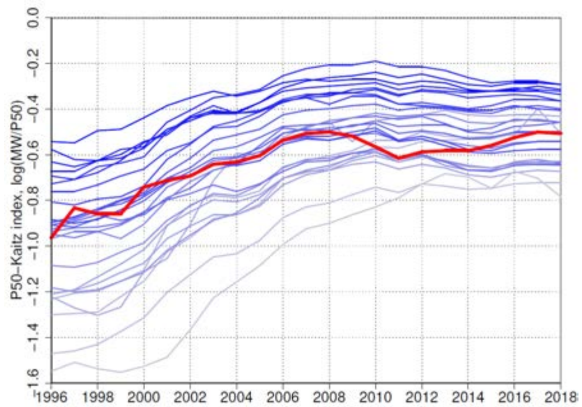


Figure D2: Variation in effective minimum wages at the state-time level
Notes: This is a copy of Appendix Figure B.10 in Engbom and Moser (2022). It shows the variation used to identify the effects of minimum wages on employment in Brazil.

shows that the current Kaitz index is a statistically significant predictor of demand-side parameters after controlling for initial characteristics at the microregion, with coefficients that vary between years. Those correlations may introduce omitted variable bias because all of those supply and demand shocks have large effects on employment rates even in the absence of minimum wage changes, as shown in Table D7. And because they correlate in differences, not only in levels, their effect is not absorbed by the state fixed effects.

To tackle those time-varying confounders, Engbom and Moser (2022) include region-specific time trends in regression models with many periods (the panel is at the yearly level, from 1996 through 2018). Intuitively, the assumption behind this approach is that the influence of these confounders on employment is well approximated by the linear trends, while the influence of the minimum wage is nonlinear. Another way of visualizing that assumption is: if one takes time differences two times for both employment rates and the Kaitz index, then the relationship between those transformed variables should reflect the impact of locational changes in the latent distribution of wages (what could be described as TFP), not the direct effect of compositional changes between groups that have different intrinsic employment rates or of biased demand shocks that affect the latent productivity distribution and employment rates differently from a locational shift.

For the minimum wage, the non-linear part of the variation comes from faster minimum wage growth in the first half of the sample. This is evident from Figure D3, which is a copy of Appendix Figure B.10 from Engbom and Moser (2022)). The red thick line shows the national average for the Kaitz-50 index, while the blue lines show the Kaitz-50 index for

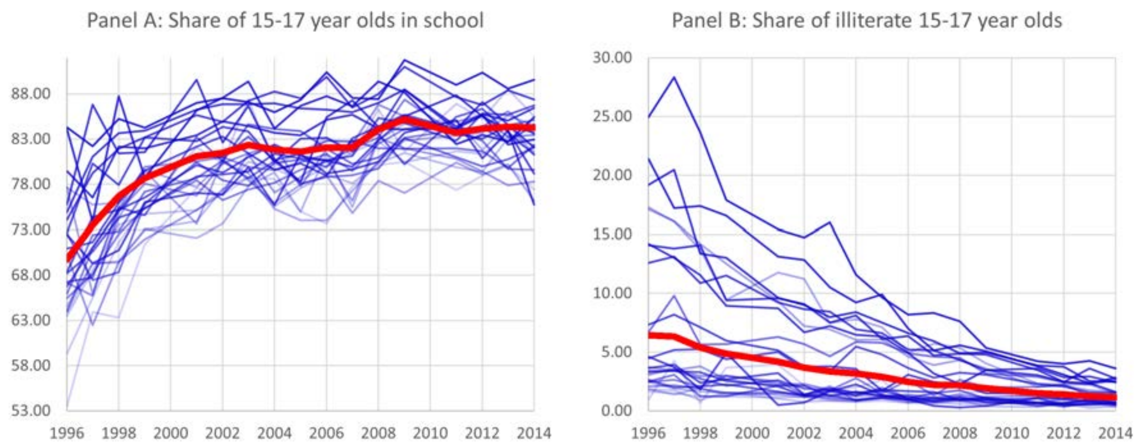


Figure D3: Evolution of educational outcomes by state
Source: PNAD survey. The series were obtained using the IpeaData online tool (available at http://www.ipeadata.gov.br).

each state.

Is the variation in minimum wages more nonlinear than the supply and demand shocks affecting the Brazilian economy? Below, I argue that this is not the case. Figure D3 shows two metrics related to the supply of young educated adults: the share of those between 15 and 17 who are in school, and the share of those between 15 and 24 who can read. Both graphs show steeper slopes early in the period, similarly to the minimum wage graph. This is an important issue, since formal employment rates vary dramatically by educational level. And among all adults, the young are more likely to be affected by the minimum wage.

A similar argument can be made for demand shocks. The variation in international commodity prices, shown in Figure D4, suggests that the influence of demand shocks may be much less smooth and monotonic than the impacts of minimum wages. In addition, Figure 2 in Costa, Garred and Pessoa (2016) shows that trends in Brazilian imports from, and exports to China are also nonlinear. The export trends is nonmonotonic, and considerably further from the a line than trends in the Kaitx-50 index. Costa, Garred and Pessoa (2016) goes on to show that shocks to Chinese supply and demand have significant labor market effects at the microregion level.

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

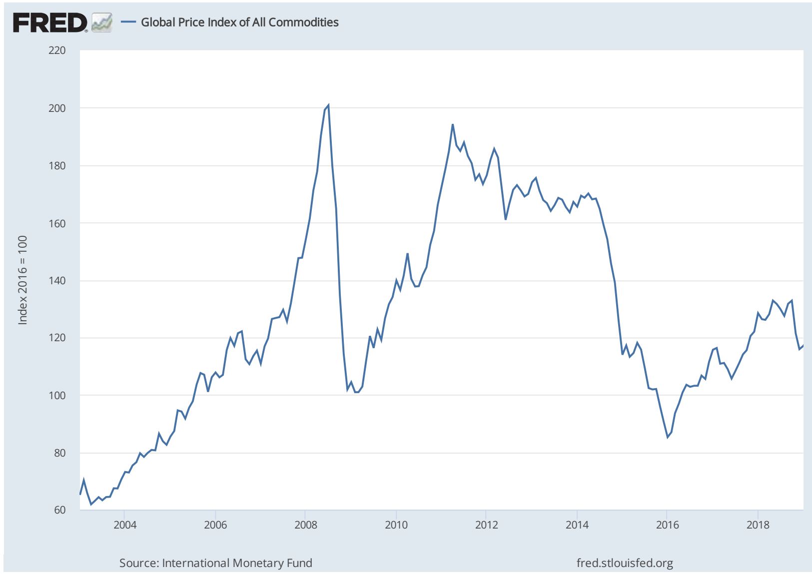


Figure D4: Global Price Index of All Commodities

a quasi-experiment that manipulates the minimum wage independently of other factors. This is a specific example of broader issues that I document in Haanwinckel (2024). That said, the analysis in that paper does suggest that time fixed effects should in general be included in this design.

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

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

Appendix Table D4 provides a test of whether the strong disemployment effects are rejected by the data. Specifically, if the employment effects predicted by the model were strongly at odds with what was observed at the microregion level, one would expect the R2 metric for the formal employment rate of workers with less than secondary in 2012 to be bad. Instead, it is 0.908.

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