Essays on Wage Inequality and Employment Informality

by

Daniel Haanwinckel Junqueira

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Patrick Kline, Co-chair
Professor Frederico Finan, Co-chair
Professor Andrés Rodríguez-Clare
Professor David Card
Professor Thibault Fally

Spring 2019

Essays on Wage Inequality and Employment Informality

Abstract

Essays on Wage Inequality and Employment Informality

by

Daniel Haanwinckel Junqueira

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Patrick Kline, Co-chair

Professor Frederico Finan, Co-chair

This dissertation develops models to understand changes in the wage distribution and employment informality rates. Common themes are imperfect competition in the labor market, firm and worker heterogeneity, imperfect substitution between different levels of skill in production, and minimum wages. In each chapter, I present a model, discuss its properties, estimate it using Brazilian data, and use it for counterfactual analysis.

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

exercise also illustrates how responses to supply and demand shocks differ qualitatively from those predicted by models with a representative firm.

The second chapter is coauthored with Rodrigo R. Soares. We develop a search model of informal labor markets with worker and firm heterogeneity, intra-firm bargaining with imperfect substitutability across types of workers, and a comprehensive set of labor regulations, including minimum wage. Stylized facts associated with the informal sector, such as smaller firms and lower wages, emerge endogenously as firms and workers decide whether to comply with regulations. Imperfect substitutability across types of workers and decreasing returns to scale enable the model to reproduce empirical patterns incompatible with existing frameworks in the literature: the presence of skilled and unskilled workers in the formal and informal sectors, the rising share of skilled workers by firm size, and the declining formal wage premium by skill level. These features also allow us to analyze the equilibrium responses to changes in the demand and supply of different types of labor. We estimate the model using Brazilian data and show that it closely reproduces the decline in informality observed between 2003 and 2012. The change in the composition of the labor force appears as the main driving force behind this phenomenon. We illustrate the use of the model for policy analysis by assessing the effectiveness of a progressive payroll tax in reducing informality.

Para Thaís, Isabela e Lucas.

Contents

List of Figures

List of Tables

Acknowledgments

I am most indebted to my advisors. You taught me more than you know. The four of you share many qualities I admire: you are intelligent, curious, measured, intolerant of nonsense, generous with time, and best of all, patient enough to put up with me year after year. I will pay it forward by doing my best as an advisor to my future students.

Pat: I deeply admire and benefited from your broad knowledge, sharp criticism, and dedication to your students. You made me a better economist by always pushing me towards excellence.

Fred: I can't count the number of times you said something along the lines of "I'm not an expert on this, but...", only to continue the sentence with fantastic advice. You were there for me from day one, giving me solid feedback on my research and soft guidance on how to navigate academic life.

Andrés: I am glad I followed your advice and enrolled in 280A in my second year. That set me on a path between fields that made my graduate experience richer. I learned a lot from your unique ability to capture the essence of an argument in an elegant model and use it to find connections that most would miss.

Dave: I am proud to be among the distinguished and large host of economists trained by you. I feel ashamed when I get carried away by math and lose my footing in the data. That comes from you, both from what you taught me and indirectly through your influence over the profession.

I also thank my coauthor and former advisor Rodrigo R. Soares for continued guidance and thoughtful discussions.

I am grateful for my amazing colleagues. I learned a lot from interacting with all of you, from the mundane (how to do something in R, how to submit this form, how to say this or that in English, how to solve this question in the problem set) to the sublime (mostly discussions about econometrics). Special thanks to those who endured me the longest: Ale, Avner, Juli, Murilo, Nick, and Yotam.

In addition to those mentioned above, my research benefited from extensive discussions with and comments from many other faculty and students at Berkeley, especially Ernesto Dal Bó, Ben Faber, Thibault Fally, Cecile Gaubert, Rupal Kamdar, Byoungchan Lee, John Loeser, Isabela Manelici, Ted Miguel, Piyush Panigrahi, Walker Ray, Evan Rose, Raffaele Saggio, Jon Schellenberg, David Schönholzer, Kevin Todd, José Vásquez, and Chris Walters. The second chapter also benefited from comments from people elsewhere: Pedro Carneiro, Gustavo Gonzaga, Marc Muendler, Renata Narita, Gabriel Ulyssea, Eduardo Zilberman, and seminar participants at IPEA-Rio, PUC-Rio, the 8^{th} IZA/World Bank Conference on Employment and Development, the 2014 PACDEV Conference, the 2^{nd} LACEA Labor Network Meeting, the 1^{st} Bay Area Labor and Public Finance Graduate Conference, the 37^{th} SBE Annual Meeting, and the 20^{th} LACEA Annual Meeting.

The research in the first chapter would not have been possible without data from the Brazilian Ministry of Labor. I thank Lorenzo Lagos and David Card for sharing code to read that data. I am also thankful to Data Zoom, developed by the Department of Economics at PUC-Rio, for providing code for reading IBGE microdata.

Finally, if I managed to achieve anything, it's because I had family and friends around. I thank my mother Fátima, my siblings Priscila and Beto, and close friends (especially Bruno) for being supportive when I decided to quit my job in pursuit of an academic career and in every moment since. I thank my in-laws Vera, Guilherme, and Gabriel for accepting me in your family and for trying to make our life a little more comfortable whenever you could. And for Thaís, Isabela, and Lucas, a finite number of words will never be enough. Life has an entirely different meaning now that I have you. Thank you for being here with me.

Chapter 1

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

1.1 Introduction

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

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

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

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

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

I derive two results on how this economy responds to structural shocks. First, the model admits a balanced growth path where technical progress is conceptualized as shifts towards more complex tasks in the production of all goods (a form of skill-biased technical change) accompanied by increased productivity. If skill levels and minimum wages rise in tandem, the shape of the wage distribution is not affected. Second, substitution in consumption creates an additional channel through which wages are affected by imbalances in the race between supply, demand, and institutions. Imbalances have direct effects on the costs of labor: for example, a higher minimum wage makes low-skilled workers more expensive relative to high-skilled workers. The ensuing changes in the prices of different goods drive substitution in consumption. In the minimum wage example, consumers substitute away from products that are intensive in low-complexity tasks, which are produced by firms using mostly low-skilled labor. Finally, because the set of tasks being produced in that economy has changed, marginal productivity gaps are affected. Without substitution between goods, a plot of the impact of the minimum wage on quantiles of the wage distribution is a decreasing curve; with substitution, it becomes less steep and can change to a U-shaped curve.

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

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

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

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

1.2 The task-based production function

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

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

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

I start this section by defining the production function and solving the managerial problem of assigning workers to tasks. The second subsection discusses cost minimization and

[1] There exists a parallel between the task-based production function developed here and models of hierarchical firms (Garicano, 2000; Garicano and Rossi-Hansberg, 2006; Antràs, Garicano and Rossi-Hansberg, 2006; Caliendo and Rossi-Hansberg, 2012), once one reinterprets tasks in my model as "problems" in those models. The key difference is that I ignore costs of information transmission within the firm, adding tractability by simplifying the assignment of workers to problems/tasks.

shows how this structure generates differences in skill intensity between firms. The third subsection derives and explains distance-dependent substitution. The final subsection presents the parametric version that is employed in the quantitative exercises of this paper. All proofs are in Appendix A.1.

Setup, definitions, and the assignment problem

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

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

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

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

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

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

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

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

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

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

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

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

The definition of the production function assumes positive input of the highest worker type. This assumption simplifies proofs and ensures the well-behaved derivatives, while not being restrictive for the applications in this paper. In general, blueprints might require at least one worker of a minimum worker type $\underline{h}$ — if none is available, lower types have zero marginal productivity. This property might be useful for models of endogenous growth and innovation.

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

[2] If tasks are freely traded, the model makes no predictions about sorting of workers to firms. A less extreme assumption — e.g. formally modeling output losses from assembling tasks produced at different firms — could be used for studying the boundaries of the firm and the effects of outsourcing.

[3] $\forall x, \tau \in \mathbb{R}_{>0}, \exists \delta \in \mathbb{R}_{>0}$ such that $x' \in [x, x+\delta) \Rightarrow |m_h(x) - m_h(x')| < \tau$.

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

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

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

$$\frac{e_{h+1}(\bar{x}_h)}{e_h(\bar{x}_h)} = \frac{f_{h+1}}{f_h} \quad h \in \{1, \ldots, H-1\} \tag{1.1}$$

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

Lower types specialize in low complexity tasks and vice-versa. Equation (1.1) means that the shadow cost of using neighboring worker types is equalized at the task that separates them. This result is useful for obtaining compensated labor demands, as described in the next subsection.[4]

Compensated labor demand and sorting of workers to firms

To study the properties of this production function, I start by considering its implications in a competitive labor market, where the cost of acquiring efficiency units of each type is given by $\boldsymbol{w} = \{w_1, \ldots, w_H\}$. When firms choose labor quantities by minimizing production costs, marginal productivity ratios equal wage ratios. It then follows from Equation (1.1) that:

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

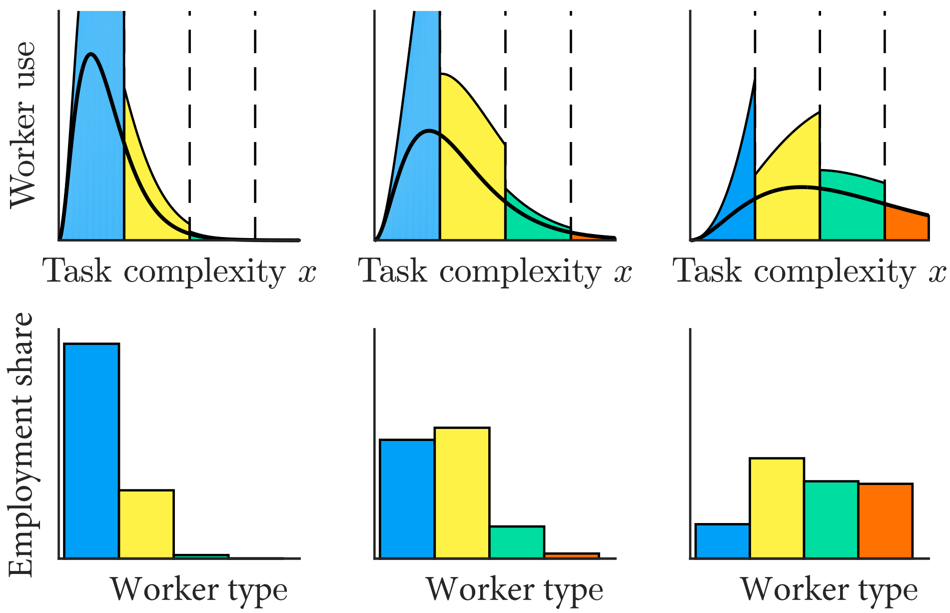
Because the ratio on the left-hand side is strictly increasing in $\bar{x}_h$, this expression pins down all task thresholds as functions of wage ratios and comparative advantage functions. Since neither are firm-specific, thresholds are common across firms in competitive economies.

The compensated labor demand is then given by:

$$l_h(q, b_g, \boldsymbol{w}) = q \int_{\bar{x}_{h-1}(\boldsymbol{w})}^{\bar{x}_h(\boldsymbol{w})} \frac{b_g(x)}{e_h(x)} dx \tag{1.2}$$

[4] In general, the task-based production function and its derivatives do not have simple closed-form representations. If one needs to evaluate output and marginal productivities as a function of labor inputs, first solve the system of H compensated labor demand equations (1.2) on q and the $H-1$ thresholds. Next, use equation (1.1) to calculate marginal productivity gaps. Finally, use the constant returns relationship $q = \sum_h l_h f_h$ to normalize marginal productivities.

Figure 1.1: Compensated labor demand in competitive labor markets



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

Figure 1.1 illustrates how differences in blueprints reflect into differences in the internal workforce composition of firms. The graphs at the top show the compensated labor demand integral above. The heavy, continuous line is the blueprint, which varies across graphs (becoming more intensive in high complexity tasks from left to right). The vertical dashed lines are the thresholds, which are common for all graphs. The colored areas are the labor demand integrals. The compensated labor demand is shown again in the bottom row, in the form of blueprint-specific wage distributions within the firm (weighted by efficiency units).

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

The concept of firms in this model is significantly different from that in the literature on labor market sorting (Shimer, 2005; Eeckhout and Kircher, 2011; Gautier, Teulings and van Vuuren, 2010; Gautier and Teulings, 2015; Lise, Meghir and Robin, 2016; Grossman, Helpman and Kircher, 2017; Lindenlaub and Postel-Vinay, 2017; de Melo, 2018; Eeckhout and Kircher, 2018). Most models in this literature focus on sorting of workers to *jobs* (or, equivalently, to firms that employ exactly one worker). Even in the ones with a concept of large firms, such as Eeckhout and Kircher (2018), any degree of within-firm wage dispersion is a sign of inefficiencies introduced by search frictions; if markets are competitive, each firm hires workers of a single type. I contribute to this literature by introducing a more realistic concept of firms as bundles of jobs (tasks in my model), coupled with a technology to acquire workers. In addition to having welfare implications, this distinction is relevant for quantitative studies where model predictions are matched to firm-related moments, such as the between-firm share of wage inequality or variance decompositions from AKM regressions.

Substitution patterns and distance-dependent complementarity

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

Proposition 1 (Curvature of the production function)**.** *The task-based production function is concave, has constant returns to scale, and is twice continuously differentiable with strictly positive first derivatives. I denote by $c = c(\boldsymbol{w}, q)$ the cost function, use subscripts to denote derivatives regarding input quantities or prices, and omit arguments in functions to simplify the expressions. Then, for any pair of worker types h, h' with $h < h'$:*

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

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

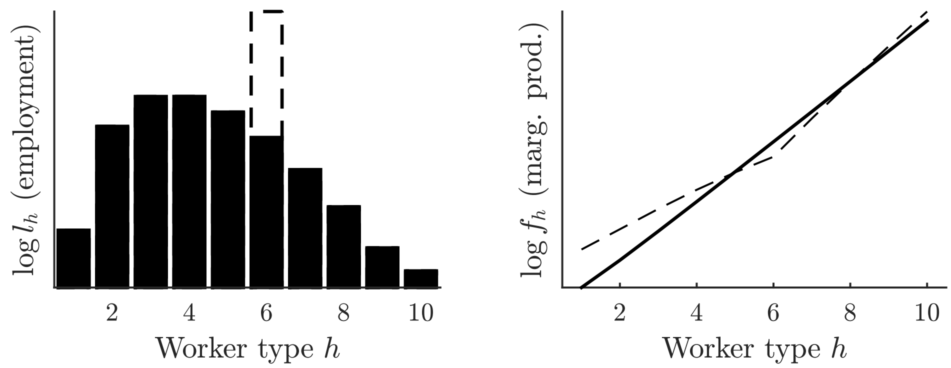
$$\textit{where} \quad \rho_h = b_g(\bar{x}_h) \frac{f_h}{e_h(\bar{x}_h)} \left[\frac{d}{d\, \bar{x}_h} \ln\left(\frac{e_{h+1}(\bar{x}_h)}{e_h(\bar{x}_h)} \right) \right]^{-1}$$

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

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

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

Figure 1.2: Distance-dependent complementarity



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

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

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

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

Close types perform similar tasks and are net substitutes; distant types perform very different tasks and are complements. The distance-dependent complementarity pattern is illustrated in Figure 1.2. The left panel shows baseline log employment by worker type (in solid bars) and a shock to employment of workers of type 6 (bar with dashed contour). The right panel shows baseline log marginal productivities (solid line) and marginal productivities after the employment shock (dashed line). Workers of type 6 suffer the largest relative decline in marginal productivity, followed by neighbor types 7 and 5. Marginal productivities increase for types that are further away, both low-skilled and high-skilled.

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

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

A minimum wage that is only binding for the lowest type will have spillover effects along the wage distribution, reducing wage gaps between any pair of worker types. This is in contrast to the CES case, where a small increase in the minimum wage would raise wages for the lowest type but keep wage gaps between other types constant. This property is useful for matching reduced form estimates of minimum wage spillovers (see Figure 1.7).

Parametric example

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

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

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

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

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

Appendix A.2 presents the mapping between marginal productivity gaps and task thresholds in this parametrization, as well as formulas for compensated labor demand integrals in terms of incomplete Gamma functions. These formulas are useful because they dispense with numerical integration, improving computational performance. Incomplete Gamma functions are readily available in software packages commonly used by economists.

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

This approach allows for modeling firm-level differences in skill intensity, skill dispersion, and productivity with a small number of parameters, while ensuring sensible substitution patterns within all firms. To understand the economic content behind those parametric restrictions, consider an example of two firms in the retail sector. One, with a low θ_g, is a small local shop, while the other, with a high θ_g, is a large online retailer. In the first one, most tasks are of low complexity, measured in terms of how they benefit from schooling: stocking shelves, operating the register, and cleaning. In those tasks, workers with little formal education can substitute for others with a college degree. Because workers with a college degree cost much more, that first firm mostly hires less educated workers. In contrast, the online retailer is intensive in tasks such as web design, system administration, and business analytics, where college-graduated workers usually perform much better. This is why those firms find it profitable to use a more skilled workforce.

1.3 Markets and wages

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

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

Factors, goods, technology, and preferences

Consider an economy with $\boldsymbol{N} = \{N_1, \ldots, N_H\}$ workers of each type h, and a large number of entrepreneurs. Entrepreneurs own entrepreneurial talent, whose total stock in the economy is T and which is used to create firms. The model is static.

There are G final goods in this economy, interpreted as differentiated varieties within an industry.[6] An entrepreneur j may set up a firm producing one good $g \in \{1, \ldots, G\}$ or

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

not enter at all. Setting up a firm requires a fixed cost F_g, paid in units of entrepreneurial talent. Once that cost is paid, the entrepreneur receives the blueprint b_g and a random draw of workplace amenities a_j from a good-specific distribution with strictly positive support and a finite mean $\bar{a}_g$. The role of workplace amenities will be explained below. Hiring and production decisions are done after the amenities draw is observed, as discussed in the next subsection.

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

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

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

$$\sum_g J_g F_g = T \tag{1.3}$$

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

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

$$\begin{gathered}
U^E\left(\{Q_g\}_{g=1}^G\right) = C\left(\{Q_g\}_{g=1}^G\right) \\
U^L_{hi}\left(\{Q_g\}_{g=1}^G, j\right) = C\left(\{Q_g\}_{g=1}^G\right)\left[a_j \exp\left(\eta_{ij}\right)\right]^{\frac{1}{\beta_h}} \\
\text{where } C\left(\{Q_g\}_{g=1}^G\right) = \left[\sum_{g=1}^G Q_g{}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{1-\sigma}} \\
\text{and } \eta_{ij} \sim \text{Extreme Value Type I}
\end{gathered}$$

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

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

Continuing with the example from Section 1.2, the small local shop and the large online retailer are interpreted as differentiated varieties in the retail sector, with elasticity of substitution σ. In addition to task requirements, these firms might differ in entry costs and the average level of amenities. The online retailer might require substantial capital investment or managerial input to set up, justifying high entry costs F_g. If $\bar{a}_g$ is higher for the large retailers, then they are also more desirable workplaces on average.

Labor markets, the problem of the firm, and equilibrium

Labor markets are based on the model of Card et al. (2018), where firms compete monopsonistically for labor. Each worker is characterized by its type $h \in \{1, \ldots, H\}$ and a quantity of efficiency units of labor ϵ. The distribution of efficiency units of labor across workers of type h is continuous with density $r_h(\cdot)$ and support over the real line.[7] Throughout this section, it is important to distinguish between quantities of workers, denoted by n, and quantities of labor, denoted by l. Worker earnings are denoted by y, while prices for labor are denoted by w.

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

Firm-level labor supply and labor costs

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

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

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

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

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

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

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

$$V_{ih}(\epsilon, j) = \underbrace{\frac{\mathbf{1}\{\epsilon \geq \underline{\epsilon}_{hj}\} \max\left\{\epsilon w_{hj}, \underline{y}\right\}}{P}}_{\text{Consumption}} \left[a_j \exp\left(\eta_{ij}\right)\right]^{\frac{1}{\beta_h}}$$
$$= \begin{cases} \frac{1}{P} \exp\left(\beta_h \log\left(\max\left\{\epsilon w_{hj}, \underline{y}\right\}\right) + \log a_j + \eta_{ij}\right)^{\frac{1}{\beta_h}} & \text{if } \epsilon \geq \underline{\epsilon}_{hj} \\ 0 & \text{otherwise} \end{cases}$$

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

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

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

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

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

$$n_h(w_{hj}, \underline{\epsilon}_{hj}, a_j) = N_h a_j \int_{\underline{\epsilon}_{hj}}^{\infty} \left(\frac{\max\left\{\epsilon w_{hj}, \underline{y}\right\}}{\omega_h(\epsilon)}\right)^{\beta_h} r_h(\epsilon) d\epsilon \tag{1.4}$$

$$l_h(w_{hj}, \underline{\epsilon}_{hj}, a_j) = N_h a_j \int_{\underline{\epsilon}_{hj}}^{\infty} \exp(\epsilon) \left(\frac{\max\left\{\epsilon w_{hj}, \underline{y}\right\}}{\omega_h(\epsilon)}\right)^{\beta_h} r_h(\epsilon) d\epsilon \tag{1.5}$$

Finally, total labor costs are given by:

$$C_h(w_{hj}, \underline{\epsilon}_{hj}, a_j) = N_h a_j \left[\int_{\underline{\epsilon}_{hj}}^{\frac{\underline{y}}{w_{hj}}} \frac{\underline{y}^{\beta_h+1}}{\omega_h(\epsilon)^{\beta_h}} r_h(\epsilon) d\epsilon + \int_{\frac{\underline{y}}{w_{hj}}}^{\infty} \frac{(\epsilon w_{hj})^{\beta_h+1}}{\omega_h(\epsilon)^{\beta_h}} r_h(\epsilon) d\epsilon\right] \tag{1.6}$$

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

Problem of the firm

Firms maximize profit by choosing posted wages and rejection cutoffs:

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

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

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

$$p_g f_h\left(\boldsymbol{l}(\boldsymbol{w}_j, \underline{\boldsymbol{\epsilon}}_j, a_j), b_g\right) \frac{\beta_h}{\beta_h + 1} = w_{hj} \qquad h = 1, \dots, H \tag{1.7}$$

$$p_g f_h\left(\boldsymbol{l}(\boldsymbol{w}_j, \underline{\boldsymbol{\epsilon}}_j, a_j), b_g\right) \underline{\epsilon}_{hj} = \underline{y} \qquad h = 1, \dots, H \tag{1.8}$$

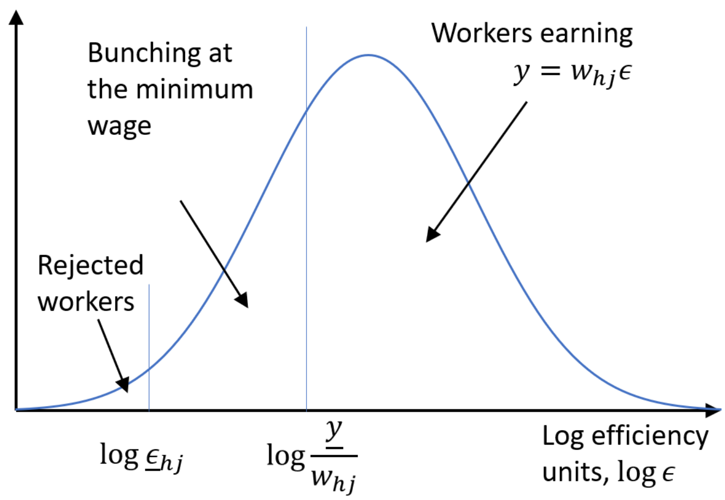
Additionally, firms producing good g choose the same wages $\boldsymbol{w}_g$ and rejection criteria $\underline{\boldsymbol{\epsilon}}_g$. Output and employment are linear in firm amenities: $q_j = \frac{a_j}{\bar{a}_g}\bar{q}_g$ and $\boldsymbol{l}_j = \frac{a_j}{\bar{a}_g}\bar{\boldsymbol{l}}_g$,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

The first order conditions represent trade-offs along two different margins: workers above the minimum wage and workers around the rejection threshold. To build intuition on the optimality condition on wages, denote by l_{hj}^{+} the sum of efficiency units at firm j supplied by workers earning more than the minimum wage. A proportional increase in posted wages $d\log w_{hj}$ brings in $(\beta d \log w_{hj}) l_{hj}^{+}$ labor units, generating $(\beta d\log w_{hj}) l_{hj}^{+} p_g f_h(\cdot)$ in additional revenues. Labor costs increase for two reasons. First, the firm pays $(\beta d \log w_{hj}) l_{hj}^{+} w_{hj}$ for the additional labor purchased. Second, a higher wage increases the wage bill for current workers by $d\log w_{hj} l_{hj}^{+} w_h$. Setting added revenues equal to additional costs yields Equation 1.7.

Equation 1.8 is the first order condition on the rejection cutoffs. A lower cutoff brings in additional workers with $\epsilon = \underline{\epsilon}_{hj}$, each of which increases revenues by $p_g f_h \underline{\epsilon}_{hj}$. When firms chose thresholds optimally, that additional revenue equals the minimum wage $\underline{y}$, which is the cost of labor at that margin.

Figure 1.3 illustrates how workers are divided in three groups according to their level of efficiency units. Those to the left of $\underline{\epsilon}_{hj}$ are rejected. Those with $\epsilon > \underline{y}/w_{hj}$ earn the wage posted by the firm times their quantity of labor units. Finally, those in the intermediate range earn the minimum wage. The first order conditions imply that these two thresholds are proportional to each other in all firms choosing labor inputs optimally, with their ratio being given by $1+1/\beta_h$. Log wage histograms simulated from the model have peaks at the minimum wage corresponding to the mass of workers between the two vertical lines. Bunching at the

Figure 1.3: Choice of rejection criterion and bunching at the minimum wage



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

minimum wage is often observed in the data (DiNardo, Fortin and Lemieux, 1996; Harasztosi and Lindner, 2018) but is not a common feature in models of wage inequality.

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

$$\omega_h(\epsilon) = \left(\sum_g J_g \mathbf{1} \left\{ \epsilon > \underline{\epsilon}_{hg} \right\} \bar{a}_g \max \left\{ \epsilon w_{hg}, \underline{y} \right\}^{\beta_h} \right)^{\frac{1}{\beta_h}} \tag{1.9}$$

Equilibrium

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

1. Markets for goods clear:

$$Q_g = \left[\frac{p_g}{P}\right]^{-\sigma} \frac{I}{P} = J_g \bar{q}_g \quad \forall g \tag{1.10}$$
$$\text{where } I = T + \sum_{g=1}^{G} J_g \sum_{h=1}^{H} C_h(w_{hg}, \underline{\epsilon}_{hg}, \bar{a}_g)$$

2. For all g, firm choices solve the set of equations (1.7) and (1.8).

3. Entrepreneurs have zero ex-ante expected profits:

$$E_{a_j|g}\left[\pi_g(a_j)\right] = p_g f\left(\boldsymbol{l}(\boldsymbol{w}_g, \underline{\boldsymbol{\epsilon}}_g, \bar{a}_g), b_g\right) - \sum_{h=1}^{H} C_h(w_{hg}, \underline{\epsilon}_{hg}, \bar{a}_g) = F_g \quad \forall g \tag{1.11}$$

4. The market for entrepreneurial talent clears (Equation 1.3).

Labor market clearing is implied by the definition of $\omega_h(\epsilon)$, which ensures that the number of job applicants to all firms (calculated using Equation 1.4) is equal to total number of workers N_h.

Solving for equilibrium can seem challenging at first glance. Using a convenient set of choice variables reduces the problem to solving a square system of $(H+1) \times G$ equations where the choice variables are firm-specific task thresholds, firm-level output, and prices for each good. The procedure below describes how to calculate that system of equations:

1. Start with values for mean output $\bar{q}_g$ and task thresholds $\bar{\boldsymbol{x}}_g = \{\bar{x}_{1g}, \ldots, \bar{x}_{Hg}\}$ 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 1.2 in the text, or Equation A.2 in Appendix A.2 if using the exponential-Gamma parametrization).

3. Find marginal products of labor f_{hg} via the non-arbitrage conditions (1.1) and the constant returns to scale relationship $\sum_h f_{hg}\bar{l}_{hg} = \bar{q}_g$.

4. Employ the first order conditions of the firm (1.7) and (1.8) to find wages w_{hg} and rejection cutoffs $\underline{\epsilon}_{hg}$, respectively.

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

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

7. Obtain $\omega_h(\epsilon)$ using expression 1.9.

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

   a) The deviation between $\bar{l}_{hg}$ found in step 2 and that implied by the labor supply curve (1.5).

   b) The deviation between profits and entry costs in Equation 1.11.

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

Determinants of the wage distribution

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

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

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

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

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

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

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

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

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

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

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

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

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

To provide a concrete example of how firms differ in equilibrium, consider the Exponential-Gamma parametrization introduced in Subsection 1.2. Under that parametrization, goods are fully described by five scalars: blueprint complexity θ_g, blueprint shape k_g, blueprint productivity z_g, mean amenities $\bar{a}_g$, and the ratio of entry costs to mean amenities $F_g/\bar{a}_g$. These scalars map directly into five empirical measures for the set of firms producing that good, respectively: mean worker education, dispersion in worker education, share of total workforce employed by those firms, mean firm size, and firm-level wage premia.

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

Proposition 3 (Race between technology, education, and minimum wages)**.** *Start with a baseline economy characterized by parameters* $\left(\{e_h, N_h, \beta_h\}_{h=1}^{H}, \{b_g, F_g, \bar{a}_g\}_{g=1}^{G}, T, \sigma, \underline{y}\right)$ *and consider a new set of parameters denoted with prime symbols. Assume* e_h *are decreasing functions to simplify interpretation (more complex tasks are harder to produce). Let* Δ_0*,* Δ_1 *and* Δ_2 *denote arbitrary positive numbers and consider the following conditions:*

1. $N_h' = \Delta_0 N_h$ $\forall h$ *and* $T' = \Delta_0 T$*: The relative supply of factors remains constant.*

2. $e'_h(x) = e_h\left(\frac{x}{1+\Delta_1}\right)$ $\forall h$: *Workers become better at all tasks and the degree of comparative advantage becomes smaller for the current set of tasks (e.g. both high school graduates and college graduates improve at using text editing software, but the improvement is larger for high school graduates).*

3. $b'_g(x) = \frac{1}{(1+\Delta_1)(1+\Delta_2)} b_g\left(\frac{x}{1+\Delta_1}\right)$ $\forall g$: *Production requires fewer tasks, but the composition of tasks moves towards increased complexity.*

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

If these conditions are satisfied, the equilibrium under the new parameter set is identical to the initial equilibrium, except that prices for goods are uniformly lower: $p'_g = p_g/(1+\Delta_2)$ *and* $P' = P/(1+\Delta_2)$.[8]

Proposition 3 delineates technological progress in this economy. Production becomes more efficient by using tasks that are more complex. At the same time, the skill of workers increase, changing the set of tasks where skill differences are relevant. If minimum wages remain as important, then there is a uniform increase in living standards. Wage differences between worker groups and between firms for the same group remain stable.

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

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

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

---

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

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

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

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

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

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

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

To partially address problems related to informality, I restrict my analysis to Rio Grande do Sul, the largest state in Southern Brazil. Except for the Southeast, the Brazilian South is the region with the lowest rates of employment informality in the country. The Southeast is less interesting for this exercise, however, because higher wages in that region make the minimum wage shock less relevant.[9] I discuss the implications of ignoring the informal sector before presenting the results of the counterfactual exercises. In addition, Appendix A.3 contains a thorough analysis of inequality and education patterns using a different data source that includes the informal sector.

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

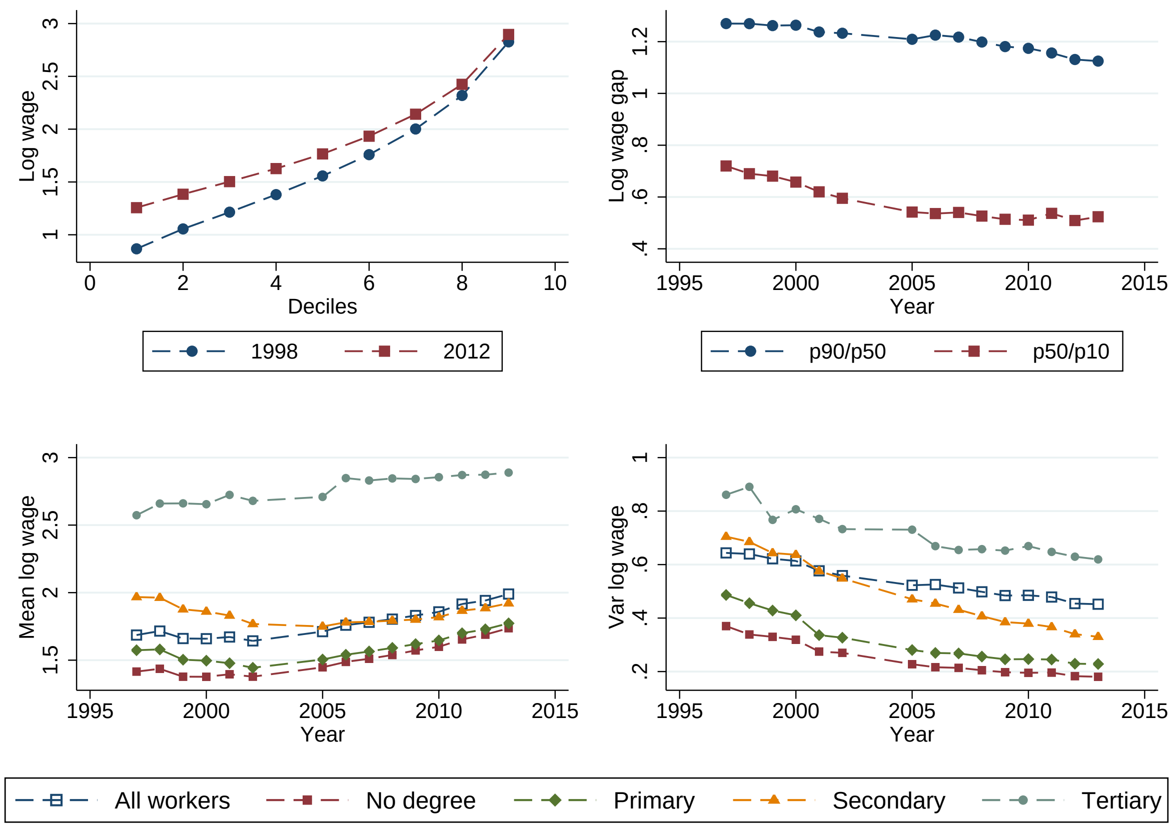
Because I am interested in equilibrium effects, the sample I use has few restrictions. I select adults of both genders between 18 and 54 years of age, who are not currently in school, and who are working in December having been hired in November or earlier. I only consider one job per worker per year. The resulting data set has 1,494,186 workers and 148,203 firms in 1998, and 2,398,391 workers and 238,545 firms in 2012. For each worker, I calculate the hourly wage based on their monthly earnings and contract hours, before winsorizing the bottom and top percentiles of the wage distribution. Summary statistics are provided in Table A.1, located in Appendix A.3.

Target moments

Figure 1.4 demonstrates the evolution of wages in the *Rio-Grandense* economy. The top left panel shows that, from 1998 to 2012, real wages have increased for all deciles of the log wage distribution, and particularly so for the lowest deciles. Almost all commonly used measures show a reduction in inequality: upper-tail or lower-tail percentile gaps (top-right panel), differences in mean log wage between workers with secondary education (that is, those complete high-school and college dropouts) and less educated workers, and the variance of log wages — for the sample as a whole and within each educational group. The single

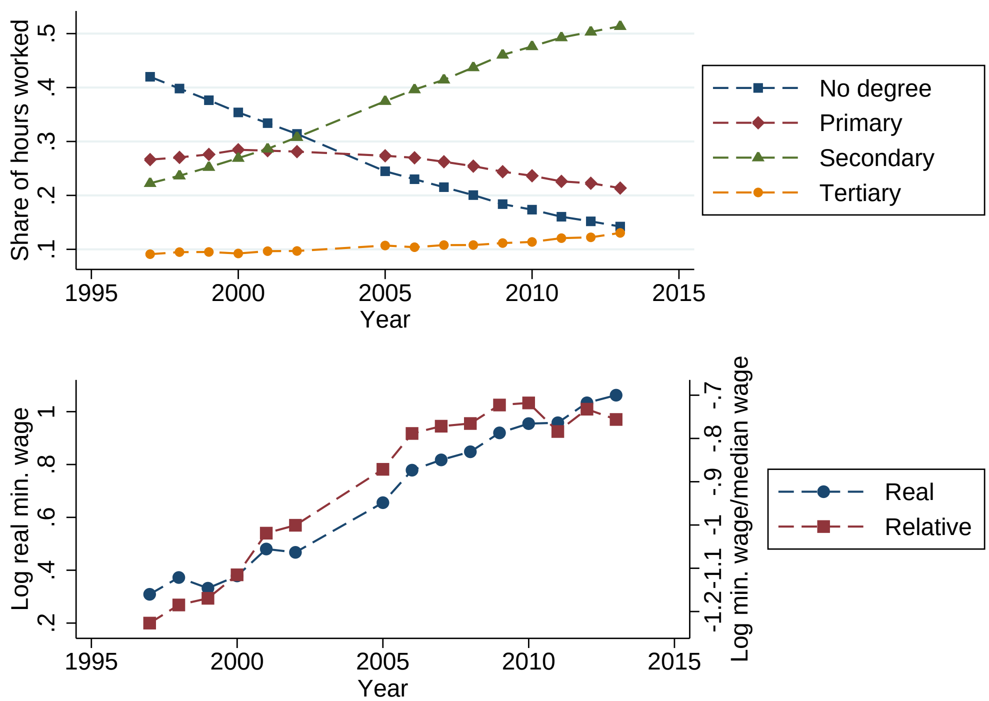
[9] I use a single state in the South because the estimator of variance components of Kline, Saggio and Sølvsten (2018) performs better in well-connected labor markets (in terms of worker transitions between firms). Table 1.5 shows that the sample size is large enough to generate precise estimates.

Figure 1.4: Measures of wage dispersion in Rio Grande do Sul, Brazil



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

Figure 1.5: Changes in educational achievement and minimum wages



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

exception is the gap between secondary and tertiary education (workers that had completed college and beyond), which rose until 2006 and subsequently remained stable through to the end of the period studied. In Appendix A.3, I show that wage inequality trends are similar in a different data set that includes informal workers.

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

The literature studying wage inequality in Brazil highlights two candidate explanations for these patterns: increased educational achievement and minimum wages. Figure 1.5 shows that both factors are relevant in Rio Grande do Sul. The first graph displays the fraction of hours worked by employees in each educational group. The pattern is striking: workers with less than a complete primary education (that is, less than eight years of schooling) supply 40 percent of the hours in 1998, but only around 15 percent in 2012. On the other hand, the group with a complete secondary education (high school and college dropouts) increased its participation level by almost 30 percentage points. Moreover there is a substantial increase in college completion in relative terms (from 9.4 percent to 12.2 percent), though they remain a fraction of the formal workforce.

In a strict sense, these are not changes in the supply of labor, but are instead the observed employment shares for each educational group. Even with the exogenous labor supply in the model, these two concepts differ because the minimum wage creates involuntary non-employment. Figure A.3 in Appendix A.3 shows similar trends in the share of all adults belonging to each of these educational groups, regardless of whether they participate in the labor force or not. That fact indicates that the source of changes in schooling achievement of the workforce are changes in the education levels for the whole population, not changes in selection patterns into employment.

The bottom graph in Figure 1.5 shows the large and steady increase in the national minimum wage in Brazil. The same figure shows that the minimum wage increased much faster than median wages in Rio Grande do Sul until 2006. The estimation procedure will target levels and changes in five measures that capture the degree to which the minimum wage is binding: the shares of workers in each educational group earning up to the minimum wage plus 25 log points, along with the minimum wage relative to mean log wages.

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

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

$$\begin{aligned} \log y_{st}(p) - \log y_{st}(50) = &\beta_1(p) \left[\log \underline{y}_t - \log y_{st}(50)\right] + \beta_2(p) \left[\log \underline{y}_t - \log y_{st}(50)\right]^2 \\ &+ \zeta_{0s}(p) + \zeta_{1s}(p) \times time_t + \zeta_2(p) \times (time_t)^2 + u_{st}(p) \end{aligned} \tag{1.12}$$

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

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

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

Table 1.1: Reduced form estimates of minimum wage spillovers

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

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

Table 1.1 shows ordinary least squares and instrumental variables estimates of the marginal effect of minimum wages over different quantiles of the wage distribution. I estimate specifications in levels and in differences. The specification in differences presents much stronger first stages (measured by the Cragg-Donald (1993) F statistic). In addition, it shows no spillovers in the upper tail, a criterion that has been used for model selection when studying the impact of minimum wages on the wage distribution (e.g. Autor, Katz and Kearney (2008) and Cengiz et al. (2018)). For these reasons, it is my preferred specification.

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

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

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

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

$$\text{Var}(\log y_{it}) = \text{Var}(\nu_i) + \text{Var}\left(\psi_{J(i,t)}\right) + 2\text{Cov}\left(\nu_i, \psi_{J(i,t)}\right) + \text{Var}(\delta_t) + 2\text{Cov}\left(\nu_i + \psi_{J(i,t)}, \delta_t\right) + \text{Var}(u_{it}) \quad (1.13)$$

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

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

Estimating the variance decomposition (1.13) is not a trivial task. Andrews et al. (2008) show that a simple "plug-in" estimator of the covariance composition using estimates of the fixed effects from ordinary least squares (OLS) regressions is biased. These authors provide a correction method that assumes the homoskedastic residuals. More recently, Kline,

Table 1.2: Variance decomposition from two-way fixed effects model

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

Notes: This table shows the variance decomposition described in Equation (1.13) obtained using the Leave-out estimator of Kline, Saggio and Sølvsten (2018). Numbers in parentheses are asymptotic standard errors. The estimation labeled 1998 uses data for two years, 1997 and 1999. Similarly, the 2012 estimation uses data for 2011 and 2013. The decomposition includes both workers who move between firms and stayers. Each worker-year observation has the same weight. See Appendix A.3 for details and sample sizes.

Saggio and Sølvsten (2018) show that heteroskedasticity causes bias and proposed a leave-out estimator that corrects for it. These authors also discuss how to conduct inference on the variance decomposition terms. I use the latter estimator (henceforth denoted by KSS) because the variance of residuals vary systematically across worker groups both in the data and in the model. Appendix A.3 provides details about the procedure.

Table 1.2 shows the KSS estimates of variance components. The variance of both firm and worker effects decline over time, helping to explain the fall of wage inequality. There is also a sizable and statistically significant increase in the covariance of worker effects and firm effects. As a result, the correlation between worker effects and firm effects increases substantially.

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

Fitting the model

Parameters

I use a simple, parsimonious version of the model to fit these moments. I employ the exponential-Gamma parametrization of comparative advantage functions and blueprints described in Section 1.2. The complete list of estimated parameters is presented in Table 1.4.

Table 1.3: Target moments and model fit

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

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

Table 1.4: Estimated parameters

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

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

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

[10] I calibrate one of the β_h because, without that normalization, their levels are weakly identified. The main source of identifying variation for the β_h are levels and changes in within-group variance of log wages, because in the model these parameters affect wage differentials between firms. However, the parameter set being estimated also includes the difference in entry costs between the two firm types, which also affect between-firm wage differentials for similar workers in the model. Under correct specification, all β_h can still be identified jointly with the entry cost gap because they imply the size of bunching at the minimum wage. Figure 1.6, however, shows that the distribution of log wages is not perfectly specified and that the model cannot exactly match the observed bunching. Thus, using a calibrated value for either one of the β_h or the entry cost gaps can lead to more reliable estimates. I chose β_4 for that purpose because, among Brazilian workers, those with college are probably the most similar to workers in developed countries where

The estimated model has only two goods (and, equivalently, two firm types), making it parsimonious and simplifying its interpretation. I make further restrictions on parameters related to goods. The shape parameter of blueprints is assumed to be constant over time and common for both goods. In addition, because I am not interested in predictions about firm size or the number of firms, I restrict my attention to the ratio of entry costs to mean amenities — more specifically, the log difference of this ratio between firm types. Finally, the elasticity of substitution between goods is calibrated to $\sigma = 2$. This parameter is technically identified from minimum wage spillovers; specifically, from the secondary impact of that shock on wages discussed in Proposition 4. In this particular case, however, that effect is small relative to the imprecision in the estimated minimum wage spillover (see Figure 1.7). Thus, that parameter is weakly identified in practice, and calibration is a more transparent approach.[11]

I now introduce the definition of the demand shock as a combination of three components. First, following the tradition in this literature, I allow for a skill-biased technical shock that increases productivity gaps between worker groups. In the task-based production function, this can be modeled as a proportional increase in task complexities $d\log\theta$, as shown in Proposition 3. The second component is a possible change in the relative productivity of firm types, $d\log z_2/z_1$. This component affects the overall employment share of each firm type. The third component is a change in the relative ratio of entry costs to amenities, $d\log((F_2/\bar{a}_2)/(F_2/\bar{a}_2))$. That relative ratio is a key determinant of cross-firm wage dispersion for similar workers.

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

Estimation procedure

The model is estimated by minimizing the least squares distance between the observed moments in Table 1.3 and the model equivalent of these moments. Each of the three groups of moments have the same importance in estimation. To evaluate the loss function, I need a function that maps the vectors of parameters to the vectors of moments predicted by two equilibria of the model, one corresponding to 1998 and another corresponding to 2012. To find these equilibria given a set of candidate models, I use a slightly modified set of equilibrium conditions: instead of imposing that the number of job applicants equals the total number of workers of each type (which is unobserved in the data), I impose instead

the estimates cited in Manning (2011) were obtained.

[11] I assume that goods are net substitutes because the model-based curves in Figure 1.7 suggest that the higher the substitution, the closer the predicted spillovers in the upper tail are to reduced-form estimates. Still, I chose a relatively small value for the elasticity of substitution as a way of being conservative about the extent of changes in the composition of firm types following structural shocks.

that the number of employed workers in the model equals observed employment.[12]

The first group of moments in Table 1.3, along with the measures of how binding the minimum wage is in the third group, are calculated from the model-implied wage distributions for each worker type. Minimum wage spillovers are calculated as follows. First, I take each of the estimated equilibria and calculate a counterfactual where the log minimum wage increases by 10^{-6} relative to the entry input. Next, consistent with the reduced form specification (Equation 1.12), I calculate the changes in each quantile p over the change in the minimum wage, both of which are relative to the median wage: $\Delta[y(p)-y(50)]/\Delta[\underline{y}-y(50)]$. Finally, I use the average of spillovers across both periods (1998 and 2012) as the model-based analogues of the reduced form estimates.

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

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

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

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

[13]Formally, I set $\epsilon_{it} = \epsilon_i^F + \epsilon_{it}^T$, with the transitory component ϵ_{it}^T orthogonal to the fixed component ϵ_i^F. The variances of each of these components is calibrated so that (i) the variance of ϵ_{it} is the parameter S^2; and (ii) the estimated share of the residuals in the model-based variance decomposition is 0.9.

[14]OLS estimates are consistent in this regression because the number of firm fixed effects being estimated does not grow with the sample size.

Estimates, goodness of fit, and discussion

Table 1.4 shows the estimated parameters and Table 1.3 illustrates the moments predicted by the estimated model. Overall, the model fits the data well, particularly for the first two groups of moments (levels and trends in wage inequality and components of the variance decomposition based on the two-way fixed effects regressions). That the model fits within-group variances well is particularly interesting given the restriction that the variance of labor efficiency units is the same for all groups and both periods.

Figure 1.7 shows that minimum wage spillovers from the model are similar to the reduced form estimates, though the model understates positive spillovers in the lower tail and overstates negative spillovers in the upper tail. The same figure also shows how the elasticity of substitution parameter substitution between goods affects comparative statics, following the logic introduced in Proposition 4. More substitution between good leads to a less negative spillovers in the upper tail of the distribution. Those differences, however, are small.[15]

The fit of the model can be verified visually by comparing observed wage distributions with model-generated ones in Figure 1.6. Overall, the model captures the most salient features of the data. The fit is worse close to the minimum wage, with the model over-predicting bunching in this area.

Now I discuss the estimated parameters shown in Table 1.4 and the implied equilibria in the model. I start with worker-related parameters. Comparative advantage α_h is increasing in education, as expected. These parameters are identified from changes in mean log wage gaps. The taste for consumption β_h is estimated to be decreasing with education. Workers with no complete degree have a high taste for consumption, implying that labor markets are close to competitive for that worker group. For workers with more education, the implied wage mark-downs relative to marginal productivities range from 12 percent (for workers with basic schooling) to 20 percent (for workers with a college degree). The variance of efficiency units S^2 is 0.447, close to the total variance of log wages in 2012. Both S^2 and β_h are identified essentially from levels and changes in the variance of log wages within worker groups.

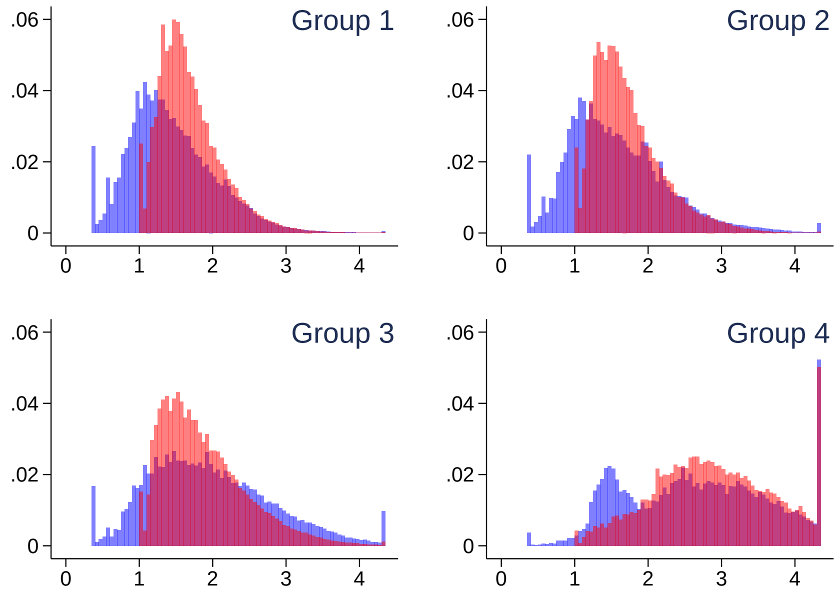
Good $g = 2$ is more intensive in task complexity, has higher fixed costs, and requires fewer tasks in production than good $g = 1$. The estimated skill-biased component of the demand shock is positive, stretching the distribution of task requirements for both goods to the right. The demand shock also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

The resulting employment patterns, along with the mean log wages for workers of each type employed at firms producing each good, are shown in Figure 1.8. Firms producing the second good are more skill intensive in both periods. In the first period, firms producing $g = 2$ pay higher wages to all worker types except the least educated one. The size of the wage premium is increasing in worker type. These differences stem from a combination of increased demand for skilled workers along with a higher ratio of entry costs to amenities

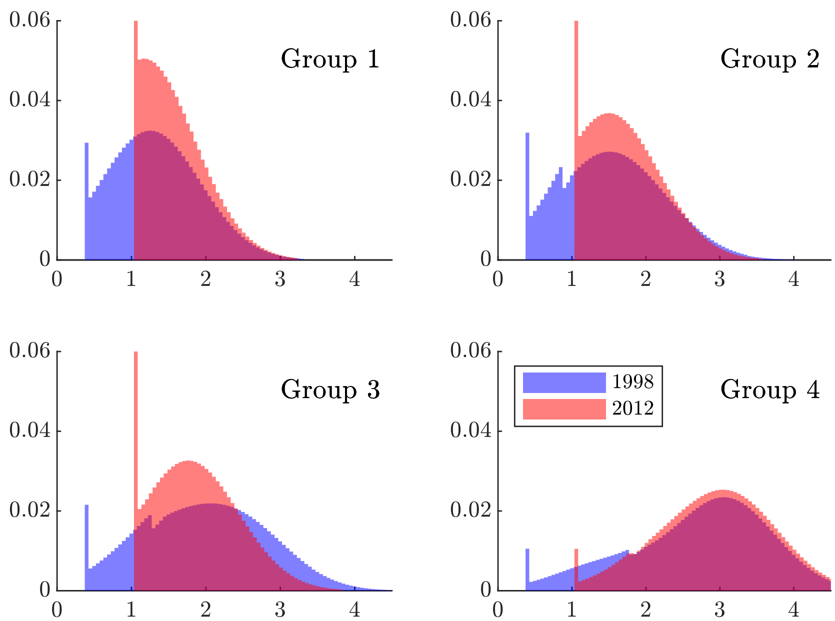
[15] In this exercise, productivity gaps are adjusted so that employment shares of each firm type remain the same after the change in σ. As a result, all other moments are the ones shown in Table 1.3.

Figure 1.6: Distribution of log wages, data and model

(a) Data

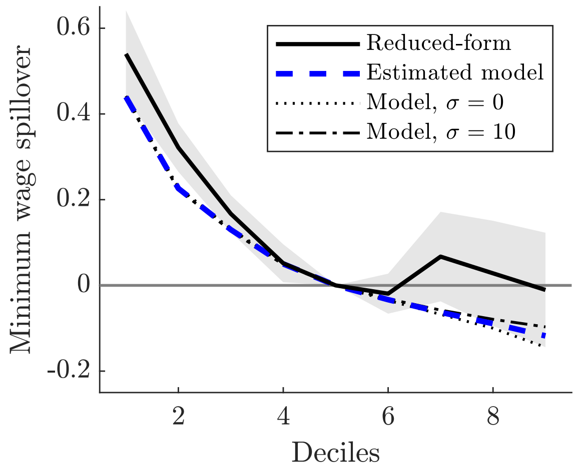


(b) Simulation from estimated model



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

Figure 1.7: Minimum wage spillovers



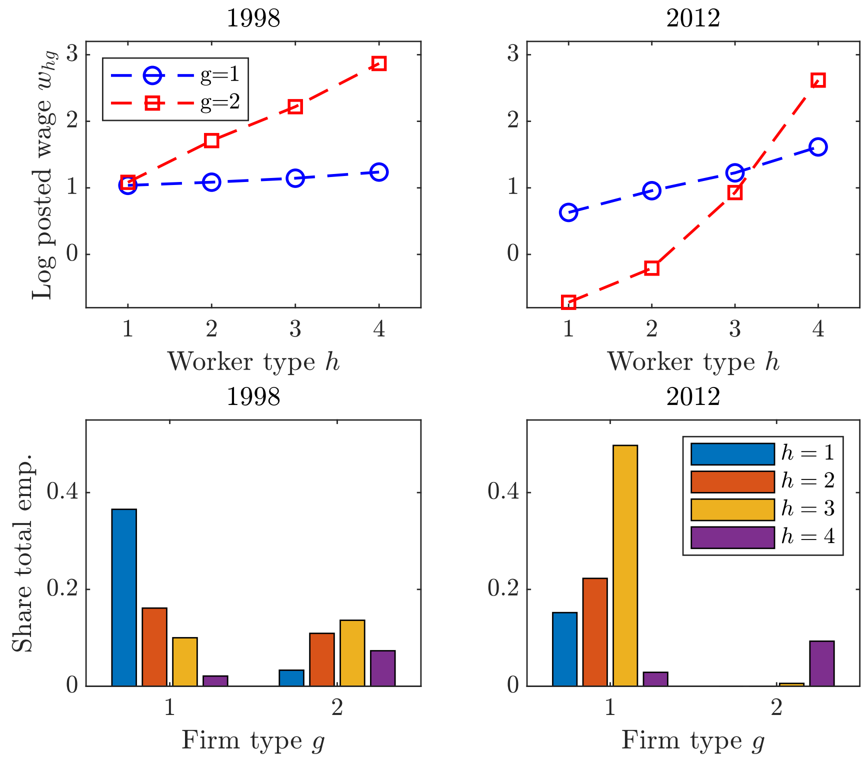
Notes: This figure shows the impact of changes in the "effective minimum" (statutory minimum wage minus median wage) on different deciles of the wage distribution, relative to the mean. The solid line shows marginal effects from the instrumental variable estimation of Equation (1.12), following Autor, Manning and Smith (2016). The shaded area represents 95 percent confidence intervals for these estimates. The dashed line shows spillovers predicted from the estimated model. The dash-dot line and the dotted line demonstrate the spillovers from a model similar to that of the estimated one, with the exception that substitution between goods is either ruled out (that is, $\sigma = 0$) or amplified ($\sigma = 10$) relative to the baseline estimation ($\sigma = 2$).

in the second firm type relative to the first, as described in Proposition 2. These differences explain both the substantial share of firm effects in the variance of log wages and why the estimated firm effects from the log-additive model correlate with worker effects.

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

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

Figure 1.8: Wages and employment by firm type and worker type



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

Disentangling the role of supply, demand, and minimum wages

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

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

Table 1.5: Model-based decomposition: supply, demand, and minimum wages.

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

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

to the estimated model in the second period, which includes changes in all dimensions.

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

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

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

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

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

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

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

1.5 Conclusion

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

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

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

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

Chapter 2

Workforce Composition, Productivity, and Labor Regulations in a Compensating Differentials Theory of Informality

This chapter is coauthored with Rodrigo R. Soares. A previous version was posted online as IZA Discussion Paper No. 9951 on May 2016.

2.1 Introduction

Labor market informality has been a major policy concern worldwide for several decades. Informal employment is not protected by labor legislation, cannot be taxed, and does not entitle workers to social security benefits. These constitute challenges to policy making in terms of the optimal design and effectiveness of both the social protection and tax systems. In developing countries, these challenges are magnified by the limited enforcement ability of governments and the sheer size of informal employment, well above 30% of the labor force in most cases. Specific programs and institutional efforts targeted at reducing labor informality have typically met with limited success (Perry et al., 2007).

Surprisingly, this historical pattern of persistently high informality was sharply reversed in most of Latin America in the early 2000s. In a half-dozen countries, informality rates among salaried workers were reduced by one-fifth or more in a period of roughly 10 years (Tornarolli et al., 2012). These shifts remain largely unexplained and cannot be accounted for by current models of informality. The decline in labor informality in Brazil, which provides the data for our quantitative exercises, is particularly puzzling. Informality among salaried workers was reduced by 10.7 percentage points between 2003 and 2012, from an initial level of 30%. At the same time, the minimum wage increased by 61% in real terms, at least twice the growth rate of GDP per capita, while changes in labor legislation and payroll taxes were negligible. But Brazil also experienced other relevant economic transformations during this period, including substantial increases in average years of schooling and TFP. In principle, these transformations may have had their own equilibrium effects on informality, through changes in the demand and supply of different types of labor and the ensuing impact on relative wages and unemployment.

The main difficulty in assessing the relevance of this latter possibility comes from the absence of an adequate theoretical framework. The modern informality literature is unable to analyze the implications of supply-demand interactions across different types of labor due to is reliance on traditional search models, which assume one-to-one matches between workers and firms or constant marginal productivity of labor. These assumptions immediately rule out complementarities across different types of labor and, therefore, equilibrium responses to changes in the relative supply of different types of workers.

In this paper, we develop a search and matching model of informality that allows for worker and firm heterogeneity, decreasing returns to scale, imperfect substitutability between different types of labor within the firm, a realistic set of labor regulations (including minimum wage), and explicit compliance decisions by workers and firms. We estimate the model using data from Brazil and show that it closely reproduces the changes in informality during the 2000s. This quantitative exercise also shows that the educational composition of the labor force and TFP can have first order implications for labor market equilibrium outcomes – including informality, unemployment, and relative wages – through their effects on the demand and supply of different types of labor. The incorporation of heterogeneous labor and decreasing returns to scale allows the model to assess how informal labor markets respond to changes in aggregate variables in ways that would have been impossible under the frameworks commonly used in the previous literature or, alternatively, with reduced-form empirical analyses.

In order to accommodate decreasing returns to scale and imperfect substitutability between different types of labor within a search model, we draw from the intra-firm bargaining theory proposed by Cahuc, Marque and Wasmer (2008), who build on Stole and Zwiebel (1996*a*), and extend it in three directions. First, we characterize an equilibrium where labor can move between the formal and informal sectors. Second, we consider firms with different productivity levels, as opposed to a single representative firm. And third, we incorporate a more realistic set of labor regulations, including minimum wages, which adds a non-trivial degree of complexity to the characterization of the solution.[1]

In the model, workers can be either skilled or unskilled and search simultaneously for formal and informal jobs when unemployed. Firms are heterogeneous in a skill-biased productivity parameter, so that more productive firms are also more intensive in skill. Firms first decide on whether to comply with labor regulations and then, at each moment, on how many skilled and unskilled vacancies to post. By not complying with regulations, firms avoid

[1]Carbonnier (2015) has independently developed a model that adds payroll taxes to Cahuc, Marque and Wasmer (2008). It does not include mandated benefits nor minimum wages in the wage bargaining problem (minimum wages are modeled in that paper as an exogenous price for low skilled workers).

payroll taxes and are not subject to the minimum wage, but face an informality penalty that is increasing in firm size (representing the probability of being audited and the associated fine). Labor regulations also include mandated benefits, which from the perspective of employees make formal jobs more valuable than informal jobs for a given wage. Finally, wages are set by intra-firm bargaining under non-binding contracts, so that changes in firm size lead to wage renegotiation with all workers in the firm.

The model leads to an equilibrium where firms and workers self-select into the formal and informal sectors following a compensating differentials logic. Firms do not want to comply with labor regulations, but non-compliance is too costly for large firms. Workers want to receive employment benefits, but may be willing to accept informal jobs and leave unemployment for a sufficiently high wage. The only labor market distortions are those introduced by imperfectly enforced labor regulations and the search and matching frictions. The marginal informal firm is technologically indistinguishable from the marginal formal firm, and skilled and unskilled workers employed in both sectors are identical. So there is no sense in which firms and workers allocated to different sectors are intrinsically different, as the classic labor market duality hypothesis would suggest (see Cain, 1976).

In a steady-state equilibrium, firms with lower productivity employ fewer workers and choose to operate informally. These firms also employ a lower fraction of skilled workers. In general, informal workers are compensated for the lack of mandated benefits by receiving higher wages, but this equalizing differentials condition can be broken by minimum wages. If the minimum wage binds for unskilled workers, they strictly prefer to hold a formal job but are willing to accept informal offers in equilibrium to avoid unemployment. In this equilibrium, the formal wage premium decreases in the skill level, becoming negative for skilled individuals. Average wages are higher in the formal sector due to workforce composition and to the binding minimum wage. But, for skill levels for which the minimum wage does not bind, workers are indifferent between formal and informal employment.

In the quantitative section of the paper, the model is used to analyze the evolution of informality in the Brazilian labor market from 2003 to 2012 and to assess the effectiveness of alternative policies aimed at reducing informality. We estimate the model using data from the Brazilian labor market in 2003 and then examine whether the estimated model is able to replicate the evolution of labor market outcomes between 2003 and 2012. The model reproduces several stylized facts from the cross-sectional distribution of workers across firms and compliance statuses: size distribution of firms, wage patterns across and within the formal and informal sectors, and unemployment. We analyze the role of changes in tax rates, mandated benefits, enforcement of labor regulation, minimum wages, workforce composition, and aggregate productivity in explaining the trends observed in the past decade. By assessing the contribution of each of these factors one at a time, we verify that our comparative statics exercises are roughly in line with the evidence available from reduced-form empirical studies. Once all factors are accounted for, the model reproduces qualitatively all the changes observed in the data, including those related to wages and employment by sectors and skill levels. Quantitatively, the model reproduces 85% of the decline in informality and 69% of the decline in the unemployment rate observed in the period. The predicted evolution of wages also matches the data with reasonable precision.

We find that changes in workforce composition are the most important factor behind the reduction in informality in Brazil: without increases in skill levels, the informality rate would have gone up by 4 percentage points instead of declining. To provide some direct empirical evidence in support of this conclusion, we also conduct a preliminary statistical analysis using Census data from 1991 to 2010. Our analysis shows that there is a positive correlation between average schooling in a local labor market and the probability that workers in that labor market are employed formally, even conditional on workers' own education. This correlation has not been explored before and is consistent with the equilibrium mechanism implied by the model.

Our last quantitative exercise illustrates the use of the model for policy analysis. We examine two policies that subsidize formal low wage employment as a means to reduce informality. In the first policy, the subsidy is implemented in the form of lower tax rates for low wage positions, as in a progressive payroll tax. In the second, the subsidy is instead a direct government transfer to low wage formal workers, similar to a current policy adopted in Brazil (*Abono Salarial*). Our results show that the first alternative can reduce informality and increase government revenues, while the second one is much less cost-effective. The reason behind the sharp contrast in outcomes of these apparently similar policies lies in the binding minimum wage. While a reduction in payroll taxes induces employers to create formal jobs, there are no incentives for employers under the second policy, since they do not benefit from the government transfer to workers if wages cannot adjust downward.

In addition to the theoretical points and the quantitative exercises mentioned before, the paper makes two conceptual contributions to the informality literature. First, it shows that both the cross-sectional and time-series variations in informality are consistent with a model that does not impose structural differences in technology across sectors. The model reproduces several stylized facts related to informality and its recent evolution resorting only to regulatory distortions and to search and matching frictions commonly associated with the functioning of the labor market. Second, it rationalizes three interrelated and widely documented patterns that are incompatible with previous informality models: the presence of skilled and unskilled workers in both the formal and informal sectors, the rising share of skilled workers by firm size (and formality status), and the declining formal wage premium by skill level (becoming null or negative at the top). Many authors suggest that the heterogeneity in the formality wage premium indicates that the informal sector is composed of two distinct tiers. For the more productive workers at the top tier, informality is a matter of opportunity, which is reflected on their wages being equal to or higher than they would be in the formal sector. For the bottom tier, informality is strictly worse than formal employment, since informal workers earn lower wages and lack valuable mandated benefits. In our model, the two tiers are clearly identified by the two skill levels, and the pattern of decreasing wage gap results from the binding minimum wage for unskilled workers.[2]

[2] Bargain and Kwenda (2011) find this pattern in fixed-effects models using data from Brazil, Mexico and South Africa. Botelho and Ponczek (2011) reach similar conclusion with Brazilian data under different

Our model builds upon many search models from the informality literature, but differ from them in key aspects. Boeri and Garibaldi (2007) and Boeri, Garibaldi and Ribeiro (2011) propose simple models with worker heterogeneity, but without the possibility of substitutability between different types of labor and with poor institutional characterizations. In both papers, the equilibrium displays complete segregation of workers by skill level across the formal and informal sectors. Albrecht, Navarro and Vroman (2009) introduces uncertainty about workers' productivity in the formal sector and a richer institutional setting, but maintains the one-to-one matching between workers and firms, in addition to assuming strong structural differences between sectors and no compliance decision on the side of the firms. Ulyssea (2010), Bosch and Esteban-Pretel (2012), and Meghir, Narita and Robin (2015) have more sophisticated compliance decisions and are better equipped in institutional details, but forgo worker heterogeneity. Ulyssea (2010) still assumes substantial structural differences between sectors, while Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015) assume that formal and informal firms differ only in their choice to abide by labor regulations.[3] On the institutional side, Ulyssea (2010) incorporates unemployment insurance and severance payments, and Meghir, Narita and Robin (2015) accounts for both these dimensions and minimum wages.[4]

---

specifications (also using panel data), and observe that the formal wage premium decreases as workers become older and more educated. Lehman and Pignatti (2007) find similar results for the Ukrainian labor market. The idea of a two-tiered informal sector goes back at least to Fields (1990). Günther and Launov (2012) develop an econometric model of selection to test the hypothesis of heterogeneity inside the informal sector. They find that there are two distinct groups in the informal sector in Cote d'Ivoire. Some of these authors, as well as others, have used the term segmentation to describe the bottom tier of the informal sector. By that, they mean that wages are not fully determined by individual productivity and compensating differentials. This interpretation, present in Fields (1975) and Rauch (1991), is different from the original concept of segmented labor markets, as described in Dickens and Lang (1985) or Cain (1976). In the case we discuss, increases in education (or, more generally, productivity) can lead every worker to better jobs, a view that contrasts with labor market duality. In addition, the significant flow of workers in and out of the informal sector, particularly among those with lower skills, undermines the hypothesis of strong non-economic barriers of entry to the so-called primary sector. To our knowledge, Araujo and Ponczek (2011) present the only alternative model that explains the decreasing wage gap among salaried workers, but in a very specific setting (one-to-one random matching model with asymmetric information, where workers can take employers to court). Bargain et al. (2012) account for heterogeneity in income gaps between formal and informal self employed workers.

[3]This perspective is supported by the experiment in de Mel, McKenzie and Woodruff (2013) and also by other empirical evidence showing that firms change their compliance decision in response to changes in tax rates (Monteiro and Assunção, 2012; Fajnzylber, Maloney and Montes-Rojas, 2011) or in the intensity of enforcement of labor regulation (Almeida and Carneiro, 2012).

[4]Galiani and Weinschelbaum (2012) model a competitive labor market with heterogeneous firms and workers and self-selection of both into formal and informal sectors following a compensating differentials logic. But they have a single, homogeneous, labor input (workers are heterogeneous in their endowment of this input) and, given the competitive labor markets assumption, cannot account for unemployment. Marrufo (2001) develops a similar competitive model where firms use a single type of labor and workers choose in which sector to work, but she models workers' choices as a Roy model − therefore implicitly assuming structural differences across the formal and informal sectors − and does not allow for endogenous compliance decisions on the side of the firms. The competitive model in Amaral and Quintin (2006) has

The critical features that set our model apart from the rest of the literature are imperfect substitutability across different types of labor and decreasing returns to scale. By considering skilled and unskilled workers and linking them through firms that use both types of labor, embedded within a rich institutional setting, our model reproduces empirical patterns incompatible with previous theoretical models of informality. In addition, it allows us to study the equilibrium effects of changes in aggregate variables – such as workforce composition and TFP – in ways that would otherwise have been impossible.

The remainder of the paper is organized as follows. Section 2 sets the background by describing some stylized facts from the Brazilian labor market and explaining why the recent increase in formalization is a puzzle under existing theories of informality. Section 3 presents the model and discusses some of its properties. Section 4 describes the estimation of the model using Brazilian data. Section 5 uses the estimated model to analyze the evolution of labor market outcomes in Brazil between 2003 and 2012 and conducts some policy experiments. Section 6 concludes the paper.

2.2 Empirical Context

The term "informality" is used to describe many different aspects of non-compliance with regulations. In this paper, we focus on the decision by firms and workers not to comply with labor law when contracting with each other, thus excluding self-employed and domestic workers from the analysis. We also follow the bulk of the literature and restrict our attention to urban informality.

In the Brazilian labor market, a salaried job position is considered formal if the worker's "labor card" (*carteira de trabalho*) is signed by the employer. This is the definition we use henceforth. An employee with a signed labor card is entitled to social security benefits, such as severance payments, pensions, and unemployment insurance, while her employer is obliged to pay social security contributions and payroll taxes. Appendix A contains a thorough description of the benefits available to formal workers and costs associated with formal employment in Brazil.

Most of our data come from the Monthly Employment Survey (*Pesquisa Mensal de Emprego*, PME), a household survey conducted by the Brazilian Census Bureau (*Instituto Brasileiro de Geografia e Estatistica*, IBGE). PME collects information on workers and their employment status in the six largest metropolitan areas in Brazil. We concentrate on the period between 2003 and 2012 due to data availability under a consistent methodology.

The average informal worker in Brazil earns a lower wage, is less educated, and works in a smaller firm than her formal counterpart. The first claim is evident from the top row in Table 2.1. While the average formal hourly wage was 4.83 Brazilian Reais in 2003 (around 1.60 US dollars), the average informal wage was 32% lower (2.67 Brazilian Reais). Table 2.2 also

labor heterogeneity and firms hiring both types of workers. However, it focuses on firm – rather than labor – informality, does not have labor market regulations, and, since it features a competitive labor market, cannot account for wage differentials across sectors or unemployment.

Table 2.1: Labor Market Outcomes, Brazil, 2003-2012

	Informality		Wage gap		Wage growth		Unemployment	
Sample	2003	2012	2003	2012	Formal	Informal	2003	2012
Whole workforce	28.4%	17.7%	-31.9%	-13.4%	13.1%	43.9%	12.6%	5.4%
By schooling:								
Less than 8 years	35.8%	25.9%	-20.2%	-11.8%	26.0%	39.3%	12.2%	4.5%
8 to 10 years	32.1%	23.6%	-21.1%	-10.5%	18.2%	33.9%	16.9%	7.4%
High school, college dropouts	24.0%	14.5%	-14.2%	-3.2%	1.6%	14.7%	13.4%	6.2%
College or more	17.3%	12.6%	-16.1%	10.8%	-12.3%	15.7%	4.3%	2.7%

Source: IBGE/PME, author's calculations.
Notes: Data is presented for October 2003 and October 2012. Informality is fraction of salaried workers in the private sector with a signed work card. Wage gap is the difference between informal and formal average wages as a fraction of formal wages. Wage gain is the relative increase in average wage from 2003 to 2012.

presents the distribution of workers across sectors, firm sizes, and educational categories. By comparing the totals along rows for each sector, the differences in average schooling become clear: 40% of informal employees had less than 8 years of schooling, while the analogous number was less than 28% in the formal sector. The differences in firm size can be seen in the column totals. While only a minority (roughly 1/16) of formal employees worked in firms with 5 workers or less, this fraction was over one third for informal employees.

These stylized facts are consistent with many papers that discuss the empirical regularities of informality in the developing world, such as La Porta and Shleifer (2008) and Maloney (2004). They have been traditionally interpreted as evidence that informality is circumscribed to low-earning, unskilled workers, but a closer look at the data reveals that this assertion is not accurate. Table 2.1 shows that the informality rate among workers with a college degree is 17.3%, not dramatically lower than the overall rate of 28.4%. Moreover, informal workers with college earn almost three times as much as the average formal employee. Note that these individuals are not self-employed professionals defaulting on taxes or social security contributions, since we have restricted our sample to wage earners. The table also suggests that there is no labor market segmentation in the traditional sense: as workers become more educated, they are more likely to be employed formally and also more likely to receive higher wages if they stay in the informal sector. Finally, the fact that some informal firms are willing to pay high wages for skilled workers shows that the technology used by these firms displays significant returns to human capital, contradicting many depictions of labor market duality in which informal firms are presented as being structurally different from formal ones.

But it is also useful to highlight that formal schooling does not seem to encompass all dimensions of skill relevant to the labor market. To illustrate this point, Table 2.3 shows the distribution of wages in the formal sector by educational level. There is a wide dispersion in wages across all levels of schooling, with the exception of college or more. For example, among those with complete high school and college drop outs, there is a fraction of 8.4% earning roughly one minimum wage, while 15.5% earn more than 5 times the minimum wage.

Table 2.2: Educational Distribution of Workers by Sector and Firm Size, Brazil, 2003

	Formal workers, by size of employer				Informal workers, by size of employer			
Worker education	2 - 5	6 - 10	11+	Total	2 - 5	6 - 10	11+	Total
Less than 8 years	36%	30%	27%	28%	49%	37%	33%	39%
8 to 10 years	24%	23%	20%	20%	25%	23%	22%	23%
High school, college dropouts	37%	41%	42%	41%	24%	35%	36%	32%
College or more	4%	6%	12%	11%	2%	5%	9%	6%
Total	1,133	1,226	13,937	16,296	2,363	731	3,196	6,290

Source: IBGE/PME, author's calculations. Salaried workers only. Employer size is reported by the worker in the household survey. The percentage values sum to one along columns. Data from October 2003.

Table 2.3: Formal Wage Distribution by Schooling Levels and Workforce Composition, Brazil, 2003 or 2012 (when indicated)

	Formal wage as multiple of minimum wage					Fraction of workforce	
Worker education	$(0, 1.2]$	$(1.2, 1.5]$	$(1.5, 2]$	$(2, 5]$	$(5, \infty)$	2003	2012
Less than 8 years	18.7%	16.7%	26.9%	35.0%	2.7%	33.8%	20.9%
8 to 10 years	15.3%	14.6%	25.6%	40.2%	4.4%	20.1%	17.1%
High school, college dropouts	8.4%	9.4%	19.4%	47.3%	15.5%	33.6%	43.1%
College or more	0.5%	0.7%	2.2%	22.2%	74.4%	12.5%	18.9%

Source: IBGE/PME, author's calculations. Salaried workers only. Data from October 2003 and October 2012.

Wage dispersion seems almost as large within as across educational categories, despite the fact that average wages – and, therefore, skills – do increase with years of schooling.

We can look at data on firm size in Table 2.2 to infer the hiring behavior of firms in both sectors. Comparisons between different columns in the same sector show that, as firm size increases, the proportion of educated workers also increases. In other words, larger firms are more likely to have a higher fraction of educated workers. An important takeaway is that this pattern is observed for workers in both sectors, suggesting again that the technologies used by formal and informal firms, at the margin, are not substantially different.

Now we turn to the evolution of informality in Brazil since the 1990s. Figure 2.1 shows that the rate of informality was rising up to 2002, but then started declining sharply.[5] In Appendix B, we show that the decline was widespread in the economy and not driven by workforce reallocation (i.e., a movement of employment to sectors of economic activity that are intrinsically more formal). What makes this pattern intriguing is the observation that, while the upward trend has been credited to increasing costs of formal employment during the 1990s, these costs continued to rise even after the reversal.[6] In particular, the minimum

[5]In Figure 2.1, we use data from the National Household Survey (PNAD) instead of the PME, because of methodological changes in PME in 2002.

[6]Barros and Corseuil (2001) explain how the 1988 Constitution significantly raised employment costs (payroll and firing costs and mandated benefits). Bosch, Goñi-Pacchioni and Maloney (2012) claim that these changes were the most important factor behind the increase in informality during the 1990s. We

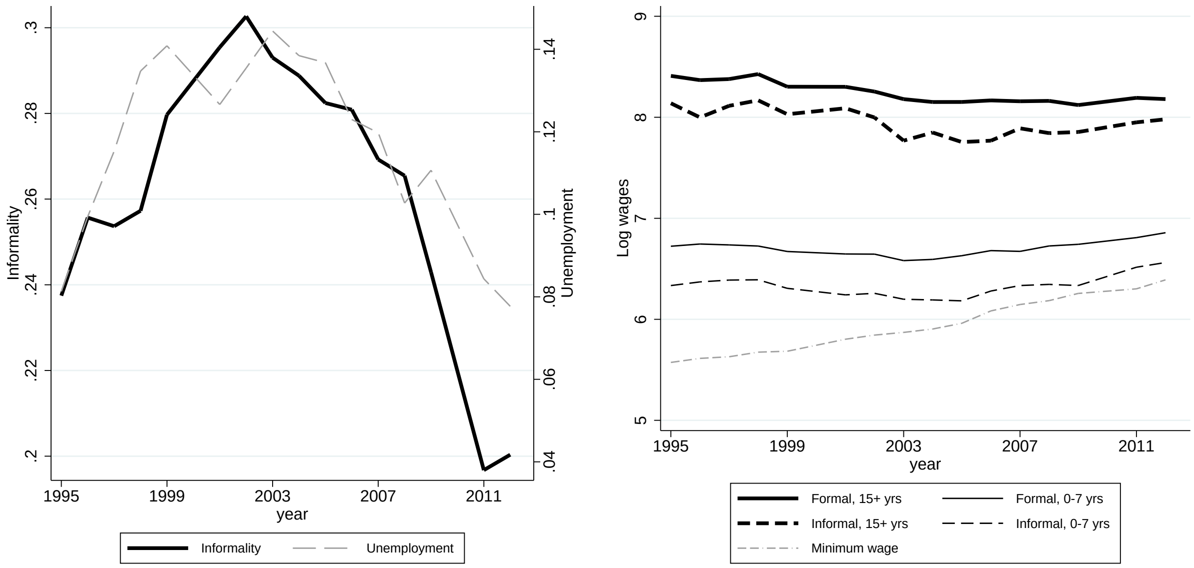


Figure 2.1: Evolution of Informality, Unemployment and Real Wages for Salaried Workers, Brazil, 1995-2012
Source: IBGE/PNAD, author's calculations. The sample is restricted to the six metropolitan regions surveyed in the IBGE/PME.

wage increased dramatically throughout the period, accumulating real gains of 60% from 1995 to the end of 2003, and another 61% from 2003 to 2012.

There is some evidence that the enforcement of labor regulation in Brazil has become more efficient, a factor that could also bring down both unemployment and informality rates.[7] But enforcement cannot account for other important shifts in labor market outcomes: Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015), for example, predict that the formal wage premium should increase as a consequence of more enforcement, which is the opposite of what happens in the data.

The changing composition of the workforce, evident in the last columns in Table 2.3, may have contributed to the patterns described here, despite rarely appearing in the literature as an important determinant of informality. Two intuitive arguments hint at this potentially important role. First, since informality is much lower among the highly educated, increases in the share of skilled workers should mechanically lead to a decline in informality due to a compositional effect (abstracting from equilibrium considerations).[8] Second, the increase in

present a brief discussion of changes in labor legislation and tax rates after 2003 in Appendix A.

[7]The effect of enforcement on unemployment is ambiguous in most models, and quantitative analyses show diverging results. While Boeri and Garibaldi (2007) and Ulyssea (2010) find that increased enforcement leads to higher unemployment, Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015) reach the opposite conclusion.

[8]In fact, Mello and Santos (2009) and Barbosa Filho and Moura (2015) find that changes in workforce composition, particularly skill level, can statistically account for a significant part of the reduction in

the relative supply of skilled workers should reduce their relative wage, leading to increases in the number and size of formal firms (which are intensive in skilled labor) and to a decline in informality conditional on schooling. When coupled with the increases in TFP observed in Brazil during this period (documented, for example, by Ferreira and Veloso, 2013), changes in the relative supply of skills seem promising as a main driving force behind the evolution of labor market outcomes.

In the next section, we develop a model that is able to incorporate all the dimensions discussed here and use it to rationalize both the cross-sectional patterns and the changes in informality observed in Brazil during the last decade.

2.3 The Model

We develop a continuous time model of labor markets with search frictions, firm and worker heterogeneity, informality, a minimum wage, and mandated benefits. In our model, the compliance decision refers to labor informality, not firm informality. Although these concepts are highly correlated in the data, there are some important differences which are reflected in our modeling choices. We focus on payroll taxes, ignoring sales and profit taxes. Moreover, we do not consider the possibility of an intensive margin choice of labor informality within firms, as proposed in Ulyssea (2014). Instead, firms make one single formality decision encompassing all of their job relations. From now on, we use the term "informal firm" or "formal firm" to refer to establishments that offer informal or formal jobs, respectively.

We use a matching framework because it is the simplest way to model unemployment and sectoral wage differentials in this context. The key outcomes in our analysis are skilled-unskilled wage gaps, formality wage premia, unemployment rate, and informality rate. We abstract from wage dispersion within sector-skill level combinations because it has already been studied in the informality literature (see e.g. Meghir, Narita and Robin 2015) and plays no central role in the phenomena we want to analyze. That way, we avoid additional layers of complexity, keeping the analysis transparent and parsimonious. As a result of this modeling choice, the model makes fewer predictions at the worker level than wage posting/on-the-job search approaches (such as those related to transition rates between sectors, wage ladders, and within-sector firm size wage premia).[9] On the other hand, the relative simplicity of the labor market structure allows us to employ a richer technology for firms in our model, leading to interesting supply and demand interactions across skill levels that are yet unexplored in this literature.

Before describing the model in detail, we first provide a sketch of its basic logic. There is a continuum of measure 1 of infinitely-lived, income-maximizing workers with identical preferences. Workers can be either skilled or unskilled, and the fraction η of skilled workers

informality rates in Brazil from 2002 to 2007.

[9] The model generates a firm size wage premium for similar workers that is completely explained by formality status. Pratap and Quintin (2006) and Badaoui, Strobl and Walsh (2010) discuss the relationship between the formality wage premium and the firm size wage premium in developing countries.

in the population is exogenous. There is a measure m of firms and all firms are risk-neutral profit maximizers. They use both types of labor in producing the single consumption good in the economy.

There are four aggregate variables that are taken as given by firms and workers and pinned down by equilibrium conditions. The first two are labor market tightnesses for skilled and unskilled workers, θ_s and θ_u. These variables are important for firms and workers because they determine the probability that vacancies posted by firms are filled, and, accordingly, the probability that unemployed workers find a job. The other two variables are the values of unemployment for skilled and unskilled workers, U_s and U_u. These are the outside options of workers when bargaining, and so are important determinants of wages. The bargained wage is, for each firm, a function of the number of workers currently employed, as firm size affects the marginal productivities of the different types of workers. The problem of the firm is then to choose a vacancy posting strategy – or, equivalently, firm size – conditional on its specific wage function and on its compliance decision, made at the beginning of time. Workers accept or reject the offers they receive from firms and bargain over wages. An equilibrium is found by determining the values of θ_s, θ_u, U_s and U_u that are consistent with the aggregate behavior of firms and workers.

Labor Markets

We model search frictions following Pissarides (2000). There are two separate labor markets, one for each skill level. Firms need to post vacancies in order to find workers, paying an instantaneous cost ξ per vacancy. The number of matches taking place at each moment is given by a matching function $M(V_i, u_i)$, where V_i and u_i are the measures of open vacancies and unemployed workers in the job market $i \in \{s, u\}$, for skilled and unskilled workers, respectively. We make the standard assumptions that $M(\cdot)$ is increasing in its arguments, concave and has constant returns to scale. This enables us to use the more convenient form $q(\theta_i)$ for the instantaneous probability of filling a vacancy. This means that, over a short time interval dt, the probability that a vacancy gets matched to an unemployed worker is $q(\theta_i)dt$. θ_i is the labor market tightness in market i, that is, the ratio of vacancies to unemployed workers: $\theta_i = \frac{V_i}{U_i}$, $i \in \{s, u\}$. The probability that an unemployed worker finds a job in a small time interval dt is given by $\theta_i q(\theta_i)dt$.

We make no distinction between formal and informal firms in the search process. The aggregate $V_i = V_i^{for} + V_i^{inf}$ is the sum of all vacancies posted by formal and informal firms, and unemployed workers search simultaneously in both sectors. After a worker is matched to a vacancy, the probability that this vacancy is offered by a formal firm is given by $\phi_i = \frac{V_i^{for}}{V_i}$, which is simply the fraction of vacancies posted by formal firms in market i. With this assumption, as with many others, we try to minimize the structural differences between formal and informal sectors and focus instead on the regulatory asymmetries. Our modeling of the search process is most similar to that in Bosch and Esteban-Pretel (2012). Other models with undirected search, such as Ulyssea (2010) and Meghir, Narita and Robin

(2015), assume exogenous differences in the matching technology across sectors.

Problem of the Firm

Firms are endowed with a production function $F(z, n_s, n_u) = F^z(n_s, n_u)$, assumed to be twice differentiable, where n_s and n_u denote units of skilled and unskilled labor. The term z is an exogenous productivity parameter distributed across firms according to a distribution function $G(z)$. We assume that $F^z(\cdot)$ is strictly concave in (n_s, n_u) for any z in the support of $G(z)$, and increasing in z. Moreover, we assume that $\sigma_{z,n_s} < \sigma_{z,n_u}$, where $\sigma_{i,j}$ denotes the partial elasticity of substitution between inputs i and j. Given fixed hiring costs, firms with higher z employ relatively more skilled workers. The parameter z is most easily interpreted as entrepreneurial talent, as in Lucas (1978), with the idea that entrepreneurs cannot efficiently manage a large number of skilled workers if they are not highly talented themselves. For simplicity, there is no firm entry or exit.

Due to search frictions, firms cannot directly choose the amount of labor inputs employed in production. Instead, the control variable is the number of vacancies posted at each instant, $v_s(t)$ and $v_u(t)$. Firms also decide on whether to comply with labor regulations or not. For simplicity, we assume that this decision is taken at the beginning of time and cannot be changed thereafter. If a firm complies, it must pay taxes τ over its total payroll. If a firm chooses instead to hire workers informally, it avoids payroll taxes but incurs in an informality penalty $\rho(n)$, where n is the total number of workers hired by the firm. We assume that $\rho(n)$ is strictly increasing and convex. As in Meghir, Narita and Robin (2015), we do not specify how the informality penalty emerges. In general, it can be seen as the product of the probability of being caught by labor inspectors and the monetary value of the corresponding sanction. It can also encompass the lack of access to some public goods available to formal firms, such as courts.

Skill-biased productivity and the informality penalty are the ingredients behind the aggregate differences that arise in equilibrium across the formal and informal sectors. First, the penalty induces larger firms to formalize. Since larger firms are the most productive ones, it follows that the formal sector has higher average productivity due to selection. Finally, due to skill bias in productivity, there is a higher proportion of skilled workers in formal firms. Still, there are skilled workers employed in the informal sector as well.

Normalizing the price of the final good to 1, the instantaneous profit function of the firm with productivity z, according to its compliance decision j, is

$$\pi^{z,j}(n_s, n_u, v_s, v_u) = \begin{cases} F^z(n_s, n_u) - (1+\tau)\sum_{i=s,u} n_i w_i^{z,for}(n_s, n_u) - (v_s + v_u)\xi, & \text{if } j = for, \text{ and} \\ F^z(n_s, n_u) - \sum_{i=s,u} n_i w_i^{z,inf}(n_s, n_u) - \rho(n_s + n_u) - (v_s + v_u)\xi, & \text{if } j = inf, \end{cases}$$

where $w_i^{z,j}(n_s, n_u)$ is the wage that the firm pays to workers of type i, according to its compliance status j, and the current number of employees, n_s and n_u, and ξ is the cost of posting a vacancy, assumed to be the same across types of workers and sectors (again, in order to minimize structural differences between sectors). We describe how the wage function $w_i^{z,j}(n_s, n_u)$ is determined in the next subsection. From left to right, instantaneous profits are given by total production minus total payroll, payroll taxes (in the case of formal firms) or the informality penalty (for informal firms), and the costs of vacancy posting.

Job relations are destroyed at exogenous separation rates s^{for} and s^{inf}, which depend on the compliance decision. This allows the model to capture the empirical pattern of higher labor turnover among informal firms.[10] The dynamics of labor quantities inside each firm are given by

$$\dot{n}_i = v_i q(\theta_i) - s^j n_i, \text{ with } i \in \{s, u\} \text{ and } j \in \{for, inf\}.$$

The instantaneous variation in the number of workers of type i is equal to the number of vacancies multiplied by the probability that each vacancy is filled, minus the rate of job destruction. In this equation, we implicitly assume that every match turns into a job relation. Later in the paper we show that all job offers are accepted in equilibrium.

The problem of the firm in its recursive Bellman formulation is given by

$$\begin{aligned} \Pi^z \quad &= \quad \max_{j\in\{for,inf\}} \Pi^{z,j}(n_s, n_u), \text{ with} \\ \Pi^{z,j}(n_s, n_u) \quad &= \quad \max_{\{v_s,v_u\}} \left(\frac{1}{1+rdt}\right)\left\{\pi^{z,j}(n_s, n_u, v_s, v_u)\,dt + \Pi^{z,j}(n_s^+, n_u^+)\right\} \\ \text{s.t.} \quad & n_i^+ = n_i + \dot{n}_i dt = \left(1 - s^j dt\right) n_i(t) + v_i q(\theta_i)\,dt, \quad i = s, u. \end{aligned} \tag{2.1}$$

For a firm with productivity z, given a compliance decision j, the total present value of profits is the sum of instantaneous profits earned at the end of the small time interval dt plus the present value of profits after dt. The discount rate r is the same for all firms. Given its initial conditions and productivity, the firm makes the compliance choice that maximizes total profits.

Denote by $J_i^{z,j}(n_s, n_u)$ the marginal value of an additional worker of type i in a firm of type z, with compliance status j: $J_i^{z,j}(n_s, n_u) = \frac{\partial \Pi^{z,j}(n_s,n_u)}{\partial n_i}$. We derive the first order conditions for the firm's problem in Appendix C. From now on, we restrict attention to steady-state solutions where the numbers of workers of different types are constant in each

[10] See the turnover analysis in Gonzaga (2003) and Bosch and Maloney (2010), and also the calibration results in Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015). The existence of high dismissal costs in the formal sector provides strong incentives for keeping an employee. Albrecht, Navarro and Vroman (2009) develop this argument formally, using a search and matching model with endogenous job destruction and an informal sector. Moreover, as mentioned in the introduction, our target equilibrium is the one in which the minimum wage is binding for unskilled workers, who strictly prefer formal employment. Thus, formal employees should also have stronger incentives to maintain the job relation. It would be interesting to use a model with endogenous separation rates, but, in our setting, we do not believe that the gains would offset the additional analytical complexity.

firm. By imposing $\dot{n}_i = 0$ in the F.O.C.'s, the expressions simplify to:

$$(r+s^j)J_i^{z,j}(n_s,n_u) = \begin{cases} F_i^z(n_s,n_u) - (1+\tau)\left[w_i^{z,for}(n_s,n_u) + \sum_{l=s,u} n_l \frac{\partial w_l^{z,for}(\cdot)}{\partial n_i}\right] & \text{, for } j = for \\ F_i^z(n_s,n_u) - \rho'(n_s+n_u) - \left[w_i^{z,inf}(n_s,n_u) + \sum_{l=s,u} n_l \frac{\partial w_l^{z,inf}(\cdot)}{\partial n_i}\right] & \text{, for } j = inf \text{, and} \end{cases} \tag{2.2}$$

$$J_i^{z,j}(n_s,n_u) = \frac{\xi}{q(\theta_i)}, \tag{2.3}$$

with $F_i^z(n_s,n_u) = \frac{\partial F^z(n_s,n_u)}{\partial n_i}$.

Equation 2.2 is an intuitive description of the marginal value of a worker as the discounted sum of expected rents, taking into account the discount rate r and the separation hazard rate s^j. The instantaneous rent is given not only by the difference between marginal product and wage, but also by the effect of this additional employee on the wages of all other workers currently employed by the firm, due to changes in marginal productivities. At the time of the hiring decision or bargaining, previous vacancy costs are sunk and thus do not appear in this expression.

Equation 2.3 is the optimality condition in a steady state. Its interpretation is straightforward: the value of the marginal worker must be equal to the expected cost of hiring another worker, which is the cost ξ per vacancy multiplied by the expected number of vacancies needed to hire a worker. By combining both expressions, we find the first order condition of the firm derived in Cahuc, Marque and Wasmer (2008), in which marginal product equals a generalized notion of marginal cost:

$$\underbrace{F_i^z(n_s^z,n_u^z)}_{\substack{\text{Marginal}\\\text{productivity}}} = (1+\tau)\underbrace{w_i^{z,for}(n_s^z,n_u^z)}_{\text{Own wage}} + (1+\tau)\underbrace{\sum_{l=s,u} n_l \frac{\partial w_l^{z,for}(\cdot)}{\partial n_i}}_{\substack{\text{Effect on other}\\\text{workers' wages}}} + \underbrace{(r+s^{for})\frac{\xi}{q(\theta_i)}}_{\text{Hiring costs}}.$$

We denote the optimal labor choices in firm z as n_s^z and n_u^z (as opposed to arbitrary choices n_s and n_u). The case for informal firms is analogous, just omitting the payroll tax τ and adding the marginal effect of n_i on the informality penalty $\rho(n_s^z + n_u^z)$.

Wage Determination

Wage is determined through Nash bargaining, with workers and firms sharing the rents created by the match. The share of the surplus appropriated by a worker is given by the exogenous parameter σ, which corresponds to the bargaining power of workers. Differently from the standard model in Pissarides (2000), we do not assume homogeneous labor nor constant returns to scale in the production function, and allow workers and firms to engage in renegotiation after the initial match. As discussed in Stole and Zwiebel (1996*a*), these assumptions imply that changes in firm size lead to wage renegotiation due to changes in marginal productivities, and this must be anticipated by firms in their hiring decisions. We follow the solution developed by Cahuc, Marque and Wasmer (2008), who analyze this type of problem in a context with search frictions.

Also differently from many models of informality, such as Ulyssea (2010) and Bosch and Esteban-Pretel (2012), we do not allow formal and informal workers to have different bargaining powers. Once more, this reflects our strategy of minimizing structural differences across sectors. Adding this degree of freedom can be a straightforward way to create a formality wage premium. In our model, worker heterogeneity and minimum wages play this role, while also allowing for a richer pattern of wage dispersion.

We first describe how wages are determined in the absence of a binding minimum wage. Following, we explain how the introduction of a binding minimum wage changes the results. Define $E_i^j(w)$ as the value that workers of type $i \in \{s, u\}$ place on holding a job position of type $j \in \{for, inf\}$ that pays wage w. Also, call U_i the opportunity cost of the worker – that is, the expected present value of being unemployed, which is taken as given by firms and workers. Note that, in a context of mandated benefits which possibly include unemployment benefits, we might be worried that U_i should be a function of factors related to eligibility, such as having worked in a formal firm before or not having reached the maximum number of payments. We avoid this additional complication by including the expected value of unemployment benefits in the expressions for $E_i^{for}(w)$, instead of in U_i, as done by Ulyssea (2010). Since workers are assumed to be risk neutral, this greatly simplifies the solution without loss of generality.

We can write the flow equations that define the value of employment at formal and informal firms with wage w as:

$$rE_i^{for}(w) = a_i w + b_i + s^{for}\left[U_i - E_i^{for}(w)\right], \text{ and} \tag{2.4}$$

$$rE_i^{inf}(w) = w + s^{inf}\left[U_i - E_i^{inf}(w)\right], \tag{2.5}$$

where a_i and b_i represent mandated benefits that may increase (or decrease) the value of holding a formal job.

The value $E_i^j(w) - U_i$ is the rent earned by workers of type i when they accept a job offer in sector j. For firms, the marginal value of a worker of type i is given by $J_i^{z,j}(n_s, n_u)$, which was discussed in the previous subsection. So the Nash sharing rule imposes that the wage function $w_i^{z,j}(n_s, n_u)$ must satisfy[11]

$$(1-\sigma)\left[E_i^j\left(w_i^{z,j}(n_s, n_u)\right) - U_i\right] = \sigma J_i^{z,j}(n_s, n_u) \tag{2.6}$$

[11]Our bargaining expression in the presence of payroll taxes differs from that in Mortensen and Pissarides (2001) because we define σ as the effective bargaining share of workers, while they define their bargaining parameter β as the exponent in the generalized Nash bargaining solution. To convert from their notation to ours, one should use the expression $\sigma = a_i\beta/\left[a_i - (1 + t_i - a_i)(1-\beta)\right]$. Thus, in our comparative statics

where $i \in \{s, u\}$ and $j \in \{for, inf\}$, for all z, n_s, and n_u.

Due to the derivative terms in expression 2.2 (for $J_i^{z,j}$), the set of Nash bargaining equations results in a system of nonlinear differential equations. In Appendix D, we adapt the solution in Cahuc, Marque and Wasmer (2008) to account for two sectors, heterogeneous firms, mandated benefits, and payroll taxes. The resulting wage functions are

$$w_i^{z,for}(n_s, n_u) = \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{1}{1+\tau_i}\int_0^1 \epsilon^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}} \frac{\partial F^z\left(\epsilon^{\frac{1+\tau_s}{a_s}\frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u}\frac{a_i}{1+\tau_i}} n_u\right)}{\partial n_i} d\epsilon$$
$$w_i^{z,inf}(n_s, n_u) = (1-\sigma)rU_i + \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma}} \frac{\partial H^z(\epsilon n_s, \epsilon n_u)}{\partial n_i} d\epsilon,$$

with $c_i = [(1-\sigma)a_i + \sigma(1+\tau_i)]$ and $H^z(n_s, n_u) = F^z(n_s, n_u) - \rho(n_s + n_u)$. Notice that we allow for skill-specific payroll taxes (τ_s and τ_u) in this solution, since we use this result later on in our policy experiments.

As in the solution of the standard bargaining problem with search frictions, wages are a weighted sum of the reservation wage, rU_i, and a term related to the productivity of the marginal worker. In the standard search and matching model, where marginal productivities are not related to firm size, the wage equation reduces to $w_i^{z,for}(n_s, n_u) = \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{\sigma}{c_i}\frac{\partial F^z}{\partial n_i}$ (with $b_i = 0$ and $c_i = 1$ for informal firms). However, with decreasing returns to scale, heterogeneous labor, and intra-firm bargaining, the second term is not simply the marginal productivity of the input considered, but instead a weighted average of infra-marginal productivities, with weights $\epsilon^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}}$ higher for points closer to the margin. We refer the reader to Stole and Zwiebel (1996*b*), Stole and Zwiebel (1996*a*), and Cahuc, Marque and Wasmer (2008) for a detailed discussion of the characterization of this type of solution. In Appendix D, we derive our results and compare them to those from Cahuc, Marque and Wasmer (2008).

Before turning to the outcome of wage bargaining in equilibrium, we introduce a minimum wage into the model. We add it to the baseline model, as opposed to an extension, because minimum wage changes are a first order issue in the Brazilian labor market, as explained in Section 2.2.

If the bargained wage in a formal firm for one type of worker – typically, the unskilled – is lower than the minimum wage, then the minimum wage restriction is binding. The Nash bargaining equation is not satisfied anymore for unskilled workers; indeed, in this situation, these workers receive a share of rents larger than σ. This also implies that the previous wage function for skilled workers is not valid anymore, since the term $\frac{\partial w_u^{z,for}}{\partial n_s}$ in equation 2.2 is equal to zero (marginal changes in the number of skilled workers do not affect wages of unskilled workers, which are binding at the minimum wage). In Appendix D, we show that the wage equation for skilled workers in the formal sector when the minimum wage binds for unskilled workers is

exercises below, the share of rents accruing to workers is always constant, whereas it could vary under Mortensen and Pissarides (2001)'s definition.

$$w_s^{z,for}(n_s, n_u) = \frac{1-\sigma}{c_s}(rU_s - b_s) + \frac{1}{1+\tau_s}\int_0^1 \epsilon^{\frac{1-\sigma}{\sigma}\frac{a_s}{1+\tau_s}} \frac{\partial F^z(\epsilon n_s, n_u)}{\partial n_i} d\epsilon.$$

From the perspective of a firm, whether the minimum wage binds is not only a function of parameters, but also of firm size. This introduces a discontinuity in the first order condition of the problem of the firm. Consider a case where there are complementarities between labor types, as the one in our quantitative exercise. Without a minimum wage, hiring an additional skilled worker decreases skilled wages and increases unskilled wages, and the reverse is true for hiring an unskilled worker. This effect is taken into account in the value of the marginal worker of both types, $J_s^{z,for}$ and $J_u^{z,for}$. However, when the minimum wage becomes binding for unskilled workers, the effect of firm size on unskilled wages disappears, leading to a discontinuous increase in $J_s^{z,for}$ and a discontinuous decrease in $J_u^{z,for}$. The increase in $J_s^{z,for}$, in turn, causes a discrete increase in skilled wages, which might give an incentive for firms to strategically reduce the number of unskilled workers or increase the number of skilled workers – just enough so that bargained unskilled wages are slightly above the minimum wage.

In Appendix D, we show that, because of this discontinuity, there might not be a solution to the first order conditions when the unconstrained (freely bargained) unskilled wage is slightly lower than the minimum wage. In these cases, firms engage in the strategic manipulation of firm size described above.[12] In our quantitative exercises, we deal explicitly with this issue by assuming that firms in this situation choose employment figures that (i) satisfy the first order condition for skilled workers and (ii) lie immediately to the "left" (in terms of n_u) of the region of the (n_s, n_u) space where the minimum wage binds for unskilled workers. Details are laid out in Appendix D.

Now we turn to the analysis of wage determination in equilibrium. If we plug the firm's first order condition (equation 2.3) in the bargaining expression (equation 2.6), and take into account that the bargaining equation is not satisfied if the minimum wage is binding, we have:

$$\begin{aligned}(1-\sigma)\left[E_i^{for}\left(w_i^{z,for}\left(n_s^{for}, n_u^{for}\right)\right) - U_i\right] &\geq \sigma\frac{\xi}{q(\theta_i)}, \quad i \in \{s,u\} \\ (1-\sigma)\left[E_i^{inf}\left(w_i^{z,inf}\left(n_s^{inf}, n_u^{inf}\right)\right) - U_i\right] &= \sigma\frac{\xi}{q(\theta_i)}, \quad i \in \{s,u\}\end{aligned}$$

The first expression is a strict inequality only if $w_i^{z,for} = \bar{w}$,. Because the $E_i^j(\cdot)$ functions

[12] It is not trivial to infer the partial equilibrium consequences of the binding minimum wage on the demand for skilled labor. On the one hand, the minimum wage increases the cost of unskilled labor, which reduces the return to skilled labor due to complementarity between the two inputs. On the other hand, the discontinuity mentioned above increases the return to unskilled labor, going in the opposite direction. In simulation exercises we performed, the effect on the demand for skilled labor was always negative, though in general it should depend on the degree of complementarity between the two factors. Panel A of Appendix Figure B.2 can help understand this discussion.

are monotonic in the wage argument, these equations can be rewritten as:

$$\begin{aligned} w_i^{z,for}\left(n_s^{for}, n_u^{for}\right) &= \max\left\{\frac{1}{a_i}\left[rU_i - b_i + (r+s^{for})\frac{\sigma}{1-\sigma}\frac{\xi}{q(\theta_i)}\right], \bar{w}_i\right\}, \quad i \in \{s,u\} \\ w_i^{z,inf}\left(n_s^{inf}, n_u^{inf}\right) &= rU_i + (r+s^{inf})\frac{\sigma}{1-\sigma}\frac{\xi}{q(\theta_i)}, \quad i \in \{s,u\}. \end{aligned}$$

These expressions do not depend on either firm size or productivity. So, in equilibrium, there are only four wages in this economy: w_s^{for}, w_u^{for}, w_s^{inf} and w_u^{inf}.

This result comes from the assumptions that the matching technology and the cost of posting a vacancy do not depend on firm size or productivity within sector/skill combinations, and that $E_i^j(w)$ does not depend directly on firm size or productivity either (holding sector and wage constant). The intuition behind it is that, regardless of productivity, all firms adjust the number of employees so as to equate the marginal value of workers to the expected search cost, which does not depend on productivity or firm size. Thus, the value added by the marginal worker in equilibrium is the same across the productivity distribution.

Equilibrium

So far, we have described the behavior of firms taking θ_i and U_i as given. In equilibrium, these values have to be consistent with the aggregate behavior of firms and workers. The labor market tightness, as explained in subsection 2.3, is given by the ratio of vacancies to unemployed workers. Define the measure of workers of type i employed in sector j as

$$N_i^j = m\int_{-\infty}^{\infty} n_i^z \,\mathbf{1}\left(\text{Firm } z \text{ chooses compliance } j\right) dG(z).$$

Since, in equilibrium, $\dot{n}_i = 0$ for all firms, $v_i^z = s^j n_i^z / q(\theta_i) \implies V_i^j = s^j N_i^j / q(\theta_i)$. We can therefore find the expressions that pin down θ_i,

$$\theta_s = \frac{s^{for}N_s^{for} + s^{inf}N_s^{inf}}{q(\theta_s)\left(\eta - N_s^{for} + N_s^{inf}\right)} \quad \text{and} \quad \theta_u = \frac{s^{for}N_u^{for} + s^{inf}N_u^{inf}}{q(\theta_u)\left(1 - \eta - N_u^{for} + N_u^{inf}\right)}. \tag{2.7}$$

To find the equilibrium value of U_i, we write the standard flow value equation for the reservation wage:

$$\begin{aligned} rU_i &= \theta_i q(\theta_i)\left[\phi_i E_i^{for}(w_i^{for}) + (1-\phi_i)E_i^{inf}(w_i^{inf}) - U_i\right] \qquad (2.8) \\ &= \begin{cases} \dfrac{\sigma}{1-\sigma}\xi\theta_i & \text{, if } w_i^{for} > \bar{w} \text{ and} \\ \dfrac{\theta_i}{1+\frac{\phi_i\theta_i q(\theta_i)}{r+s^{for}}}\left[\phi_i q(\theta_i)\dfrac{a_i\bar{w}+b_i}{r+s^{for}} + (1-\phi_i)\dfrac{\sigma}{1-\sigma}\xi\right] & \text{, otherwise.} \end{cases} \end{aligned}$$

For simplicity, since we incorporate unemployment benefits in the parameters a_i and b_i, we assume that individuals derive no utility flow from unemployment. The instantaneous return of being unemployed is the expected value of finding a job and leaving unemployment. In case a worker finds a job, which happens with probability $\theta_i q(\theta_i)$, there is a probability $\phi_i = \frac{V_i^{for}}{V_i^{for}+V_i^{inf}} = \frac{s^{for}N_i^{for}}{s^{for}N_i^{for}+s^{inf}N_i^{inf}}$ that the match is with a formal firm. The second expression is the result of inserting the first order condition of the firm, equation 2.3, in 2.8.

An equilibrium in our model is defined as a set of wage functions $w_i^{z,j}(n_s, n_u)$, schedules of firm decisions $j(z)$ and n_i^z, labor market tightnesses θ_i, and unemployment values U_i, such that:

1. the wage functions solve the system of differential equations given by expressions 2.2 and 2.6;
2. the labor schedules n_s^z and n_u^z solve equation 2.3 given the compliance decision $j(z)$ and the wage functions;
3. the compliance decisions $j(z)$ maximize the present value of discounted profits in problem 2.1;
4. the labor market tightnesses are consistent with equation 2.7; and
5. the unemployment values are consistent with equation 2.8.

Note that we do not impose government budget balance in our definition of equilibrium. This choice is motivated by our intended application, since the Brazilian government collects a surplus from salaried job positions (even after paying for mandated benefits and unemployment insurance). In all comparative exercises in the next sections, we show how this surplus varies with changes in model parameters.

Discussion: Compensating Differentials

In the equilibrium bargaining outcomes shown above, is it immediate to see that $E_i^{For}\left(w_i^{for}\right) \geq E_i^{inf}\left(w_i^{inf}\right)$. This expression holds as an equality if the minimum wage is not binding for skill level i. In this case, we can use the definition of $E_i^j(w_i^j)$ to show that

$$w_i^{inf} = \frac{r+s^{inf}}{r+s^{for}}\left(a_i w_i^{for} + b_i\right) - \frac{rU_i\left(s^{inf} - s^{for}\right)}{r+s^{for}}.$$

In words, wages in both sectors adjust to exactly compensate workers for the differences in benefits and job duration across sectors. If the minimum wage is not binding and jobs in both sectors have the same expected duration ($s^{for} = s^{inf}$), then the difference between formal and informal wages is equal to the value that workers attribute to mandated benefits.

If the expected duration in the formal sector is longer, as we see in the data, then the wage differentials should be even higher to compensate for that. If the minimum wage is binding, on the other hand, then this equation is no longer valid: informal wages are lower than the value needed to make workers indifferent between sectors, and formal jobs are strictly preferred. However, workers still accept informal job offers, since it is too costly to remain unemployed and wait for a good job. In this case, formal jobs are rationed in equilibrium and compensating differentials do not hold exactly. Still, informal wages have to be high enough to compensate for the expected benefits of formal jobs, once one also considers the lower probability of obtaining such positions.

On the side of the firms, with a continuous distribution of z, the marginal formal firm is identical to the marginal informal firm. However, employment decisions and wages may differ substantially due to regulatory distortions. It remains true, though, that the marginal firm is indifferent between operating in the formal and informal sectors and is willing to change its compliance status given marginal changes in the parameters.

2.4 Fitting the Model

We fit the model to the Brazilian labor market in 2003, calibrating some of the parameters and estimating others using a minimum distance procedure. We choose 2003 as the baseline year because it is close to the reversal of the informality trend (Figure 2.1) and it is when the second wave of the Informal Urban Economy survey (*Economia Informal Urbana*, ECINF) was conducted by the Brazilian Census Bureau (IBGE). The ECINF targeted small urban firms, most of which were unregistered, thus providing an estimate of the number of informal firms in the economy. We use the survey's micro data in the next section, but, since the ECINF is relatively small and was not repeated after 2003, it is not our main source.

Most of the data we use come from the Monthly Employment Survey (*Pesquisa Mensal do Emprego*, PME), also conducted by IBGE. The PME is a household survey that provides information on employment, wages, occupational choice, formality status, and other characteristics of the workforce, including educational attainment. Because there was an increase in the minimum wage on April 1st, 2003, we restrict the sample to the months of April through December of that year.[13] We use two other data sources from IBGE: the Central Registry of Firms (*Cadastro Central de Empresas*, CEMPRE), a registry of formal firms, and the annual projections of the size of the workforce.

In this section, we show that the model is able to match a series of moments in the data as of 2003. However, the main test of the model is whether it is also able to replicate changes in the Brazilian labor market outcomes from 2003 to 2012, given the changes in observable parameters. This assessment is done in Section 2.5.

[13] When using 2012 data in the next section, we also restrict the sample to the months of April through December to maintain consistency.

Functional Forms

We assume that the production function takes on the following two-level CES functional form:

$$F(z, n_s, n_u) = A\left[Bz{n_s}^{\gamma} + (1-B){n_u}^{\gamma}\right]^{\frac{\alpha}{\gamma}},$$

where A, B, α, and γ are parameters. A is a standard total factor productivity term, while B indicates the relative weight of skilled versus unskilled labor. We restrict the exponent α to be smaller than one, so that the function has decreasing returns to scale in $(n_s,\, n_u)$ for any given z. This production function implies that an entrepreneur with $z = 0$ can still generate output, but only uses unskilled labor. We assume that γ belongs to the interval $(0, 1]$ to ensure that the parameter z denotes skill-biased productivity. In the limiting case where $\gamma = 1$, increases in z only raise the productivity of skilled labor. If $\gamma \in (0, 1)$, unskilled workers are more productive in a firm with a higher z and with more skilled workers.[14]

The parameter z is assumed to follow a Generalized Pareto distribution, to account for the fact that the majority of firms are small but a large part of the workforce is employed by large firms (see IBGE, 2005). We set the location parameter to zero, so that the smallest firms have z arbitrarily close to zero. Also, we normalize the scale parameter to $1 - T$, where T is the shape (tail) parameter, so that average productivity is normalized to one.[15] Increases in T are thus mean-preserving spreads that add probability mass to extreme values of productivity. The cumulative distribution of productivity is given by[16]

$$G(z) = 1 - \left(1 + \frac{Tz}{1-T}\right)^{-\frac{1}{T}}.$$

Since the informality penalty must be increasing and convex, we use a quadratic function, $\rho(n) = Cn^2$. In the specification of the matching technology, we follow the literature and use a Cobb-Douglas function. We thus have $q(\theta) = D\theta^{-E}$, where D is the matching scale and E is the matching elasticity.

Finally, the valuation of fixed benefits by workers takes the form:

$$b_i = \left(b_i^F + s^{for} b_i^D\right)\bar{w}.$$

The term b_i^D is the present value of the expected unemployment insurance flow, measured in multiples of the minimum wage $\bar{w}$, and b_i^F represents transfers received by the worker (also measured in multiples of the minimum wage). The details on the computation of these benefits, along with those on a_i and τ, are provided in Appendix A.

[14] If $\gamma = 0$, the production function collapses to a Cobb-Douglas and the elasticity of substitution between any two pair of inputs, including z, will be the same. If $\gamma < 0$, unskilled labor is a better complement to z than skilled labor.

[15] Allowing for other values for the scale parameter would not add information to the model, since the changes in the scale of z can be offset by changes in the parameters A, B, and γ in the production function.

[16] For computational purposes, we set an upper bound to the distribution and discretize it to $100,000$ atoms. When solving for an equilibrium numerically, the problem of the firm is solved for 20 levels of z and interpolated for the $100,000$ types using cubic splines. These and many other computational details are listed and discussed in Appendix E.

Calibrated Parameters

Table 2.4 presents a first subset of the parameter values we use.

A non-trivial problem in our calibration exercise is how to map observed traits at the individual level to skills in the model. In the model, skills map directly into wages. In the relevant case from the perspective of the quantitative analysis, formal sector minimum wages bind only for unskilled workers. This gives an empirical counterpart of skills for formal workers that does not match perfectly with schooling. Unskilled workers in the model represent workers in the data who receive close to the minimum wage when employed in the formal sector. If they receive significantly more than the minimum wage in a formal job, then they must correspond to skilled workers in the model. As mentioned in section 2.2, there is a wide dispersion of wages for each level of schooling in the data, indicating that the definition of skill in the model does not map easily into schooling (despite being highly correlated with it).

Our approach is to combine an aggregate definition of the share of skilled workers with the individual level implications of the model in terms of the relationship between wages and skills. We assume that the measure η of skilled workers corresponds to the fraction of the workforce with 8 or more years of schooling, but let the quantitative model determine the allocation of workers of different skill levels to the formal and informal sectors based on the distribution of wages observed in the data. Though inevitably somewhat arbitrary, our choice of 8 or more years of schooling to represent skilled workers is based on the distributions of schooling and wages in the Brazilian labor market, discussed in section 2.2, and on the definition of skills that arise from the model (earning more than the minimum wage in the formal sector).[17]

We impute a value for the measure of firms m using the total number of salaried workers and the number of firms, both formal and informal. The PME asks unemployed workers what was the nature of their last employment. We use this information to proxy for the fraction of unemployed workers who are looking for salaried jobs. We estimate that salaried workers, either employed or unemployed, account for 73% of the workforce. Since the PME covers only the 6 main metropolitan regions in Brazil, we multiply this fraction by the total size of the workforce in 2003, calculated by IBGE, to get the total number of salaried workers. We obtain the number of formal firms from CEMPRE and the number of informal firms from ECINF, excluding self-employed workers. The measure m is the ratio of firms to salaried workers.

The job destruction rates s^j are taken from estimates of the duration of employment spells in Gonzaga (2003), who uses the same data set in a similar period.[18] The values for the payroll tax rate and benefits are calculated in Appendix A, according to the methodology suggested by Souza et al. (2012). The discount rate for workers and firms is assumed to be

[17] We cannot let the quantitative model determine the shares of skilled an unskilled workers directly because we want to explore their exogenous change as a driver of reductions in informality.

[18] We use employment duration for formal and informal workers in 2002 (the last year available in Table 1 of Gonzaga 2003) and convert it into a monthly hazard rate of job destruction.

Table 2.4: Parameters Imputed from the Data or from the Literature

Parameter	**Value**	**Source**
η (measure of skilled workers)	0.662	Share 8+ years of schooling
m (measure of firms)	0.0905	Ratio of firms to workforce
s^{for} (formal hazard rate)	0.030	Gonzaga (2003)
s^{inf} (informal hazard rate)	0.082	Gonzaga (2003)
τ (payroll tax rate)	0.7206	Appendix A
a_s, a_u (variable benefits)	0.235, 0.306	Appendix A
b_s^F, b_u^F (fixed benefits)	0.02, 0.05	Appendix A
b_s^D, b_u^D (unemp. insurance)	7.48, 4.00	Appendix A
r (discount rate)	0.008	Real interest rate
D (matching scale)	0.30	Ulyssea (2010)
E (matching elasticity)	0.50	Ulyssea (2010)
σ (worker bargaining power)	0.5	

the real interest rate. We use the same values for the parameters of the matching function as Ulyssea (2010). Finally, we assume symmetric bargaining, meaning that the bargaining power of workers is set to 0.5.

Minimum Distance Estimation

We use a minimum distance procedure to estimate the remaining seven parameters displayed in Table 2.5. The algorithm minimizes differences between a set of eight moments taken from the data, listed in Table 2.6, and the equivalent values implied by the model. Formally, the minimum distance estimator is defined as:

$$\hat{x} = \underset{x \in X}{\operatorname{argmax}} \, [\hat{\pi} - h(x)]' \, W \, [\hat{\pi} - h(x)] \tag{2.9}$$

where $\hat{\pi}$ is the vector of the logarithms of the targets in Table 2.6, x is a vector of the seven parameters being estimated, $h(x)$ is the mapping from the parameter space X to the model outcomes corresponding to the moments $\hat{\pi}$ (measured in logs), and W is a weighting matrix. We use logs rather than levels to define the distance measure in relative terms, thus reducing concerns regarding the scaling of moments.

In this section, we focus on the discussion of the choice of moments and the results of the estimation. Appendix E contains a complete description of the estimation procedure. This description includes the procedure used to solve the model numerically, the selection of starting points, the minimization algorithm used, the choice of a weighting matrix, and the calculation of the standard errors of $\hat{x}$. The estimates discussed in this section used the identity weighting matrix, but results are similar when we use the optimal weighting matrix.

The targets were selected from observable characteristics that are either important for our analysis or informative about parameters that we cannot directly observe. The first two

Table 2.5: Estimated Parameters - Minimum Distance Procedure

Parameter	**Value**	**SE**
A (productivity)	10.2388	0.1674
B (technology bias)	0.6247	0.0042
α (dec. returns)	0.5005	0.0042
γ (CES param.)	0.2800	0.0035
C (informality cost)	0.0796	0.0019
ξ (search cost)	1.0050	0.0236
T (firm dist. shape)	0.1539	0.0070

targets, unemployment and informality rates, are directly observable in the PME data set. The next four targets refer to wage differentials across types of workers and sectors. For all workers in the data, we compute hourly earnings in their main job and divide by the hourly equivalent of the minimum wage. For workers in the formal sector, we consider those who earn up to 120% of the minimum wage as unskilled, and others as skilled. With this definition, we compute the average wage for skilled formal workers, as well as the fraction of unskilled workers in the formal sector, after "Winsorizing" the top and bottom 0.5% of the distribution of hourly wages. In the informal sector, we cannot distinguish between skilled and unskilled workers, and so we compute only the average wage among all informal employees. However, we can set a reasonable target for the informal wage penalty among unskilled workers from Bargain and Kwenda (2011). Using the same PME data set and quantile fixed-effects regressions, they find that, for salaried workers at the quantile 0.2 of the wage distribution, the wage penalty associated with informality is around 7.5%.

The labor share of income is defined in the model as the fraction of total production (net of search costs and informality penalties) that is not firm profits nor government surplus. Although not particularly related to our analysis, this is a sensible way to add information to pin down the concavity of the production function, since the latter is directly related to profits. We calculate the empirical counterpart of this measure using the National Accounts System, applying the corrections proposed in Gollin (2002). The last target, the fraction of salaried workers employed in firms with 10 or fewer employees, is set as a means to determine the shape parameter of the productivity distribution. We use 10 workers as the threshold to match the employer size question in the PME survey, which has "11 or more employees" as the top bracket.

Table 2.6 shows that the estimated model matches all target variables with considerable accuracy. Moreover, the standard errors of the estimated parameters are very small, due mostly to the very large sample sizes from the PME survey.

Before we proceed to the next subsection, it is interesting to use our baseline specification to characterize some properties of the equilibrium, particularly as it relates to the cross-

Table 2.6: Moments Used in Estimation

Outcomes	**Model Value**	**Target Value**	**Target SE**
Unemployment	12.7%	12.6%	0.11%
Share informal workers	29.1%	28.4%	0.20%
Formal skilled wage	4.09	4.00	0.02
Unskilled formal workers	11.8%	11.7%	0.18%
Informal unskilled wage	0.929	0.925	0.004
Avg. informal wage	2.45	2.52	0.02
Labor share of income	52.6%	52.8%	0.28%
% workers in firms 10 or less	23.5%	23.5%	0.18%

Note: Wages in multiples of the minimum wage in 2003, the numeraire in the model.

Table 2.7: Firms in the Model

Percentile	z	**Size**	**Fraction Skilled**	**Formal?**
Smallest	0.00	0.96	0.0%	No
50%	0.62	1.85	9.9%	No
75%	1.31	3.87	26.9%	No
90%	2.34	9.01	53.2%	No
95%	3.22	15.7	70.1%	No
97.5%	4.20	46.7	76.9%	Yes
99%	5.67	119.6	83.5%	Yes
Top 0.01%	17.0	4,899	95.9%	Yes

Note: Wages in model units (one model unit is equivalent to the minimum wage in 2003).

sectional distribution of firms. Each row in Table 2.7 describes firms in a specific position in the distribution of productivity. The top row refers to the smallest firms in the model and the bottom row refers to the largest ones. The columns show the productivity parameter, the number of workers, the fraction of skilled workers, and the compliance status. The model generates an equilibrium where the fraction of skilled workers increases monotonically with firm size (in both the informal and formal sectors) and formal firms are larger than informal ones. This profile reproduces patterns observed in the data (as in Table 2.2) but incompatible with previous search models of informality: the presence of skilled and unskilled workers in both sectors and a higher share of skilled workers in formal firms. Also interestingly, the smallest firms in the model have approximately one employee, even though this is not imposed as a restriction.

2.5 Quantitative Results

The Recent Reduction in Informality in Brazil

We use the model to analyze the behavior of the Brazilian labor market between 2003 and 2012. First, we look at the main exogenous changes observed during the period and analyze how each of them separately affected the labor market. In order to validate the performance of the model, when possible, we confront these comparative statics exercises with the empirical evidence currently available from reduced-form estimates. Then we evaluate whether the model is able to account for the aggregate movements in informality, unemployment, and wages by considering changes in all exogenous variables simultaneously.

Throughout the analysis, we often refer to Table 2.8, where each row contains a particular labor market outcome. The first column describes how the Brazilian labor market changed from 2003 to 2012 using the same data and definitions used in the calibration. Each following column considers how changes in one or more parameters affect labor market outcomes in the model, by comparing the baseline calibration with a new steady-state equilibrium where only the parameters in question are set to their 2012 levels.

In the period we study, the unemployment rate fell by 7.2 percentage points and the informality rate dropped by 10.7 points. Average wages increased by 28%, but, as mentioned in section 2.2, the gains were larger for low-skill formal workers and for informal workers. Informal wages, for example, increased by 42%, as compared to a wage growth of 22% for formal skilled workers.

Minimum Wage

The minimum wage increased by 61.2% from 2003 to 2012. The effects of a change of this magnitude in the calibrated model are shown in column 2 of Table 2.8. Wages for skilled workers in both sectors are only marginally affected. However, for informal workers, wages fall by 3.2%. The reason for this decline is the reduced demand for unskilled labor by formal firms, which increases unemployment and lowers the outside option of workers being hired by informal firms.

From the changes in the minimum wage alone, informality increases by 5.7 percentage points and unemployment by 1.6 percentage point. The increase in unemployment seems small when compared to the magnitude of the increase in the minimum wage. The reason is that part of the effect on unemployment is attenuated by marginal firms entering informality (and thus not being subject to the minimum wage anymore), and also by the fact that informal unskilled wages decrease, leading to increased labor demand by informal firms.

This logic resembles the traditional view of the informal sector, where for some workers informality is an alternative to unemployment (Fields, 1975; Rauch, 1991; Boeri and Garibaldi, 2007, for example). In our model, this applies to unskilled workers when the minimum wage binds, in the sense that formal jobs are strictly preferred to informal ones,

Table 2.8: Quantitative Experiments, Changes in the Brazilian Labor Market between 2003 and 2012

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Changes in:	Data	Model							
Outcomes		Minimum wage	Payroll tax	Benefits	Enforcement	Fraction skilled	All but productivity	Productivity	All
Unemployment (p.p.)	-7.2	1.6	0.0	0.0	0.4	-6.0	-3.1	-3.2	-5.0
Informality (p.p.)	-10.7	5.7	-0.3	0.2	-3.4	-12.6	-8.9	-1.2	-9.1
Wages (%):									
Average	28.4	0.6	0.2	0.0	-0.8	1.2	2.2	24.3	28.4
Formal, skilled	22.4	-1.5	0.3	0.0	0.0	-8.8	-9.6	26.5	14.2
Formal, unskilled	61.2	61.2	0.0	0.0	0.0	9.6	61.2	0.0	61.2
Informal	42.2	-3.2	-0.4	0.3	-8.4	48.5	16.2	44.4	61.7
Product^a (%)	27.0	-1.8	0.1	0.0	0.3	8.0	6.9	26.6	35.5
Govt. net revenues (%)^b	-	-23.8	-1.1	1.4	3.9	7.1	-15.5	40.1	29.7

Notes: ^{a}For column 1, product is GDP per capita. For the remaining columns, product is total production in the model net of search costs and the informality penalty. ^{b}In the baseline calibration, the government appropriates 8.4% of total production by salaried workers. Numbers in this line represent relative changes over that baseline amount, not changes in fraction of total production appropriated.

but are also more difficult to find, so unskilled workers accept informal job offers to avoid unemployment.

The increase in informality following a rise in the minimum wage generated by the model, accompanied by a more timid increase in unemployment, is in line with evidence from the Brazilian labor market. Though there are no well identified studies of the labor market response to increases in the minimum wage currently available, the existing evidence, such as Foguel, Ramos and Carneiro (2001) and Lemos (2009), seems to indicate that informality tends to rise and employment responds only mildly – if at all – to minimum wage increases. Our comparative statics reproduces the qualitative patterns documented by the empirical literature on minimum wages in Brazil.

It is also worth mentioning that there is a reduction of 1.8% in aggregate production following the increase in the minimum wage. Net government revenues experience a more sizable decline of 23.8%. This is mainly because some benefits that accrue to all formal workers, such as unemployment insurance, are indexed by minimum wages. On the other hand, revenues from labor taxes increase only for unskilled workers, and increased informality and unemployment reduce the tax base.

Payroll Taxes

The only change in labor market regulations from 2003 to 2012 was the phasing out of a temporary additional contribution to the worker's severance payment fund (*Fundo de Garantia por Tempo de Servico*, FGTS). As described in Appendix A, we calculate that this change decreased the final payroll tax rate only slightly, from 72.06% of the nominal wage to 71.43%. Column 3 shows that, as standard models would predict, informality falls following the reduction in the payroll contribution. Wages rise for all workers, except for those who receive exactly the minimum wage. This is a consequence of the axiomatic bargaining approach, through which workers receive part of the increased profits of firms. Product rises and government revenues decline. All effects are quantitatively small.

Mandated Benefits

There were minor changes in mandated benefits, specifically in the formulas for calculating the income tax and social security contributions, which are both deducted from the wage of formal employees and thus are included in our parameter a_i. However, on average, they did not result in sizable changes in deductions. When we recalculate the parameters a_i and b_i using 2012 data (Appendix A), we find that the differences are negligible. Hence, they do not have any relevant impact on labor market outcomes, as column 4 from Table 2.8 shows.

Enforcement of Regulation

We use data from the Ministry of Labor to estimate changes in the enforcement of labor regulations from 2003 to 2012. Reports of labor inspections, available in MTE (2013), show that the number of workers targeted by inspections rose during the last decade both in absolute terms and as a fraction of the workforce.[19] We use the relative increase as a proxy for increases in enforcement in the model. We find that the fraction of the workforce targeted by inspections rose by about 33.9% from 2003 to 2012. We therefore raise the parameter C in the model by this same proportion.

The fifth column of Table 2.8 shows how this change impacts our baseline calibration. First, informality decreases by 3.4 percentage points, as expected. We argued in section 2.2 that the effects of increased enforcement on unemployment are ambiguous in many models, and this is also true in ours. There is an extensive margin effect because firms that change their compliance decision may hire more workers, and also an intensive margin effect because the remaining informal firms hire fewer workers. In our calibration, unemployment increases by 0.4 percentage point with the increase in enforcement. The qualitative responses of informality and unemployment generated by the model are consistent with reduced-form evidence from exogenous variations in labor inspections provided by Almeida and Carneiro (2012). The only noticeable change in wages is a decline in earnings for informal workers. In this respect, our model replicates the results found in Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015). Government revenues increase, but one should be cautious about drawing additional implications from this result since we do not take into account operational costs associated with increased enforcement.

Workforce Composition

Over the last decades, there has been a consistent increase in school attendance among Brazilian school-aged children, reaching near universalization of primary schooling by the late 1990's.[20] This has led to a corresponding improvement in the educational composition of the workforce not only because young adults are now more educated than previous generations but also because more individuals enter the labor market at later ages. At the same time, demographic changes associated with historical reductions in fertility and population aging are leading to an older and, therefore, more experienced workforce. From 2003 to 2012, the fraction of the workforce with complete elementary education – in Brazil, 8 or more years of schooling – increased from 66.0% to 78.9%. In column 6 of Table 2.8, we assume that this change of 12.9 percentage points corresponds to the increase in the fraction of skilled workers in the model.

We find that the predicted changes are in line with our discussion from section 2.2. Both unemployment and informality decrease sharply as a consequence of a more skilled workforce, falling, respectively, by 6 and 12.6 percentage points. Wages for informal workers increase by 48.5%, while they decrease for skilled formal workers by 8.8%. This is a direct consequence of the relative increase in the supply of skilled workers. The labor market for skilled workers becomes less tight (and the reverse happens for unskilled workers). Because firms hire more

[19] Other indicators, such as total revenues from fines, also increased during the period. For a thorough discussion of enforcement of labor regulation in Brazil, see Cardoso and Lage (2005).

[20] The share of children 7-14 years old attending school goes from 80.9% in 1980 to 97% in 1999 (de Oliveira, 2007).

skilled labor in the new equilibrium, the productivity of unskilled work increases due to complementarities in the production function. The combination of a tighter labor market for unskilled labor and higher productivity is behind the steep increase in the informal wage. Wages for unskilled formal workers also rise, by 9.6%, meaning that the minimum wage is not binding anymore in the new equilibrium. The model therefore predicts that, absent the observed increases in minimum wages between 2003 and 2012, the minimum wage would have become non-binding under the 2012 composition of the Brazilian labor force.

The large effect of the workforce composition on informality works through both the intensive and extensive margins. With the reduction in the market tightness for skilled workers, formal firms, which are intensive in skilled labor, face stronger incentives to grow than informal firms, shifting part of the skilled labor force from the informal to the formal sector. Since skilled and unskilled labor are complements in the production function, this also leads to a higher productivity of unskilled labor in the formal sector and, therefore, to a shift of part of the unskilled labor force as well to the formal sector. At the extensive margin, an analogous phenomenon happens. The marginal informal firms, which are close to indifferent between the formal and informal sectors, start choosing formality due to the increased incentives to grow from the increased supply of skilled workers.

We know of no reduced-form empirical study that analyzes the aggregate labor market effect of changes in the educational composition of the labor force. Various papers, such as Menezes Filho, Mendes and de Almeida (2004), describe the strongly positive individual-level correlation between schooling and formality. Other papers, such as Barbosa Filho and Moura (2015), assume a stable individual-level relationship between schooling and informality and perform Oaxaca-Blinder type exercises analyzing the role of demographic changes as determinants of changes in informality. But no paper allows for the possibility that changes in the educational composition of the labor force directly affect labor market equilibrium outcomes, conditional on individual schooling. This highlights the relevance of the type of analysis conducted in this paper, where we can systematically address the endogenous labor market response to this type of compositional changes.

In Appendix F, we provide some reduced-form evidence related to these qualitative predictions of the model. We use Brazilian census data from 1991, 2000, and 2010 and look at equilibrium outcomes at the local labor market (micro-region) level. Exogenous changes in the educational composition of the labor force are difficult to obtain in this setting, so we interpret the results simply as correlations between changes in composition in each local labor market and labor market equilibrium outcomes. The results show that an increase in the fraction of skilled workers is associated with increases in formality, as predicted by the theory. In particular, this result holds conditional on individual level schooling, meaning that it does not reflect only a mechanic increase in formality due to a higher and stable probability of formal employment among more educated workers. A higher fraction of skilled workers is positively associated with the probability of formal employment even for given educational levels. Results related to employment are less robust. We do find a positive and significant correlation between the fraction of skilled workers and employment under some specifications, but most results are quantitatively small and not statistically significant. This may be due to a different utility from unemployment between skilled and unskilled workers, a possibility not considered in the model. We refer the interested reader to the detailed discussion in the appendix.

Estimating Changes in Productivity

Now we consider the performance of the model when the five dimensions discussed above are brought together. The results are shown in column 7. These changes explain 83% of the observed decline in informality, but only 43% of the decline in unemployment. Also, average wages and product increase by far less than observed in the data, explaining in both cases just a fraction of the actual change. The increase in GDP per capita in the data – listed as product in the table – reflects in part an overall increase in TFP in the Brazilian economy during this period. Ferreira and Veloso (2013), for example, estimate an yearly growth of TFP in Brazil between 1.5% and 2.5% per year from 2003 to 2009. These observations suggest that there was an increase in overall productivity in the economy that, not surprisingly, cannot be captured by the model, which does not display capital accumulation or technological change. To calibrate TFP gains in the model, we raise the parameter A until the increase in average wages – taking into account all changes during the period – matches that observed in the data. We find that TFP in the model must increase by 25.3% between 2003 and 2012 in order for the model to reproduce the increase in average wages in the data. This number falls close to the upper bound of the cumulative growth in TFP that would be obtained from the Ferreira and Veloso (2013) estimates.

Before we assess the performance of the model including the increase in productivity, we study the effects of productivity gains in isolation. Column 8 shows that unemployment declines by 3.2 percentage points and wages rise by 24.3% when productivity increases. There is also a reduction of 1.2 percentage point in informality, consistent with many other models where informal employment is countercyclical. This particular pattern generated by the model – with unemployment and informality being countercyclical, but the former responding more than the latter to changes in aggregated conditions – is also consistent with the empirical evidence for Brazil presented in Bosch and Esteban-Pretel (2012). Wages rise for most workers, but particularly in the informal sector. This is because most unemployed workers in the baseline calibration are unskilled, and thus the decline in unemployment has larger effects on the tightness of the unskilled labor market and, therefore, on the informal sector. Wages do not rise for formal unskilled workers because the minimum wage is still binding after the productivity gain.

Explaining the Evolution of Labor Market Outcomes

In column 9, we consider changes in minimum wages, taxes, benefits, enforcement, skills, and productivity simultaneously. The qualitative implications of the model, in terms of direction of predicted changes, matches precisely the pattern of movements observed in the Brazilian labor market between 2003 and 2012: reductions in unemployment, reductions in informality,

Table 2.9: Individual Contribution of Each Factor, Changes in the Brazilian Labor Market from 2003 to 2012

	All	All changes, except:					
Outcomes:	changes	Minimum wage	Payroll tax	Benefits	Enforcement	Fraction skilled	Productivity
Unemployment (p.p.)	-5.0	-6.5	-5.0	-5.0	-5.0	0.3	-3.1
Informality (p.p.)	-9.1	-11.0	-8.8	-9.3	-6.1	4.0	-8.9
Wages (%):							
Average	28.4	27.5	28.1	28.4	29.7	25.7	2.2
Formal, skilled	14.2	14.8	13.9	14.2	14.2	24.8	-9.6
Formal, unskilled	61.2	38.3	61.2	61.2	61.2	61.2	61.2
Informal	61.7	79.6	61.8	61.5	69.9	19.8	16.2
Product[b] (%)	35.5	35.7	35.4	35.5	35.0	24.5	6.9
Govt. net revenues (%)[c]	29.7	42.1	31.3	27.0	24.8	17.6	-15.5

Notes: [a]Change from 2003 to 2007 (IBGE/SCN is only data available up to 2007). [b]Product is total production in the model net of search costs and the informality penalty. [c]Numbers in this line represent relative changes in government surplus over the baseline amount in 2003, as in Table 2.8.

and increases in average wages, with proportionally larger gains for informal and unskilled workers. Quantitatively, the model does a good job in explaining the reduction in informality, generating a decline of 9.1 percentage points while the observed decline was 10.7 points. It also predicts a decline in unemployment of 5 percentage points, which corresponds to 69% of the observed decline of 7.2 points. Predictions regarding wages are close to the empirical patterns, though the model underestimates by more than a third the gains for formal skilled workers. Overall, the model is able to explain quantitatively the main outcomes of the Brazilian labor market with a reasonable degree of precision.

Going back to the discussion in section 2.2, we can use the model to determine which factor was the main driver behind the reductions in informality and unemployment. Table 2.8 already addressed this issue, by looking at the effect of each factor one at a time. In Table 2.9, we conduct the opposite comparative statics exercise: we analyze what happens in the model when all but one of the factors discussed before is taken into account. We find that the declines in both unemployment and informality would have been considerably larger – respectively, 6.5 and 11 percentage points – if the minimum wage had not increased. We also reinforce the idea that the change in labor force composition was the main driver behind the observed reductions in informality: without a larger fraction of skilled workers, informality would have increased by 4 percentage points, instead of declining by 9.1, and unemployment would have remained roughly stable. In short, the model is unable to reproduce the reductions in informality and unemployment when changes in labor force composition are ignored. The relevance of enforcement (informality penalty) is of second order: without changes in this parameter, the decline in informality would have been three percentage points smaller. As before, the effects of changes in payroll taxes and benefits are negligible.

To strengthen our argument and to show that changes in workforce composition are strictly necessary to replicate the patterns observed in the data, we conduct an additional exercise. Suppose that we want to explain the evolution of labor market outcomes in the model without resorting to changes in the fraction of skilled workers. Since we directly observe minimum wages, payroll taxes, and benefits in the data, we have two degrees of freedom in this exercise: aggregate productivity and enforcement (informality penalty). We therefore choose the total factor productivity parameter and the informality penalty such that the model, with a fixed composition of the labor force, reproduces the same declines in informality and unemployment from column 9 in Table 2.9. In order to match these numbers, productivity would have to increase by 101% and the costs of informality would have to increase by around 216%. No estimates of productivity and enforcement currently available suggest increases remotely similar to these magnitudes. In addition, under this scenario, product per capita and average wages would have gone up by close to 100%, and wage increases would have been roughly homogeneous across sectors and skill levels. These results are clearly at odds with the data, suggesting that changes in workforce composition are really essential in any attempt to rationalize the changes in labor market outcomes observed in Brazil between 2003 and 2012.

Finally, this exercise also shows that the impact of productivity on informality may depend on the initial level of unemployment. While an increase in A starting from the baseline model led to a mild decline in informality (column 8 Table 2.8), the same change led to no noticeable impact using parameters of 2012 (the difference between columns 9 and 7 in the same Table). In our model, increases in productivity can lead to more formalization because firms hire more workers and the informality penalty is increasing in firm size. On the other hand, more productivity leads to higher wages, and thus increased taxes. If the economy has high unemployment, firms can hire more workers without putting too much pressure on wages, since marginal productivities decrease with firm size. In this case, the firm size effect dominates and informality is reduced. If instead unemployment is low, firms cannot grow much with gains in productivity and wages increase more to sustain the new labor market equilibrium. Then, payroll taxes increase relative to the informality penalty and marginal firms may decide to switch to the informal sector. The net effect of increased productivity on informality is ambiguous. Our results suggest that this theoretical ambiguity may indeed be quantitatively relevant. It also shows that increases in productivity alone are not enough to rationalize the changes seen in the Brazilian context.

As in any other theory, we use several simplifying assumptions when modeling the labor market. Nevertheless, we believe that the most restrictive assumptions in our model go against our key results, in the sense that they dampen the ability of the model to explain decreases in unemployment and informality. This is clearly the case with the exogenous job destruction rate. Even after improvements in firm level productivity, the rate of dismissals remains the same, attenuating the reduction in unemployment. Likewise, assuming exogenous distribution of firm productivities and ruling out firm entry does not allow for an increase in the number of large, high-productivity firms, whose profits increase disproportionately after the changes observed from 2003 to 2012 are taken into account.[21] If high productivity firms entered the market, the ensuing increase in average productivity would be an additional force towards employment and formalization.

Policy Experiments

How to bring down informality without increasing unemployment has been a major policy challenge in developing countries. In this subsection, we use the model to assess the effectiveness of alternative labor market policies in achieving this goal, while also keeping track of the fiscal burden imposed on the government. This exercise illustrates that the framework developed in the paper can also be a useful tool for policy analysis.

The first policy we consider is a reduction in payroll taxes for low wage workers. In column 3 from Table 2.8, we showed that a lower payroll tax rate can lead to a decline in informality with no adverse effect on unemployment. On the other hand, it also leads to a reduction in government revenues that is substantial when compared to the decline in informality. However, informal firms are relatively more intensive in unskilled labor. In addition, only a fraction of government revenues come from payroll taxes on low skill workers, since their wages are lower and they account for a small fraction of formal employment. Thus, an intermediate alternative might be for governments to subsidize the employment of low wage formal workers through a progressive payroll tax, with the tax rate increasing with the wage. Proposals like this have been considered as ways to subsidize low wage workers in developed countries for different theoretical reasons (see Chéron, Hairault and Langot, 2008; Lee and Saez, 2012; Robin and Roux, 2002), but rarely feature in the informality discussion in the developing world.

In Table 2.10, we examine the progressive payroll tax policy using as starting point the model as of 2012 (column 9 in Table 2.8). In the first column, we show as a reference point the result of simply reducing the overall tax rate by 1 percentage point (to 0.7043). As argued above, although this reduction leads to reductions in informality, there are significant costs in terms of government revenue. In columns 2 and 3, we consider similar policies where the reduction in payroll taxes is restricted to low wage workers (in the model, equivalent to low skill). The policy achieves similar or better results for employment and formalization and, in addition, for some values of τ_u government revenues actually increase. The formalization induced by lower taxes among low skill workers is sufficient to induce marginal firms to comply, and thus enlarges the tax base. The taxes raised from skilled workers in firms that formalize more than offset the foregone revenue from low skill workers in infra-marginal firms. On top of that, wages increase substantially for unskilled workers in the informal sector because of a tighter labor market. This policy therefore is also likely to have positive effects on poverty alleviation.

Next, we consider an apparently similar policy in which the government increases the attractiveness of formal jobs to unskilled workers by increasing benefits for low wage earners

[21] The smallest firms have a loss of 16% in profits, while the largest ones gain 38%.

Table 2.10: Hypothetical Policy Experiments

Outcomes	(1) 1 p.p. reduction in payroll tax $\tau = 0.7043$	(2) Progressive payroll tax $\tau_s = 0.7143$, $\tau_u = 0.7043$	(3) Progressive payroll tax $\tau_s = 0.7143$, $\tau_u = 0.50$	(4) Transfer to low wage $b_u^F = 0.10$	(5) Transfer to low wage $\tau = 0.7441$, $b_u^F = 0.10$
Unemp. (p.p.)	-0.1	0.0	-0.4	0.1	0.3
Inform. (p.p.)	-0.5	-0.1	-2.8	0.0	1.4
Wages (%):					
Average	0.3	0.0	-0.1	0.0	-0.9
Formal, skilled	0.5	0.0	0.4	0.0	-1.5
Formal, unsk.	0.0	0.0	0.0	0.0	0.0
Informal	-0.1	0.1	3.1	-0.4	-0.2
Product[a] (%)	0.1	0.0	0.5	0.0	-0.3
Govt. revenues[b] (%)	-2.0	0.0	1.3	-5.7	0.0

Notes: In all columns, the reference is the model as of 2012, with $\tau = 0.7143$ and $b_u^F = 0.05$.[a]Product is total production net of search costs and the informality penalty.

in the formal sector. This policy is similar to a current program in Brazil in which the government transfers resources directly to low wage employees in the formal sector (*Abono Salarial*). In column 4, we assess the consequences of increasing the fixed payments from the government to low-skilled workers from 5% of the minimum wage to 10%. We find that there is no reduction in informality, despite the sizable costs incurred by the government. If payroll taxes are raised by about 3 percentage points, so that the program breaks even in terms of government revenue, the policy leads to increases in both informality and unemployment (column 5).

The second policy is ineffective because of the binding minimum wage. In the unconstrained case, the formal unskilled wage would drop after the increase in benefits because of rent sharing between worker and firm. This would generate incentives for the posting of more formal unskilled vacancies, and the results would come closer to those of the progressive payroll tax. With a binding minimum wage, however, wages cannot adjust downward so the supply of formal vacancies remains unchanged. The only channel left for lowering informality is the increase in informal wages, which results from an increase in the outside option of unemployed unskilled workers when bargaining (because formal jobs become more attractive).

Three important caveats should be made regarding our progressive payroll tax results. First, our model assumes that every firm hires both skilled and unskilled workers. This enables the government to increase its revenues by inducing firms to formalize through lower taxes for unskilled workers. If firms instead hire a single type of worker – either skilled or unskilled – there would be less potential to increase revenues with this policy (depending on the initial distribution of firms across formal and informal sectors). The second limitation is the assumption that there is a single compliance decision for all workers. If firms are free to make individual compliance decisions for each worker, then the policy would merely result in the formalization of low wage workers, while high wage employees would remain informal. Third, there is the possibility of under-reporting of wages in the formal sector, so as to disguise skilled workers as unskilled, which is not taken into account in the model.

We believe that these concerns are not enough to compromise the qualitative implications of the analysis, though the quantitative results from Table 2.10 should not be taken at face value. To assess the relevance of the first two issues, we examine data from the ECINF, which surveyed small firms in the formal and informal sectors in Brazil. For each of the small firms covered by ECINF, we have information on number of employees, formal status, wages, and schooling levels. Regarding the first point, we examine the degree of wage dispersion within firms in the informal sector. In 64% of the informal firms with five employees – the largest firms surveyed by ECINF and those more likely to be marginal firms in the informal sector – the highest wage was at least 50% above the lowest wage. In 20% of them, the highest wage was more than three times the lowest wage. The data also show that, in most of these firms, workers belong to very different educational categories. This evidence suggests that there is a substantial degree of skill heterogeneity within marginal informal firms, as implied by the model.

On the second point, the formalization of low wage workers should increase the probability of formalization of high wage workers for two reasons. If firms formalize a fraction of their workforce, they become more visible to labor inspectors and thus the cost of employing informal workers increases. Also, the existence of formal ties to some workers may make it easier for others to take the employer to court. The data support the view that most firms hire all of their workers either formally or informally. Among firms in the ECINF data set with five employees, 32% hire all workers informally, while 46% hire all of them formally. Only 22% of the firms have both formal and informal employees. This number is even lower for smaller firms.

Finally, although this policy would certainly increase incentives to under-report wages, there are already large incentives for firms to do so under current labor law, since several contributions and taxes are proportional to earnings (see Appendix A). In addition, the value of many mandated benefits is also indexed by the contractual wage, so workers have an incentive to enforce truthful reporting by firms. It does not seem to be the case that the progressive payroll tax would dramatically change the incentives for under-reporting already present in the formal sector.

2.6 Concluding Remarks

This paper studies how the interplay between workforce composition and labor market institutions, particularly minimum wages, affects informality, unemployment and wages. In order to incorporate these factors, we propose a search and matching framework in which firms use heterogeneous types of labor and face decreasing returns to scale. In addition, we model the compliance decision by firms and workers, so that agents self-select into formal and informal sectors given their individual characteristics and the institutional setting. In the model, there are no intrinsic differences between individuals and firms in the formal and informal sectors, and all market imperfections are generated by labor regulations and search and matching frictions.

The model is used to reproduce the cross-sectional characteristics of the Brazilian labor market and to study the decline in informality rates observed between 2003 and 2012. We show that the model is able to replicate important features of informal labor markets, particularly wage patterns and rates of unemployment and informality. Following, we use changes in tax rates, benefits, minimum wage, enforcement of regulation, workforce composition, and productivity, to show that the model replicates with considerable precision the evolution of labor market outcomes in Brazil. The improvement in the educational composition of the labor force is the most important factor behind the sharp decline in informality among salaried workers observed during the period, though changes in minimum wages and productivity are also key for rationalizing other patterns observed in the data. The search and matching framework we develop is essential for these issues to be simultaneously taken into account in the analysis.

We also perform additional exercises to analyze the impact of two policies aimed at reducing informality. First, we show that decreasing the payroll tax rate for low wage workers can have positive effects on both employment and formalization, while at the same time increasing government revenues. On the other hand, a subsidy to formal unskilled workers is not cost-effective. The discrepancy between these two policies comes from the binding minimum wage, which prevents downward adjustments of formal wages and the creation of more formal jobs in the second alternative. The model indicates that a change from flat to progressive payroll taxes could be an effective way to fight informality in the developing world. This application highlights the potential use of the model for policy analysis and the quantitative relevance of the new dimensions it brings to the table.

Bibliography

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

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

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

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

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

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

Albrecht, James, Lucas Navarro, and Susan Vroman. 2009. "The Effects of Labour Market Policies in an Economy with an Informal Sector." *The Economic Journal*, 119(539): 1105–1129.

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.

Amaral, Pedro S., and Erwan Quintin. 2006. "A competitive model of the informal sector." *Journal of Monetary Economics*, 53(7): 1541–1553.

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.

Araujo, Luis, and Vladimir P. Ponczek. 2011. "Informal wages in an economy with active labor courts." *Textos para discussão, Escola de Economia de São Paulo, Getulio Vargas Foundation (Brazil)*, 294.

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

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

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

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

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

Badaoui, Eliane El, Eric Strobl, and Frank Walsh. 2010. "The formal sector wage premium and firm size." *Journal of Development Economics*, 91(1): 37–47.

Barbosa Filho, Fernando Holanda, and Rodrigo Leandro Moura. 2015. "Evolução Recente da Informalidade do Emprego no Brasil: Uma Análise Segundo Características da Oferta e Demanda de Trabalho." *Pesquisa e Planejamento Econômico*, 45: 101–123.

Bargain, Olivier, and Prudence Kwenda. 2011. "Earnings Structures, Informal Employment, and Self-Employment: New Evidence From Brazil, Mexico and South Africa." *Review of Income and Wealth*, 57: S100–S122.

Bargain, Olivier, Eliane El Badaoui, Prudence Kwenda, Eric Strobl, and Frank Walsh. 2012. "The Formal Sector Wage Premium and Firm Size for Self-employed Workers." *IZA Discussion Paper*, 6604.

Barros, Ricardo, and Carlos Henrique Corseuil. 2001. "The Impact of Regulations on Brazilian Labor Market Performance." *Inter-American Development Bank Research Network Working paper*, R-427.

Boeri, Tito, and Pietro Garibaldi. 2007. "Shadow Sorting." *NBER International Seminar on Macroeconomics*, 2005: 125–163.

Boeri, Tito, Pietro Garibaldi, and Marta Ribeiro. 2011. "The Lighthouse Effect and Beyond." *Review of Income and Wealth*, 57: S54–S78.

Bosch, Mariano, and Julen Esteban-Pretel. 2012. "Job creation and job destruction in the presence of informal markets." *Journal of Development Economics*, 98(2): 270 – 286.

Bosch, Mariano, and William F. Maloney. 2010. "Comparative analysis of labor market dynamics using Markov processes: An application to informality." *Labour Economics*, 17(4): 621–631.

Bosch, Mariano, Edwin Goñi-Pacchioni, and William Maloney. 2012. "Trade liberalization, labor reforms and formal-informal employment dynamics." *Labour Economics*, 19(5): 653 – 667.

Botelho, Fernando, and Vladimir Ponczek. 2011. "Segmentation in the Brazilian Labor Market." *Economic Development and Cultural Change*, 59(2): 437–463.

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

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

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

Cahuc, Pierre, Francois Marque, and Etienne Wasmer. 2008. "A Theory of Wages and Labor Demand with Intra-firm Bargaining and Matching Frictions." *International Economic Review*, 49(3): 943–972.

Cain, Glen G. 1976. "The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey." *Journal of Economic Literature*, 14(4): 1215–1257.

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

Carbonnier, Clément. 2015. "Payroll Taxation, qualifications, wages and unemployment rates in a frictional labor market with productive interactions between segments." Unpublished manuscript.

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

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

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

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

Cardoso, Adalberto, and Telma Lage. 2005. "A Inspeção do Trabalho no Brasil." *Revista de Ciências Sociais*, 48(3): 451–490.

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

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

Chamberlain, Gary. 1994. "Quantile regression, censoring, and the structure of wages." In *Advances in Econometrics: Sixth World Congress*. Vol. 2, , ed. Christopher Sims, 171–209.

Chéron, Arnaud, Jean-Olivier Hairault, and François Langot. 2008. "A quantitative evaluation of payroll tax subsidies for low-wage workers: An equilibrium search approach." *Journal of Public Economics*, 92(3-4): 817–843.

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

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

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

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

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

de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2013. "The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka." *American Economic Journal: Applied Economics*, 5(2): 122–50.

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

de Oliveira, Romualdo Portela. 2007. "Da universalização do ensino fundamental ao desafio da qualidade: uma análise histórica." *Educação & Sociedade*, 28(100): 661–690.

Dickens, William T., and Kevin Lang. 1985. "A Test of Dual Labor Market Theory." *The American Economic Review*, 75(4): pp. 792–805.

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

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

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

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

Fajnzylber, Pablo, William F. Maloney, and Gabriel V. Montes-Rojas. 2011. "Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program." *Journal of Development Economics*, 94(2): 262 – 276.

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

Ferreira, Pedro C., and Fernando Veloso. 2013. "O Desenvolvimento Econômico Brasileiro no Pós-Guerra." In *Desenvolvimento Econômico: Uma Perspectiva Brasileira.* , ed. Fernando Veloso, Pedro C. Ferreira, Fabio Giambiagi and Samuel Pessôa, 129–165. Rio de Janeiro:Campus.

Fields, Gary S. 1975. "Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs." *Journal of Development Economics*, 2(2): 165 – 187.

Fields, G. S. 1990. "Labour market modelling and the urban informal sector: Theory and evidence." In *The informal sector revisited.* , ed. D. Turnham, B. Salomé and A. Schwarz, 49–69. Organisation for Economic Cooperation and Development.

Foguel, Miguel, Lauro Ramos, and Francisco Carneiro. 2001. "The Impacts of the Minimum Wage on the Labor Market, Poverty and Fiscal Budget in Brazil." *IPEA Textos para Discussao*, 839: 1–42.

Galiani, Sebastian, and Frederico Weinschelbaum. 2012. "Modelling Informality Formally: Households and Firms." *Economic Inquiry*, 50: 821–838.

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

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

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

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

Günther, Isabel, and Andrey Launov. 2012. "Informal employment in developing countries: Opportunity or last resort?" *Journal of Development Economics*, 97(1): 88–98.

Gollin, Douglas. 2002. "Getting Income Shares Right." *Journal of Political Economy*, 110(2): 458–474.

Gonzaga, Gustavo. 2003. "Labor Turnover and Labor Legislation in Brazil." *Economía*, 4(1): 165–222.

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

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

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

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

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

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

IBGE. 2005. "Estatísticas do Cadastro Central de Empresas 2003." Available at http://www.ibge.gov.br/home/estatistica/economia/ cadastroempresa/2003/default.shtm.

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

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

Kovak, Brian. 2013. "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" *American Economic Review*, 103: 1960–1976.

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

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

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

La Porta, Rafael, and Andrei Shleifer. 2008. "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity*, 2008(2): 275–363.

Lee, David, and Emmanuel Saez. 2012. "Optimal minimum wage policy in competitive labor markets." *Journal of Public Economics*, 96: 739–749.

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.

Lehman, Hartmut, and Norberto Pignatti. 2007. "Informal Employment Relationships and Labor Market Segmentation in Transition Economies: Evidence from Ukraine." *IZA Discussion Paper*, 3269.

Lemos, Sara. 2009. "Minimum Wage Effects in a Developing Country." *Labour Economics*, 16: 224–237.

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

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

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

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

Lucas, Robert E. Jr. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics*, 9(2): 508–523.

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

Maloney, William F. 2004. "Informality Revisited." *World Development*, 32(7): 1159–1178.

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

Marrufo, Grecia. 2001. "The Incidence of Social Security Taxes in Economies with Partial Compliance: Evidence from the SS Reform in Mexico." PhD diss. University of Chicago.

Meghir, Costas, Renata Narita, and Jean-Marc Robin. 2015. "Wages and Informality in Developing Countries." *American Economic Review*, 105(4): 1509–46.

Mello, Rafael F., and Daniel D. Santos. 2009. "Aceleração educacional e a queda recente da informalidade." *Boletim Mercado de Trabalho IPEA*, 39: 27–33.

Menezes Filho, Naercio Aquino, Marcos Mendes, and Eduardo Simoes. de Almeida. 2004. "O Diferencial de Salarios Formal-Informal no Brasil: Segmentação ou Viés de Seleção?" *Revista Brasileira de Economia*, 58(2): 235–48.

Monteiro, Joana C.M., and Juliano J. Assunção. 2012. "Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises." *Journal of Development Economics*, 99(1): 105 – 115.

Mortensen, Dale T., and Christopher Pissarides. 2001. "Taxes, Subsidies and Equilibrium Labour Market Outcomes." *CEPR Discurssion Paper 2989*.

MTE. 2013. "Resultados da Fiscalização do Trabalho - Nível Brasil - 2003 a 2012." Available at http://portal.mte.gov.br/fisca_trab/estatisticas.htm.

Perry, Guillermo, William F. Maloney, Omar Arias, Pablo Fajnzylber, Andrew Mason, and Jaime Saavedra. 2007. *Informality: Exit and Exclusion.* World Bank.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory.* MIT Press.

Pratap, Sangeeta, and Erwan Quintin. 2006. "Are labor markets segmented in developing countries? A semiparametric approach." *European Economic Review*, 50(7): 1817–1841.

Rauch, James E. 1991. "Modelling the informal sector formally." *Journal of Development Economics*, 35(1): 33 – 47.

Robin, Jean-Marc, and Sébastien Roux. 2002. "An equilibrium model of the labour market with endogenous capital and two-sided search." *Annales d'Economie et de Statistique*, , (67): 257–307.

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

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

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

Souza, André Portela, Sérgio P. Firpo, Vladimir Ponczek, Eduardo Zylberstajn, and Felipe Ribeiro. 2012. *Custo do Trabalho no Brasil: Proposta de uma nova metodologia de mensuração.* FGV/EESP.

Stole, Lars A., and Jeffrey Zwiebel. 1996*a*. "Intra-Firm Bargaining under Non-Binding Contracts." *The Review of Economic Studies*, 63(3): 375–410.

Stole, Lars A., and Jeffrey Zwiebel. 1996*b*. "Organizational Design and Technology Choice under Intrafirm Bargaining." *The American Economic Review*, 86(1): 195–222.

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.

Tornarolli, Leopoldo, Diego Battistón, Leonardo Gasparini, and Pablo Gluzmann. 2012. "Exploring trends in labor informality in Latin America, 1990-2010." Center for Distributive, Labor and Social Studies (CEDLAS), Universidad Nacional de La Plata.

Ulyssea, Gabriel. 2010. "Regulation of entry, labor market institutions and the informal sector." *Journal of Development Economics*, 91(1): 87 – 99.

Ulyssea, Gabriel. 2014. "Firms, Informality, and Development: Theory and Evidence from Brazil." Unpublished manuscript.

Appendix A

Appendix to Chapter 1

A.1 Proofs

Section 2: Task-based production function

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

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

Lemma 3. *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. □

Lemma 4. *Any candidate solution with slack of labor is not optimal.*

Proof. Consider two cases:

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

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

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

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

Proof of Lemma 1, except non-arbitrage condition. From Lemma 4, we know that any optimal solution must not have any slack. The same Lemma implies that any candidate solution satisfying the conditions in Lemma 3 is also not optimal. So any optimal solution must be such that for any two tasks $x_1 < x_2$ and two types $h_1 < h_2$, $m_{h_2}(x_1) > 0 \Rightarrow m_{h_1}(x_2) = 0$ and $m_{h_1}(x_2) > 0 \Rightarrow m_{h_2}(x_1) = 0$. This property can be re-stated as: for any pair of types $h_1 < h_2$, there exists at least one number ${}_{h_1}\bar{x}_{h_2}$ such that $m_{h_2}(x) = 0 \; \forall x < {}_{h_1}\bar{x}_{h_2}$ and $m_{h_1}(x) = 0 \; \forall x > {}_{h_1}\bar{x}_{h_2}$. By combining all such requirements together, there must be $H - 1$

numbers $\bar{x}_1, \ldots, \bar{x}_{H-1}$ such that, for any type h, $m_h(x) = 0 \; \forall x \notin [\bar{x}_{h-1}, \bar{x}_h]$ (where $\bar{x}_0 = 0$ and $\bar{x}_H = \infty$ are introduced to simplify notation).

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

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

Proof of the non-arbitrage condition (Equation 1.1) is provided in the next section of this Appendix. □

Proof of Proposition 1: curvature of the task-based production function and non-arbitrage condition (Equation 1.1)

Constant returns to scale and concavity follow easily from the definition of the production function. Let's start with concavity. Suppose that there are two input vectors $\boldsymbol{l}^1$ 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 5. *The excess labor demand function is* C^2*.*

Proof. We need to show that, for all components $z_h(\cdot)$, the second partial derivatives exist and are continuous. This is immediate for the first derivatives regarding q and $\boldsymbol{l}$, 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 6. *The Jacobian of the excess labor demand function regarding* $(q, \bar{x}_1, \ldots, \bar{x}_{H-1})$*, when evaluated at a point where* $\boldsymbol{z}(\cdot) = \mathbf{0}_{H\times 1}$*, 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 5 and 6 mean that the implicit function theorem can be used at the solution to the assignment problem to obtain derivatives of the solutions to the system of equations imposed by the labor constraints. These solutions are $q(\boldsymbol{l}) = f(\boldsymbol{l}, b_g(\cdot))$ 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 1.1, along with the constant returns relationship $q = \sum_h l_h f_h$, emerge as first order conditions, after noting that the Lagrange multipliers are marginal productivities.

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

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

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

Proof of Corollary 1: Distance-dependent complementarity

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

Proof of Corollary 2: Effect of minimum wages

In a competitive economy where the minimum wage binds for only the lowest worker type, a marginal increase in the minimum wage causes unemployment of that lowest type up to the point where marginal revenue product of labor equals wages. That is exactly the same comparative statics as reducing the supply of the least skilled labor type by that corresponding amount. The proportional effect of that change on marginal productivities of all labor types is given by the negative of the elasticity of complementarity times the share of labor costs attributable to the lowest worker types. Distant-dependent complementarity constrains that effect to be decreasing over the wage distribution (because the elasticities of complementarity are increasing).

Section 3: Markets and wages

Proof of Lemma 2: Firm problem and representative firms

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

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

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

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

$$\pi_g(a_j) = \max_{\boldsymbol{l}_j} p_g f\left(\boldsymbol{l}_j, b_g(\cdot)\right) - \sum_{h=1}^{H} \tilde{C}_h\left(l_{hj}, a_j\right)$$
$$\text{where } \tilde{C}_h\left(l_{hj}, a_j\right) = \min_{w_{hj}} C_h\left(w_{hj}, \underline{\epsilon}_h(l_{hj}, w_{hj}, a_j), a_j\right)$$

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

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

where l_{hj}^{+} is the amount of efficiency units supplied by workers who earn more than the minimum wage. This derivative is strictly increasing in w_{hj}, which means that the cost-minimizing wage $w_h(l_{hj}, a_j)$ is defined by the first order condition:

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

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

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

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

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

To obtain the first order conditions in terms of w_{hj} and $\underline{\epsilon}_{hj}$, one can take the corresponding derivatives of (1.5) and (1.6). Alternatively, one can obtain the optimal cutoff by equating the marginal cost of labor above to the marginal revenue product of labor, and then use the first order condition on cost minimization to find the mark-down rule.

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

Proof of Proposition 2: Wage differentials across firms

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

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

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. □

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

Proof of part 1: From Lemma 7, $f_h(\boldsymbol{l}, b_g(\cdot)) = z_g f_h(\boldsymbol{l}, b(\cdot))$. 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 (1.7), (1.8) and the zero profits condition (1.11) imposing the conditions from this proposition:

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

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

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

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: $\underline{\epsilon}_{hg} = 0$. In addition, the labor supply curve becomes isoelastic with identical elasticities for all worker types:

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

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

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

Also, rewrite the zero profit condition as:

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

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

The corresponding prices that lead to zero profits are:

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

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

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

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

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

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

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

□

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

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

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

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

Proof of Proposition 4: Changes in firm costs affect the returns to skill

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

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

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

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 7, 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 (1.1) in Lemma 1, along with log-supermodularity of $e_h(x)$, that the statement to be proved is equivalent to

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

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

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

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

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

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

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

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

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 $Q'_2/Q'_1 > Q_2/Q_1$. Thus, Lemma 8 implies that wage gaps increase as stated in the Proposition.

A.2 Numerical implementation

The basic logic of obtaining compensated labor demands in this model is to use the non-arbitrage equation 1.1 from Lemma 1 to obtain thresholds as functions of marginal productivity gaps. Then, compensated labor demands can be obtained through numerical integration of Equation 1.2.

The exponential-Gamma parametrization is helpful because it provides a simple closed form solution for thresholds and the labor demand integrals. Let:

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

Then:

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

$$\begin{aligned}\ell_{hg}\left(\bar{x}_{h-1}, \bar{x}_h\right) &= \int_{\bar{x}_{h-1}}^{\bar{x}_h} \frac{b_g(x)}{e_h(x)} dx \\ &= \begin{cases} \frac{1}{z_g \Gamma(k_g)} \left(\frac{k_g}{\Upsilon_{hg}\theta_g}\right)^{k_g} \left[\gamma\left(\Upsilon_{hg}\bar{x}_h, k_g\right) - \gamma\left(\Upsilon_{hg}\bar{x}_{h-1}, k_g\right)\right] & \text{if } \Upsilon_{hg} \neq 0 \\ \frac{1}{z_g \Gamma(k_g)} \left(1 - k_g\right)^{-k_g} \left[\gamma\left(\bar{x}_{h-1}(1 - k_g)/\theta_g, k_g\right) - \gamma\left(\bar{x}_h(1 - k_g)/\theta_g, k_g\right)\right] & \text{otherwise} \end{cases}\end{aligned} \tag{A.2}$$

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

When $\Upsilon_{hg} < 0$, one needs to use the standard holomorphic extension of the incomplete Gamma function. This extension is readily available in many numerical software packages.

In Matlab, Equation A.2 can be implemented using the *gammainc* function; it will handle both positive and negative values of $\Upsilon_{hg} < 0$.

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

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

That series converges slowly if Υ_{hg} is large in magnitude. But Υ_{hg} is bounded below at -1. Thus, one can use the power series representation for negative Υ_{hg} and Equation A.2 for other values.

Calculating the production function and its derivatives — that is, solving for output and marginal productivities given labor inputs — is not needed in the equilibrium computation nor in estimation. However, it might be useful for other purposes. Those numbers are obtained from a system of H equations implied by requiring that labor demand equals labor available to the firm. The choice variables can be either $(q, \bar{x}_1, \ldots, \bar{x}_{H-1})$ or $f_1, \ldots, f_H$. Moving from thresholds and output to marginal productivities, or vice-versa, is a matter of applying the constant returns relation $\sum_h f_h = q$.

A.3 Appendix to the quantitative exercise

Summary statistics

Descriptive statistics for the RAIS dataset are presented in Table A.1. Statistics are presented for the whole sample and separately by schooling group.

Wage inequality and schooling trends using PNAD data

In this Appendix, I analyze the robustness of the main facts presented in Section 1.4 using an alternative data source, the PNAD survey. I proceed in three steps. First, I compare wage inequality and schooling trends for formal sector workers in the two datasets. Second, I expand the sample to include both formal and informal workers to check whether these trends are restricted to the formal sector. Third, I look at schooling achievement for Brazilian adults regardless of their workforce participation status, as a way of investigating whether increased schooling achievement among employed workers reflects changes in selection patterns into employment or fundamental changes in access to schooling for the whole population.

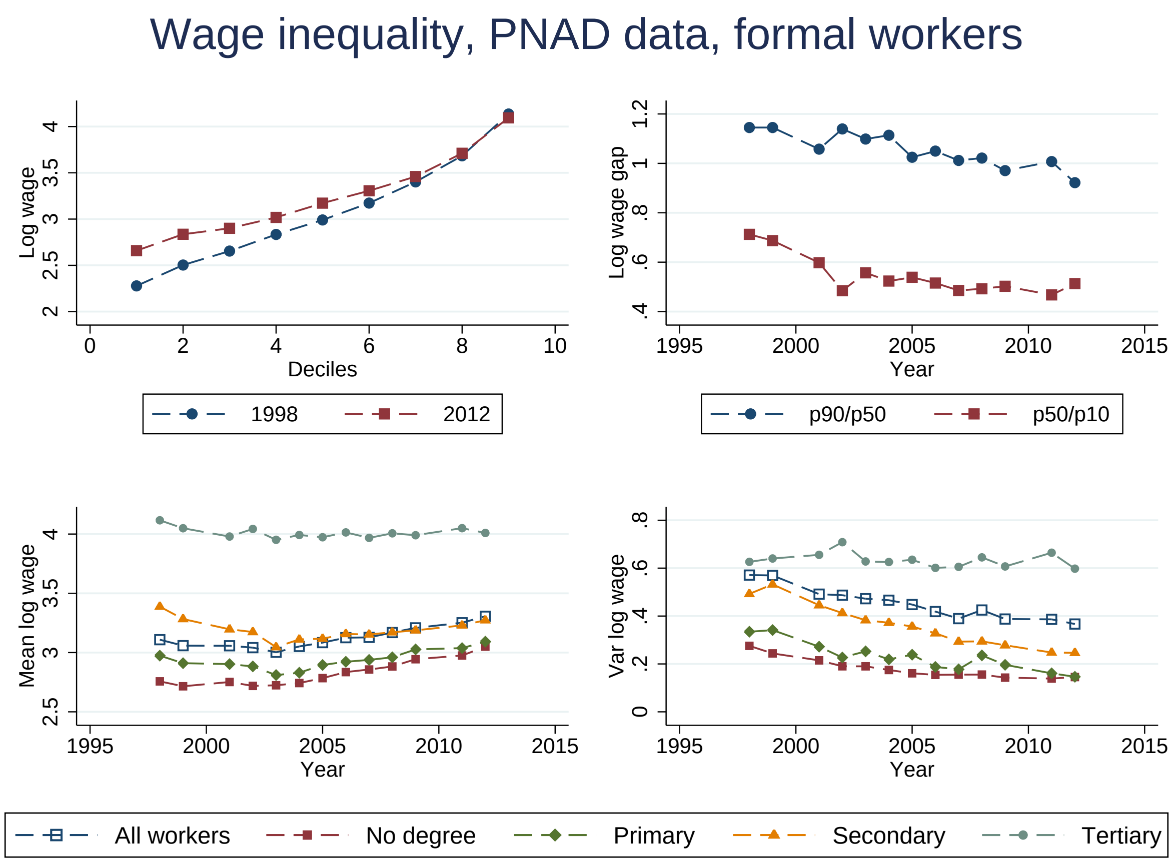
The PNAD is a household survey with national coverage administered by the by the Brazilian Statistical Bureau (IBGE). Jointly with the Census, it is one of the primary sources of nationally representative data on a series of topics that include labor market participation, earnings, and education. It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

Table A.1: Summary statistics, RAIS data.

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

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

Figure A.1: Measures of wage dispersion, PNAD data, formal sector



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

whether workers had a signed "labor card" — that is, whether the employment relationship is formally registered.

This Appendix analyzes PNAD data from 1998 through 2012. The sample I use includes adults 18 through 54 years old that are not in school, the same criterion imposed on RAIS data. I use public use software developed by PUC-Rio's Datazoom project to read the data, make it compatible across years, and deflate income variables. More information about the resulting dataset is available at Datazoom's website.[1]

Comparing RAIS data and PNAD data for formal sector workers

Figure A.1 replicates Figure 1.4 using PNAD data instead of RAIS data. The PNAD sample is constructed to match the RAIS sample, including only formal employees. Overall, the patterns are broadly similar: they show decreased wage inequality along different dimensions.

[1]Currently located at `http://www.econ.puc-rio.br/datazoom/english/index.html`.

There are two significant differences. First, the mean log wage gap between college and high school workers is stable from 1998 to 2012 in PNAD, but increasing in RAIS. Second, variances of log wages within groups and for the whole sample are larger with the RAIS data for 1998, but not for 2012. Thus, RAIS shows larger reductions in inequality using this measure.

The first panel in Figure A.3 replicates the evolution of schooling achievement of formal employees, shown in Figure 1.5. Again, the overall patterns are broadly similar: there is a substantial decline in the share of hours supplies by workers without any educational degree, accompanied by a similarly large increase in the percentage of hours supplied by workers with complete high school (secondary). There is also an increase in hours supplied by workers with college degrees. There are small changes in the shares; in particular, the PNAD shows a higher fraction of college-educated workers.

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

Inequality trends for the whole workforce

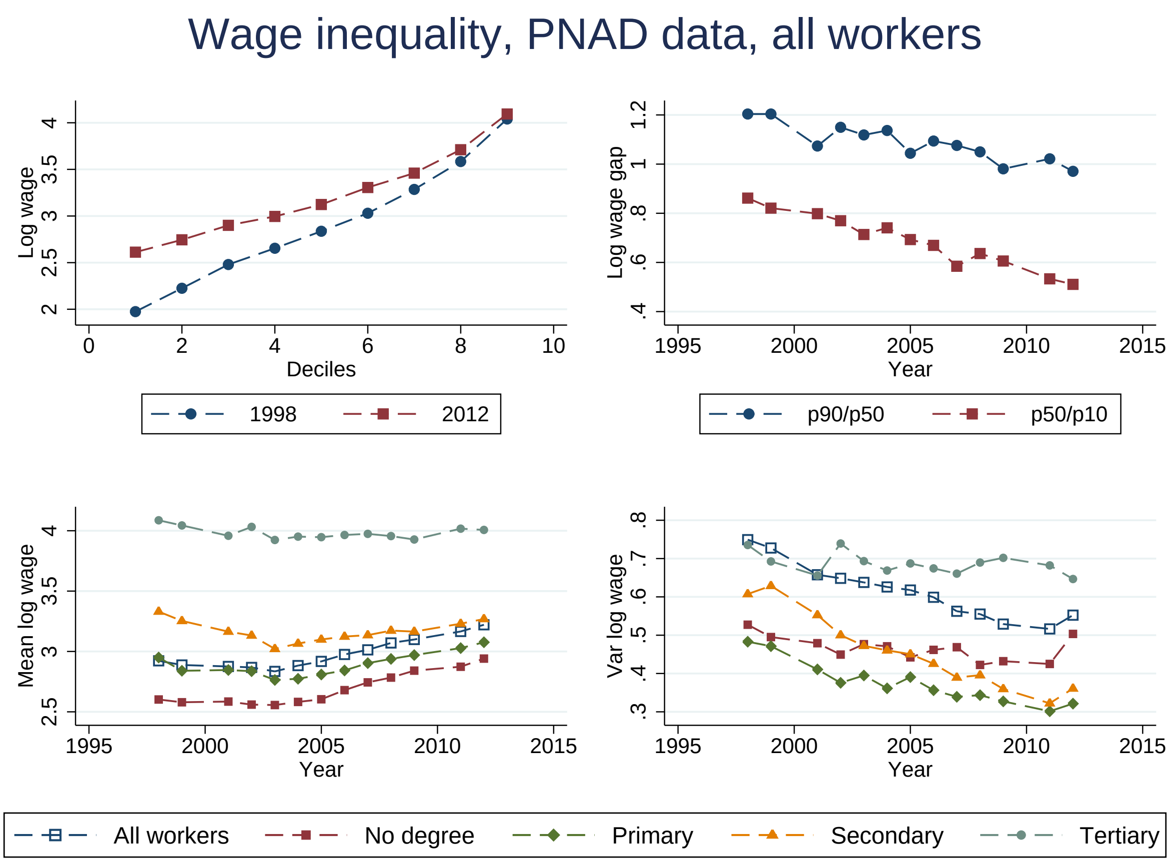
Figure A.2 is constructed similarly to Figure A.1 above, but the data includes both formal and informal workers. I use a broad definition of the informal sector that includes domestic and self-employed workers. There are no substantial changes in qualitative patterns once informal workers are taken into account. The amount of wage dispersion is higher for the whole sample than for the restricted sample, especially in the lower tail of the wage distribution. One possible candidate for these differences is the presence of the binding minimum wage.

Differences in schooling achievement between the formal sample and the full sample can be observed by comparing the first two panels in Figure A.3. Formal sector workers are a selected subsample with higher education levels. However, trends for the whole sample are, again, similar to those obtained from the formal sample.

Changes in relative labor supply

The first two panels of Figure A.3, along with Figure 1.5 in the main text, show shares of hours worked supplied by each schooling group. One might wonder whether these could reflect changes in selection patterns into employment over time (coming, e.g., from business

Figure A.2: Measures of wage dispersion, PNAD data, all workers



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

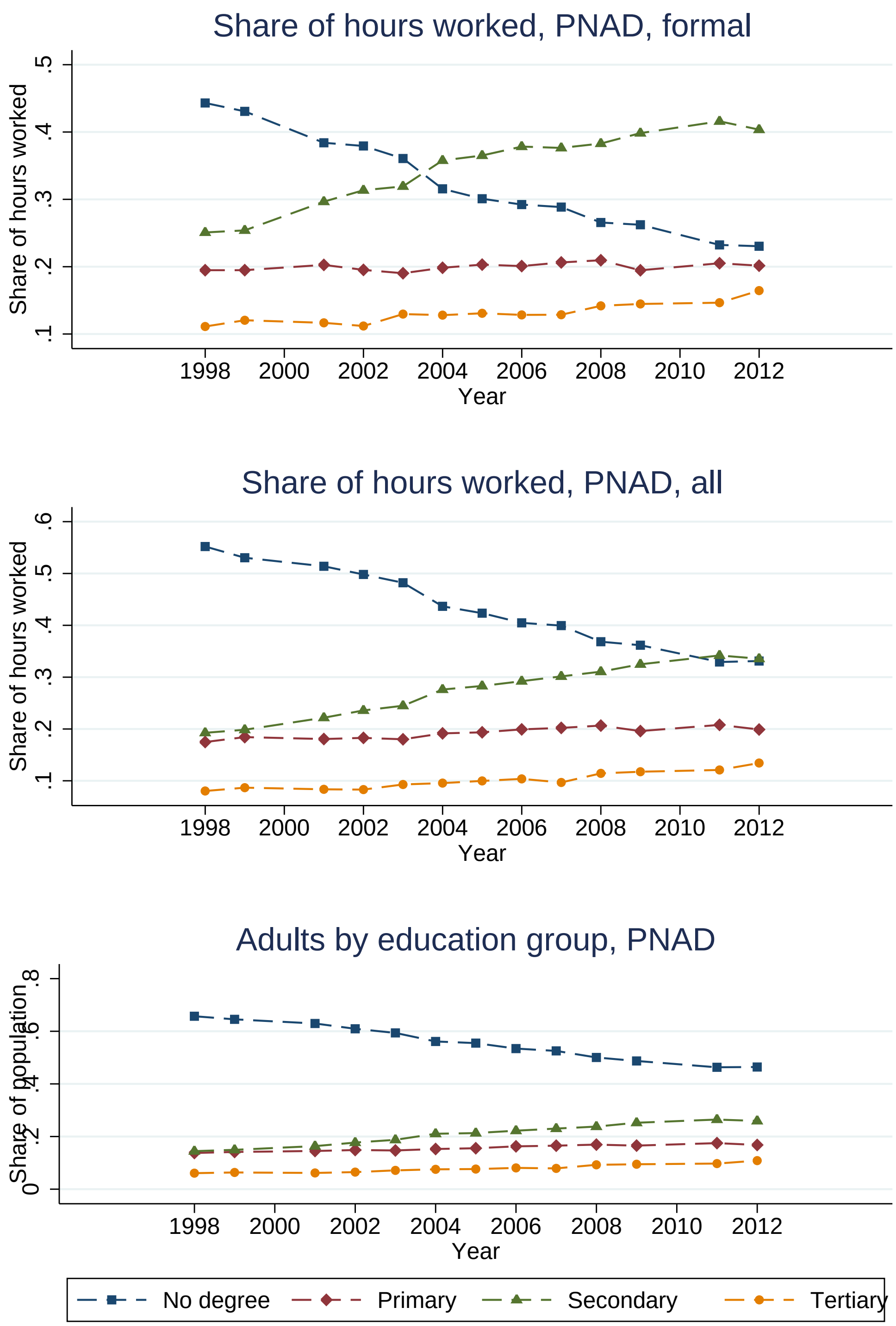
cycle fluctuations) instead of changes in labor supply. The third panel in Figure A.3 shows that this is not the case. That graph shows the share of adults out of school, aged 18 through 54, in each educational group — regardless of whether they are employed, looking for jobs, or not in the labor force. The changes in educational achievement from that figure are similar in magnitude to those in the second and first panels. The levels are different, though, suggesting selection into employment by education.

Variance decomposition using Kline, Saggio and Sølvsten (2018)

The estimation of variance components follows the methodology proposed in Kline, Saggio and Sølvsten (2018), henceforth KSS. For the 1998 period, I use data for two years: 1997 and 1999. I use non-consecutive years to increase the number of firm-to-firm transitions.

The sample used for estimation is the largest leave-one-out connected set. This concept differs from the usual connected set in matched employer-employee datasets because it

Figure A.3: Changes in educational achievement, PNAD data



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

Table A.2: Sample sizes in variance decomposition exercise

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

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

requires that firms need to be connected by at least two movers, such that removing any worker from the sample does not disconnect this set. Table A.2 presents the size of that largest connected set in each period.

I implement the variance decomposition using the code provided by KSS.[2] There are some implementation choices required in this estimation, stated below:

- Dealing with controls (year fixed effects): "Partialled out" prior to estimation (option 1 in the resid_controls argument).
- Computation of local linear regressions: stratified by grids, separate for movers and stayers (option 2 in the subsample_llr_fit argument).
- Sample selection: includes both movers and stayers (option 0 in the restrict_movers argument).
- Algorithm: Random projection method (option "JLL" in type_of_algorithm option, with epsilon=0.005).

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

Appendix B

Appendix to Chapter 2

B.1 Costs of Formal Labor and Valuation of Benefits by the Formal Employee

In this Appendix, we calculate the cost of formal employment and the valuation of mandated benefits by formal workers based on the methodology of Souza et al. (2012). In each subsection, we first show the results for the baseline calibration in October 2003. Then, we discuss the changes in regulations from 2003 to 2012 and calculate the parameters for October 2012.

In order to correctly reflect labor regulations and the differences between formal and informal jobs, it is important to have a clear grasp of what we call wage in the model and how it relates to the data. In the data set we use (PME), workers are asked to report their nominal monthly wages. If they are formal, they are asked not to include annual contributions such as the thirteenth salary. On the other hand, they report gross wages before formal deductions (such as income tax or social security contributions). However, if workers are informal, such concerns are irrelevant and the reported wage is actually what is being paid by the employer and received by the worker. On the employer side, a similar distinction must be made: while the cost of informal employment is essentially the reported wage, for formal workers the cost might be much higher once all contributions and mandated benefits are taken into account.

In the model, wages should reflect the reported wage in the PME data set, and the payroll tax (τ) and the $benefits$ term are used to adjust the costs of formal employment and the valuation of formal jobs by employees, respectively. Thus, for the purposes of the model, the payroll tax rate must encompass everything that a formal employer must pay but a informal employer must not, as a multiple of the reported wage. Likewise, the term $benefits$ is the difference between the valuation of formal jobs and reported wage. In principle, this term can be either positive or negative, depending on whether the advantages of formal employment (e.g., thirteenth salary, vacations) are quantitatively more important than the social security and income tax deductions. In the calculations below, we show that all parameters of the $benefits$ term are positive, meaning that formal jobs are preferred to informal jobs for a given reported wage.

Costs of Formal Labor

Under Brazilian labor laws, contributions paid by employees are fixed fractions of the base salary. Thus, the payroll tax rate is the same regardless of the type of worker in the model. Later, we discuss that this is not true regarding the valuation of formal jobs by employees; for instance, highly paid workers are subject to income tax, but low wage workers are not.

Table B.1 shows our calculations of the cost of formal employment in October 2003. For simplicity, we normalize the base salary to 100. Formal workers are entitled to a thirteenth salary annually and an additional stipend of 1/3 of the monthly wage when they leave for vacation. In addition, if they are dismissed, the employer must notify them at least 30 days earlier. During that period, the employee is entitled to use up to 25% of its work time in job search. As discussed in Gonzaga (2003), the advance notification is in practice an additional severance payment, since workers are not expected to devote much effort to their tasks during that month and the employer cannot rely on them.

Now we turn to the contributions that the employer is obliged to pay. These are levied over not only the nominal monthly wage, but also the additional payments described above (thirteenth salary, vacation stipend and advance notice). The first item is the monthly contribution of 8% of the wage to the worker's severance payment fund (FGTS). In the following row, we state the expected balance of this fund after 33.24 months, which is the expected duration of formal employment in the model. This information is used to calculate the severance payment, which is 50% of the total FGTS balance at the time of dismissal. Note that, of the 50% payment, 40% go to the dismissed employee and the remaining 10% are appropriated by the government. In addition, there was an additional temporary contribution to the FGTS fund of 0.5%, which expired in December 2006.

The largest cost that formal employers face is the social security contribution (INSS), which accounts for 20% of the nominal wage. Finally, there are some other smaller contributions, including mandatory insurance and contributions that are specific to the activity developed by the firm. We use Souza et al. (2012) as a reference in listing those contributions.

After all contributions are taken into account, we find that formal employers pay 57.7% more than the nominal monthly wage to each worker. However, this calculation does not take into account that formal employees are entitled to paid vacations of one month per year. Thus, although the employer pays for the 12 months in the year, each employee is only productive in 11 of them. In other words, for each 11 workers that the firm wants to use in production, 12 must be hired, because 1 in every 12 is expected to be in vacation at each time. After making the corresponding adjustments, we find that the total cost for each worker that the firm wants to use in production is 72.06% of the nominal wage in October 2003.

We then proceed to the calculation of the cost of formal employment in October 2012. The only change in regulations that affected the cost paid by the employer was the phasing

Table B.1: Cost of Formal Employment in October 2003

Item	Rationale	Value
Nominal wage (A)		**100.00**
13th salary (A.1)	1/12 of A	8.33
Vacation stipend (A.2)	0.33/12 of A	2.78
Advance notice	(A+A.1+A.2) x prob. dismissal	3.34
Raw total wage (B)		**114.45**
FGTS contribution (B.1)	8% of B	9.16
FGTS balance on dismissal (B.2)	*B.1 x average duration*	*304.36*
Severance payment	50% of B.2 x prob. dismissal	4.58
FGTS temporary extra	0.5% of B	0.57
Employer INSS contribution	20% of B	22.89
SAT, INCRA, S system	5.3% of B	6.07
Total with contributions (C)		**157.72**
Vacation adjustment	1/11 of C	14.34
Total cost		172.06
Payroll tax rate (τ)		**0.7206**

out of the temporary FGTS contribution. When we exclude that contribution, we find that the equivalent payroll tax rate in October 2012 was 71.43% of the nominal wage.

Valuation of Mandated Benefits

In this subsection we account for all characteristics of formal employment that can make it more or less attractive to workers when compared with informal employment. Differently from the previous section, some of the items we consider affect low wage and high wage workers differently, such as the income tax. Thus, we have separate valuations for low wage workers and high wage workers. Low wage workers are those who earn exactly the minimum wage. The high wage worker is a representative agent for all other formal employees.

Table B.2 shows our calculations of the value attributed to benefits and contributions that calculated as fractions of the base salary. When taken together, these regulations compose the variable benefits parameters in the $benefits$ expression, a_s and a_u. The first five rows are similar to those in Table B.1: formal workers receive not only the nominal monthly wage, but also the thirteenth salary, the vacation stipend and the advance notification in case of dismissal. Two items are then deducted from the raw total wage: the social security (INSS) deduction and the income tax (IRPF). For the low wage workers, we use the lowest brackets: zero income tax in both years and social security deductions of 7.65% (in 2003) or 8.00% (in 2012). For the high wage workers, we calculate the deductions for each individual worker in the PME data set that receives more than the minimum wage, using the corresponding tax rates and brackets in each year. Then, we calculate the average deduction per worker.

The next four items are benefits that are valuable to formal workers. The first is the FGTS fund. Workers can withdraw money from their accounts in the FGTS fund, but only in a few special occasions: dismissal, retirement and when buying a house. In addition to being illiquid, resources in the fund are also less valuable than a direct payment because their returns are lower than the market interest rate. Souza et al. (2012) consider two extreme scenarios in their exercise: one in which the valuation of FGTS funds is 100% of the nominal balance, and other where workers do not value resources in the fund at all. They then report the valuation of benefits as a range. We take an intermediate route and assume that the value of deposits in the worker's FGTS account is 50% of the employer's actual disbursement.

The remaining benefits are the severance payment, the compulsory work accident insurance (SAT) and vacations. The first two items are calculated in a similar manner as in the previous subsection, when assessing the costs of formal employment. To input the valuation of vacations by workers, we use exactly the same value calculated as the cost of vacancy for employers. In this sense, vacations can be regarded as a transfer from firm to worker. Thus, if we calculate the difference between aggregate total payroll taxes and aggregate benefits, vacations and other transfers, such as the thirteenth salary, are canceled out, and we can use the result as government surplus in the model. We find that the net valuation of variable benefits is around 30% of the base salary for low wage workers, and around 23% for high wage workers.

The fixed benefits parameters (b_s^F, b_u^F) reflect a program called *abono salarial*, which is an annual stipend equal to the minimum wage paid to low wage workers (those who receive up to two times the minimum wage per month). To be eligible for this benefit, the employee must have been employed formally for at least five years (not necessarily in the same firm). We use the PME data set and estimate that 60% of formal employees who earn less than two minimum wages are entitled to the *abono salarial*. We thus find $b_u^F = 0.05$ $(0.6 \cdot 1/12)$. Only 40% of workers defined as high wage employees earn less than twice the minimum wage in the data. Thus, we set $b_s^F = 0.02$.

Finally, we calculate the unemployment insurance parameters (b_s^D, b_u^D). Unemployed workers who were previously employed formally for at least six months are entitled to unemployment benefits. Although the size of the monthly payments vary according to the wage in the last employment, there are caps on the minimum and maximum values paid. Low wage workers will always receive exactly one minimum wage, while most others will receive the maximum value of 1.87 times the minimum wage. The number of payments may vary from 3 to 5, according to the duration of all formal jobs in the last 36 months. For simplicity, we assume that the expected present value of these payments is equivalent to four times the value of each payment. Thus, $b_s^D = 4 \cdot 1.87 = 7.48$ and $b_u^D = 4$.

B.2 Informality Trends by Economic Activity

In this Appendix, we show that the decline in the informality rate in Brazil was widespread in the economy, and also that it was not caused by reallocation of workers across sectors. In the PME survey, workers report the economic activity to which their main job belongs, choosing one of 60 categories. In Table B.3, we list 15 economic activities with the largest

Table B.2: Valuation of Variable Benefits

		October 2003		October 2012	
Item	Rationale	Low wage	High wage	Low wage	High wage
Nominal wage (A)		**240.00**	**848.00**	**622.00**	**1680.47**
13th salary (A.1)	1/12 of A	20.00	70.67	51.83	140.04
Vacation stipend (A.2)	0.33/12 of A	6.67	23.56	17.28	46.68
Advance notice	(A+A.1+A.2) x prob. dismissal	8.02	28.35	20.79	56.17
Raw total wage (B)		**274.69**	**970.57**	**711.90**	**1923.36**
INSS deduction	7.65%/7.93% (03) or 8.00%/8.27% (12) of B	-21.01	-76.97	-56.95	-159.06
Income tax (IRPF) deduction	0%/5.90% (03) or 0%/5.60% (12) of B	0.00	-57.26	0.00	-107.96
Valuation of FGTS fund	50% of employer contribution	10.99	38.82	28.48	76.93
Severance payment	40% of FGTS balance x prob. dismissal	8.79	31.06	22.78	61.55
Work accident insurance (SAT)	2% of B	5.49	19.41	14.24	38.47
Total with contributions (C)		**278.95**	**925.63**	**720.45**	**1833.29**
Vacation adjustment	Equal to the cost of vacation paid by employer	34.41	121.59	88.86	240.07
Total valuation		313.36	1047.22	809.30	2073.36
Variable benefits parameter		**0.306**	**0.235**	**0.301**	**0.234**

Table B.3: Informality Trends per Economic Activity

Economic activity	Formality rate			Share of workforce			Decomposition		
	2003	2012	Change	2003	2012	Change	Within	Between	Total
Construction	55.0	73.6	18.6	7.0	8.1	1.1	1.3	0.8	2.1
Leisure, culture, sports	55.3	65.7	10.4	2.5	2.1	-0.4	0.3	-0.2	0.0
Vehicle trading and repairs; fuel retail	60.2	73.5	13.3	4.3	3.9	-0.4	0.6	-0.3	0.3
Hospitality industry, restaurants	64.3	73.8	9.5	5.3	5.2	-0.2	0.5	-0.1	0.4
Trade and repair of personal/household objects	70.3	83.2	12.8	17.7	17.3	-0.4	2.3	-0.3	1.9
Education	72.6	81.6	9.0	4.4	4.2	-0.2	0.4	-0.1	0.3
Leather industry (including shoe crafting)	73.6	84.0	10.3	2.2	1.5	-0.8	0.2	-0.7	-0.4
Other activities	74.2	82.2	8.1	23.4	21.9	-1.5	1.9	-1.2	0.7
Terrestrial transportation	76.2	85.0	8.8	5.6	5.5	-0.1	0.5	-0.1	0.4
Food industry	77.2	86.1	8.9	2.7	2.6	-0.1	0.2	-0.1	0.1
Services for businesses	77.7	87.2	9.5	9.9	13.9	4.0	0.9	3.5	4.4
Metal crafting, including machines and equipment	78.7	83.9	5.2	2.4	1.9	-0.5	0.1	-0.4	-0.3
Health and social services	79.1	86.6	7.5	5.2	5.4	0.2	0.4	0.1	0.5
Real estate	80.8	84.2	3.4	3.5	2.6	-0.9	0.1	-0.7	-0.6
Chemical industry	88.5	92.9	4.4	2.3	1.8	-0.5	0.1	-0.5	-0.4
Automotive industry	93.1	95.9	2.8	1.5	2.1	0.7	0.0	0.6	0.7
Whole workforce	72.2	82.3	10.1	100.0	100.0	0.0	9.9	0.2	10.1

Notes: Informality is defined as proportion of workers without a signed labor card. Data does not include domestic workers, public servants or self-employed workers.

number of workers. Together, they account for 76% of the workforce in 2003, and 78% in 2012. For each activity, we compute the formality rates in 2003 and 2012, and also the share of the workforce employed therein. Note that, since the PME targets workers in large metropolitan areas, few of them are employed in agricultural or extractive activities.

The first important observation is that formality increased in all economic activities listed. The share of formal workers increased more in activities that were initially more informal, but even the automotive and chemical industries experienced important gains in formalization. However, it is still possible that part of the decline was caused from workers migrating from less formal activities to others that are intrinsically more formal. To test this hypothesis, we decompose the contribution of each sector for the increase in formalization in the following way:

$$\begin{aligned} \text{Total contribution}_i &= F_{i,2012}P_{i,2012} - F_{i,2003}P_{i,2003} \\ \text{Within contribution}_i &= P_{i,2003} \cdot (F_{i,2012} - F_{i,2003}) \\ \text{Between contribution}_i &= F_{i,2012} \cdot (P_{i,2012} - P_{i,2003}) \end{aligned}$$

where $P_{i,t}$ and $F_{i,t}$ denote the share of the workforce in and the formality rate of activity i in year t, respectively. The sum of the within contributions describe what would happen if the share of workers in each activity remained constant from 2003 to 2012, but the formality rates within each activity changed. The sum of between contributions accounts for the part of the decline in informality that can be attributed to changes in the size of each activity, given the formality rates in 2012. As can be seen in the bottom row of Table B.3, the decline in informality can be accounted for almost exclusively with changes within each activity.

The facts we show in this Appendix suggest that idiosyncratic shocks are unlikely to be the cause behind the formalization of the Brazilian labor market. This is the reason why we focus on factors that influenced the whole workforce, such as educational trends, enforcement policy and labor regulation.

B.3 Solution to the Problem of the Firm

Consider problem 2.1 and denote $\frac{\partial \Pi^{z,j}(n_s,n_u)}{\partial n_i} = J_i^{z,j}(n_s,n_u)$. The optimality of controls v_s, v_u yields:

$$-\xi + q(\theta_i)J_i^{z,j}(n_s^+, n_u^+) = 0$$

Also, differentiating the value function in n_i yields:

$$(1+rdt)J_i^{z,j}(n_s,n_u) = \frac{\partial \pi^{z,j}(\cdot)}{\partial n_i}dt + (1-s^j dt)J_i^{z,j}(n_s^+, n_u^+)$$

If we differentiate $\pi^{z,j}(\cdot)$ in n_i and restrict attention to steady-state equilibria, where $n_i^+ = n_i$, the two equations above result in 2.3 and 2.2 respectively.

B.4 Solution to the Wage Bargaining Equation

Throughout this exposition, we restrict attention to the problem of the formal firm. The solution is analogous for an informal firm, once we substitute $H(z, n_s, n_u) = F(z, n_s, n_u) - \rho(n_s + n_u)$ for the production function and set $\tau_i = b_i = 0$, $a_i = 1$. Also, for simplicity, we omit the productivity index in all functions.

The Nash bargaining equation is:

$$\sigma J_i(n_s, n_u) = (1 - \sigma)\left[E_i\left(w_i(n_s, n_u)\right) - U_i\right]$$

Replacing equations 2.2 and 2.4 in the expression above, we find the following system of nonlinear differential equations:

$$c_i w_i(n_s, n_u) = (1-\sigma)(rU_i - b_i) + \sigma\left[F_i(n_s, n_u) - (1+\tau_s)n_s\frac{\partial w_s(\cdot)}{\partial n_i} - (1+\tau_u)n_u\frac{\partial w_u(\cdot)}{\partial n_i}\right] \tag{B.1}$$

where $c_i = [(1-\sigma)a_i + \sigma(1+\tau_i)]$.

The first step to solve this system is to write it in a more convenient way. Taking the partial derivative of B.1 with respect to n_u when $i = s$ yields:

$$c_s\frac{\partial w_s(\cdot)}{\partial n_u} = \sigma\left[F_{su}(n_s, n_u) - (1+\tau_s)n_s\frac{\partial^2 w_s(\cdot)}{\partial n_s \partial n_u} - (1+\tau_u)n_u\frac{\partial^2 w_u(\cdot)}{\partial n_s \partial n_u} - (1+\tau_u)\frac{\partial w_u(\cdot)}{\partial n_s}\right]$$

where $F_{su}(n_s, n_u) = \frac{\partial^2 F(n_s, n_u)}{\partial n_s \partial n_u}$. Conversely, taking the derivative with respect to n_s when $i = u$ yields:

$$c_u\frac{\partial w_u(\cdot)}{\partial n_s} = \sigma\left[F_{su}(n_s, n_u) - (1+\tau_s)n_s\frac{\partial^2 w_s(\cdot)}{\partial n_s \partial n_u} - (1+\tau_s)\frac{\partial w_s(\cdot)}{\partial n_u} - (1+\tau_u)n_u\frac{\partial^2 w_u(\cdot)}{\partial n_s \partial n_u}\right]$$

The difference between these two equations gives us the following expression:

$$\frac{\partial w_s(\cdot)}{\partial n_u}\left[c_s - \sigma(1+\tau_s)\right] = \frac{\partial w_u(\cdot)}{\partial n_s}\left[c_u - \sigma(1+\tau_u)\right]$$

Using the definition of c_i, we obtain:

$$\frac{\partial w_s(\cdot)}{\partial n_u} = \frac{a_u}{a_s}\frac{\partial w_u(\cdot)}{\partial n_s}$$

Which we can use to write the system of equations defined in B.1 as:

$$c_i w_i(n_s, n_u) = (1-\sigma)(rU_i - b_i) + \sigma\left[F_i(n_s, n_u) - (1+\tau_i)\left(\chi_{i,s}n_s\frac{\partial w_i(\cdot)}{\partial n_s} + \chi_{i,u}n_u\frac{\partial w_i(\cdot)}{\partial n_u}\right)\right] \tag{B.2}$$

where

$$\chi_{i,j} = \frac{a_i(1+\tau_j)}{a_j(1+\tau_i)}$$

Following Cahuc, Marque and Wasmer (2008) (henceforth CMW), we first solve the equation for the case in which $\chi_{i,j} = 1$. Later, we generalize the solution. The insight in CMW is to perform a change of coordinates that allows us to express the term multiplying $(1+\tau_i)$ in equation B.2 in a simpler manner, effectively obtaining a univariate differential equation as the result. The transformation we need is:

$$\begin{aligned} n_s &= \rho \cos\phi \\ n_u &= \rho \sin\phi \end{aligned}$$

Now if we let $\hat{w}_i(\rho,\phi) = w_i(\rho\cos\phi, \rho\sin\phi)$, we can find that:

$$\begin{aligned} \rho\frac{\partial \hat{w}_i(\rho,\phi)}{\partial\rho} &= \rho\left[\cos\phi\frac{\partial w_i(\cdot)}{\partial n_s} + \sin\theta\frac{\partial w_i(\cdot)}{\partial n_s}\right] \\ &= n_s\frac{\partial w_i(\cdot)}{\partial n_s} + n_u\frac{\partial w_i(\cdot)}{\partial n_u} \end{aligned}$$

Which is the term multiplying $(1+\tau_i)$ in equations B.2 if $\chi_{i,j} = 1$. Following the same notation, let $\hat{F_{n_i}}(\rho,\phi) = \frac{\partial F(\rho\cos\phi,\rho\sin\phi)}{\partial n_i}$ denote the marginal product function in the new coordinate system. We can then rewrite the differential equations as:

$$\frac{\partial \hat{w}_i(\rho,\phi)}{\partial\rho} + \frac{c_i}{\sigma(1+\tau_i)\rho}\hat{w}_i(\rho,\phi) = \frac{1-\sigma}{\sigma(1+\tau_i)\rho}(rU_i - b_i) + \frac{1}{(1+\tau_i)\rho}\hat{F_{n_i}}(\rho,\phi) \tag{B.3}$$

We guess the following form for the solution:

$$\begin{aligned} \hat{w}_i(\rho,\phi) &= C(\rho,\phi)\rho^{-\frac{c_i}{\sigma(1+\tau)}} + D(\phi) \\ \frac{\partial \hat{w}_i(\rho,\phi)}{\partial\rho} &= C'(\rho,\phi)\rho^{-\frac{c_i}{\sigma(1+\tau)}} - C(\rho,\phi)\frac{c_i}{\sigma(1+\tau)}\rho^{-\frac{c_i}{\sigma(1+\tau)}-1} \end{aligned} \tag{B.4}$$

With $C'(\cdot) = \frac{\partial C(\cdot)}{\partial\rho}$. Plugging these expressions back in differential equation, we get:

$$\begin{aligned} D \quad (\phi) = &\quad \frac{1-\sigma}{c_i}(rU_i - b_i) = D \\ C'(\rho,\phi) &= \rho^{\frac{c_i}{\sigma(1+\tau_i)}-1}\frac{1}{1+\tau_i}\hat{F_{n_i}}(\rho,\phi) = \rho^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}}\frac{1}{1+\tau_i}\hat{F_{n_i}}(\rho,\phi) \end{aligned}$$

We can integrate the latter equation to obtain:

$$C(\rho,\phi) = \frac{1}{1+\tau_i}\int_0^\rho x^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}}\hat{F_{n_i}}(x,\phi)dx + \kappa(\phi)$$

Replacing in B.4, we get:

$$\hat{w}_i(\rho,\phi) \quad = \quad \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{\rho^{-\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}-1}}{1+\tau_i}\left[\int_0^\rho x^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}}\hat{F_{n_i}}(x,\phi)dx + \kappa(\phi)\right]$$

In order to pin down the integration constant $\kappa(\phi)$, we assume that $\lim_{\rho\to 0}\rho\hat{w}_i(\rho,\phi) = 0$, in a similar manner as CMW. The assumption means that payroll goes to zero as firm size decreases while keeping the ratio of skilled to unskilled workers constant, and it is valid as long as marginal productivities do not increase too fast as the number of worker goes to zero (technically, faster than $1/\rho$ as $\rho \to 0$). This is the case for the CES-like production function we use in our quantitative exercises. Then, the equation above implies $\kappa(\phi) = 0$.

In addition, we change the integration variable to $\epsilon = x/\rho$. With that modification, we can easily change back to the rectangular coordinates by noting that $\hat{F_{n_i}}(x,\phi) = \hat{F_{n_i}}(\epsilon\rho,\phi) = F_{n_i}(\epsilon n_s, \epsilon n_u)$. The solution is given by:

$$w_i(n_s,n_u) \quad = \quad \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{1}{1+\tau_i}\int_0^1 \epsilon^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}}\frac{\partial F(\epsilon n_s,\epsilon n_u)}{\partial n_i}d\epsilon$$

Now we consider the case in which $\chi_{i,j} = \frac{a_i(1+\tau_j)}{a_j(1+\tau_i)} \neq 1$. We perform another coordinate change, introducing a new set of variables $M_i = (M_{is}, M_{iu})$, with the goal of writing:

$$\sum_{j=s,u} M_{ij}\frac{\partial \tilde{w}_j(M_i)}{\partial M_{ij}} = \sum_{j=s,u}\chi_{ij}n_j\frac{\partial w_i(n_s,n_u)}{\partial n_j}$$

with $\tilde{w}_i(M_i) = w_i(n_s,n_u)$. Denote by $\tilde{F}(M_i) = F(n_s,n_u)$ the production function in the new coordinate system. To find M_i as a function of n_s and n_u, we assume that M_{ij} only depends on n_j. In this case,

$$\frac{\partial w_i(\cdot)}{\partial n_j} = \frac{\partial \tilde{w}_i(\cdot)}{\partial M_{ij}}\frac{\partial M_{ij}}{\partial n_j}$$

Also, we further impose that

$$M_{ij}\frac{\partial \tilde{w}_i(\cdot)}{\partial M_{ij}} = \chi_{ij}n_j\frac{\partial w_i(n_s,n_u)}{\partial n_j}$$

in order to fulfill the initial requirement on the M_i variables. Combining these expressions, we find a differential equation for M_{ij}:

$$M_{ij} = \chi_{ij}n_j\frac{\partial M_{ij}}{\partial n_j}$$

We only need one solution, the simplest being

$$M_{ij} = n_j^{\frac{1}{\chi_{i,j}}} = n_j^{\chi_{j,i}}$$

since $1/\chi_{i,j} = \chi_{j,i}$. Then, using $\partial F/\partial n_j = \chi_{j,i} n_j^{\chi_{j,i}-1} \partial \tilde{F}/\partial M_{i,j}$ and $\partial F/\partial n_i = \partial \tilde{F}/\partial M_{i,i}$ as $\chi_{i,i} = 1$, the system B.2 can be rewritten as

$$c_i \tilde{w}_i(M_{is}, M_{iu}) = (1-\sigma)(rU_i - b_i) + \sigma \left[\frac{\partial \tilde{F}(M_i)}{\partial M_{ii}} - (1+\tau_i) \left(M_{is} \frac{\partial \tilde{w}_i(M_i)}{\partial M_{is}} - M_{iu} \frac{\partial \tilde{w}_i(M_i)}{\partial M_{iu}} \right) \right] \tag{B.5}$$

System B.5 is equivalent to system B.2 in the case where $\chi_{i,j} = 1$. Thus, the solution for $\tilde{w}_i(M_{is}, M_{iu})$ is known:

$$\tilde{w}_i(M_{is}, M_{iu}) \quad = \quad \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{1}{1+\tau_i} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_i}{1+\tau_i}} \tilde{F}_i(\epsilon M_{is}, \epsilon M_{iu}) d\epsilon$$

where $\tilde{F}_i$ is the derivative of function $\tilde{F}$ with respect to its argument $i = 1, ..., n$. Switching back to the original coordinate system, we obtain:

$$w_i(n_s, n_u) = \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{1}{1+\tau_i} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_i}{1+\tau_i}} \frac{\partial F\left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i} d\epsilon \tag{B.6}$$

This wage equation is easily differentiable with regard to the number of employed workers of any type:

$$\frac{\partial w_i(n_s, n_u)}{\partial n_j} = \frac{1}{1+\tau_i} \int_0^1 \epsilon^{\frac{a_i}{1+\tau_i}\left(\frac{1-\sigma}{\sigma} + \frac{1+\tau_j}{a_j} \right)} \frac{\partial^2 F\left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i \partial n_j} d\epsilon \tag{B.7}$$

To compare the solution we found to that in CMW, write $\tilde{\sigma}_i = \frac{\sigma(1+\tau_i)}{\sigma(1+\tau_i)+(1-\sigma)a_i} = \frac{\sigma(1+\tau_i)}{c_i}$. Then, equation B.6 can be stated as:

$$a_i w_i(n_s, n_u) = (1-\tilde{\sigma}_i)(rU_i - b_i) + \frac{a_i}{1+\tau_i} \int_0^1 \epsilon^{\frac{1-\tilde{\sigma}_i}{\tilde{\sigma}_i}} \frac{\partial F\left(\epsilon^{\frac{1+\tilde{\sigma}_i}{\tilde{\sigma}_i} \frac{\tilde{\sigma}_s}{1-\tilde{\sigma}_s}} n_s, \epsilon^{\frac{1+\tilde{\sigma}_i}{\tilde{\sigma}_i} \frac{\tilde{\sigma}_u}{1-\tilde{\sigma}_u}} n_u \right)}{\partial n_i} d\epsilon \tag{B.8}$$

This expression is very similar to the solution in CMW, except for the terms a_i and $a_i/(1+\tau_i)$. Consider the case where $\alpha_i = 1+\tau_i$: the valuation of formal benefits by workers is exactly equal to the total costs incurred by firms. In this case, $\tilde{\sigma}_i = \sigma$ and the only difference between our solution and that in CMW is a term a_i multiplying w_i on the left-hand side. This factor accounts for the fact that the "true" wage in this economy is $(1+\tau_i)w_i = a_i w_i$, which is both the value that firms pay and how workers value total compensation.

If $\tau_i \neq a_i - 1$, then there is a wedge between firm disbursements and the valuation of total pay by workers, and $\tilde{\sigma}_i \neq \sigma$. Note that this does not mean that the share of rents appropriated by workers is different; instead, this is an adjustment inside the integral term to compensate for the term $a_i/(1+\tau_i)$ outside the integral, keeping the Nash bargaining equation valid. However, even in the case where σ is the same for all workers, we can have $\tilde{\sigma}_i \neq \tilde{\sigma}_j$. This would lead to non-trivial interactions between different types of labor in a similar manner to how heterogeneity in bargaining power affects wages in CMW.

Finally, note that, although we have assumed the same bargaining power for all workers, it is immediate to extend it to the more general case with type-specific bargaining power. This would lead to an expression similar to B.8, but with $\tilde{\sigma}_i = \frac{\sigma_i(1+\tau_i)}{\sigma_i(1+\tau_i)+(1-\sigma_i)a_i}$. Similarly, extending the solution to more than two types of workers would be trivial, requiring essentially a change in notation. See CMW, in particular how they define the matrix $\mathbf{NA}_i(z)$.

Minimum Wages and Wage Bargaining

The solution we found above for the wage bargaining differential equation, $w_i(n_s, n_u)$, does not take into account the possibility of a minimum wage. If we set a rule that constrains wages to be no less than a constant value, then the previous solution is only correct in the interior of the subset of the (n_s, n_u) space in which the minimum wage is less than the freely bargained wage. For other values of (n_s, n_u), the minimum wage binds for the skilled, unskilled, or both.

Figure B.1 shows an example of how wages can be affected by the minimum wage according to firm size. For small values of n_s and n_u, marginal productivities are high and bargained wages are above the minimum wage. As the quantity of either type of worker increases, it is possible that marginal productivities decrease so much that the minimum wage binds. For high values of both of inputs, it is possible that all wages equal the minimum wage. In this example, the curves are upward sloping because there is complementarity between labor types ($\frac{\partial^2 F^z(n_s,n_u)}{\partial n_s \partial n_u} > 0$). They would be straight or downward sloping if that cross derivative was null or negative, respectively.

It is also possible that, for certain values of (n_s, n_u), there is multiplicity of wages satisfying the bargaining conditions: either type of worker might receive the minimum wage, but not both. This pathology is caused by discontinuities in the marginal value of workers which we discuss below. In our applications, there is no possibility that the minimum wage binds for the skilled, no matter how many workers of this type are hired. The reason is that the first term in the wage equation B.6, related to the reservation wage, is strictly greater than the minimum wage in all simulations. Hence, we are not concerned about this multiplicity problem.

If the minimum wage binds for only one type of worker, the unconstrained solution for the other type is no longer adequate. This is because, contrary to what is implied in the wage bargaining differential equation, marginal changes in the amount of the unconstrained type do not affect wages of the constrained type. From now on, for ease of exposition and focusing on our empirical application, we restrict attention to the case in which the minimum wage binds for unskilled workers, but not for skilled workers.

To find the correct skilled wage function in this case, we observe that the differential

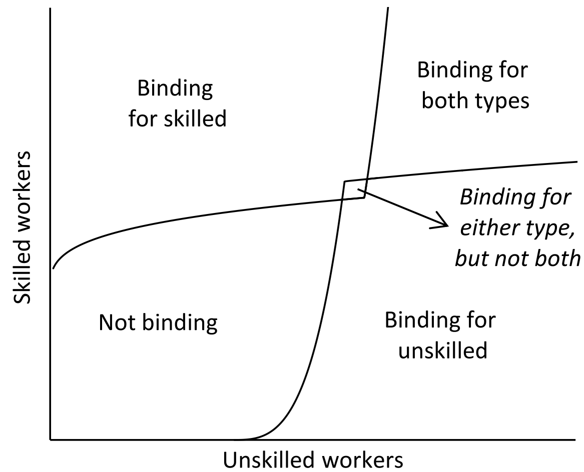


Figure B.1: Minimum Wage Status According to Firm Size

equation B.1 simplifies to:

$$c_i w_s(n_s, n_u) = (1-\sigma)(rU_i - b_i) + \sigma \left[F_s(n_s, n_u) - (1+\tau_s) n_s \frac{\partial w_s(n_s, n_u)}{\partial n_s} \right] \tag{B.9}$$

as the term $\frac{\partial w_u(n_s,n_u)}{\partial n_s}$ is set to zero. This is a univariate differential equation in n_s, similar to B.3. The solution is analogous:

$$w_s^{z,for}(n_s, n_u) = \frac{1-\sigma}{c_s}(rU_s - b_s) + \frac{1}{1+\tau_s} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_s}{1+\tau_s}} \frac{\partial F^z(\epsilon n_s, n_u)}{\partial n_i} d\epsilon$$

Note that skilled wages are still a function of the number of both skilled and unskilled workers, but not the same function as before. When the cross derivative of the production function $\frac{\partial^2 F^z(n_s,n_u)}{\partial n_s \partial n_u}$ is positive, as in our quantitative exercises, then we should expect this new wage function to be strictly greater than the unconstrained one for the same values of n_s and n_u. The reason is that, in the unconstrained case, hiring an additional skilled worker leads to an increase in unskilled wages due to the effect in the unskilled marginal productivities, which reduces the surplus being bargained over (from the point of view of the firm and the single skilled worker with whom it is bargaining). This "negative" effect does not exist (at the margin) when the minimum wage binds: the surplus is bigger, and so are bargained wages. Note that this implies a discontinuity in the wage function at the points that separate the regions where the minimum wage is or is not binding.

Minimum Wages and the Solution to the Problem of the Firm

Finally, we discuss how the existence of the minimum wage might change the problem of choosing the optimal firm size. The discontinuity in the wage function, discussed above, is caused by discrete changes in the net marginal value of workers $J_i^{for}(\cdot)$ (see equation 2.2)

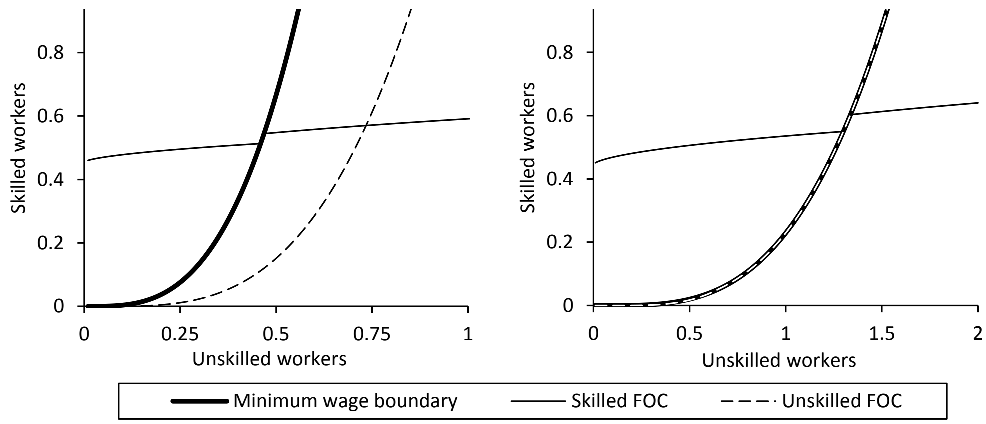


Figure B.2: Problem of the Firm and Minimum Wages

at the boundary of region of the (n_s, n_u) space where the minimum wage is binding. This discontinuity might lead to cases in which there is no exact solution to the firm's first order condition, equation 2.3. We continue to restrict attention to the case in which the minimum wage binds only for unskilled workers.

In figure B.2, we show how the minimum wage can affect the problem of the firm. In Panel A, we illustrate the problem of a formal firm with average productivity $(z = 1)$ in our baseline calibration. The heavy solid line marks the transition between a non-binding and a binding minimum wage for the unskilled workers – that is, it is the vertical line in figure B.1. The other lines are the optimality conditions for the number of skilled and unskilled workers (equation 2.3). The solid line marks the combinations of (n_s, n_u) in which the marginal value of a skilled worker, $J_s^{for}(n_s, n_u)$, is equal to the expected search cost $\frac{\xi}{q(\theta_s)}$. Above this line, there are too many skilled workers, which drives down their marginal productivity and makes the marginal value less than the search cost. The same reasoning is valid for the dashed line: to the right of it, the marginal value of unskilled workers is less than the expected search cost, and the converse is true to the left of the line. As before, the upward slope of all curves comes from complementarity between labor inputs.

The unique solution to the problem of the firm in Panel A is the point where the two first order conditions are satisfied. Since this point is to the right of the heavy solid line, the minimum wage is binding at the optimal firm size. Note that there is a discontinuity in the skilled worker's first order condition as it crosses the minimum wage boundary. Since the marginal value of skilled workers increases when the minimum wage binds for the unskilled, it becomes optimal to hire more skilled workers immediately to the right of the boundary. There is a similar discontinuity in the value of the unskilled worker, but in the opposite direction: to the right of the boundary, hiring an additional unskilled worker no longer benefits the firm by bringing down unskilled wages. However, in this case, the discrete decrease is not enough to reduce the marginal value of the unskilled to below the search cost. This is why the dashed line lies to the right of the minimum wage boundary.

Panel B describes a case in which there is no solution to the problem of the firm FOC's because of the discontinuities associated with the minimum wage. It follows from a change in the baseline model that increases overall productivity (parameter A in the quantitative experiments section), making the minimum wage binding by a smaller margin. The difference between Panel B and Panel A is that the discrete fall in the marginal value of the unskilled workers causes it to drop from a number strictly greater than the expected search costs to another strictly less than it. As a consequence, there is no point in the graph in which the unskilled first order condition is satisfied. The skilled first order condition is not satisfied either at the intersection of the three lines.

In such situation, the firm would strategically choose a point to the left of that intersection (where the minimum wage does not bind), since bargained wages for skilled workers would be discontinuously lower than immediately to the right of the intersection. There is no similar discontinuity in unskilled wage because it cannot drop below the minimum wage, and thus unskilled wages are approximately equal on both sides of the boundary. In our numerical applications, the optimal firm size in those situations is chosen by finding the point (n_s^*, n_u^*) that satisfies the first order condition for skilled workers and lies immediately to the left of the discontinuity.[1]

Note that, in the absence of the minimum wage, we would expect the firm to hire more unskilled workers, since the dashed line would lie to the right of the heavy solid line. Whether the firm would hire more or less skilled workers depend on the degree of complementarity between the two types of labor in the production function.

B.5 Numerical Procedures

In this appendix, we describe the numerical procedures required to solve for an equilibrium and to perform the minimum distance estimation presented in our quantitative exercises. The descriptions below include general overviews of the procedures as well as specific operational details. Figure B.3 provides a sketch of how each step of the numerical implementation of the paper relate to each other. In the first subsection of this Appendix, we describe the procedure to solve for equilibrium, comprising the three "lower" levels of the hierarchy described in Figure B.3. Next, we discuss the estimation procedure.

Solving for the Equilibrium

Solving the model numerically is equivalent to finding values for θ_s, θ_u, U_s and U_u that solve equations 2.7 and 2.8 up to a desired numerical precision – that is, that set a residual

[1] In practical terms, our algorithm first tries to solve the problem of the firm using a derivative-based method. If it cannot find the solution, it solves the system given by the skilled FOC and the equation $w_u^{for}(n_s, n_u) = \bar{w}$ (the solid line in figure B.2). After solving this system, the algorithm checks if the solution is such small deviations in the number of unskilled workers make $J_u^{for}(n_s, n_u) - \frac{\xi}{q(\theta_u)}$ change sign. If not, n_u is increased or decreased, depending on the sign of $J_u^{for}(n_s, n_u) - \frac{(r+s^{for})\xi}{q(\theta_u)}$, until the condition just described is satisfied. The optimal solution is the smallest value of n_u (in a finite grid with intervals given by the numerical tolerance) such that $J_u^{for}(n_s, n_u) - \frac{(r+s^{for})\xi}{q(\theta_u)} > 0$, as a strategic firm would choose.

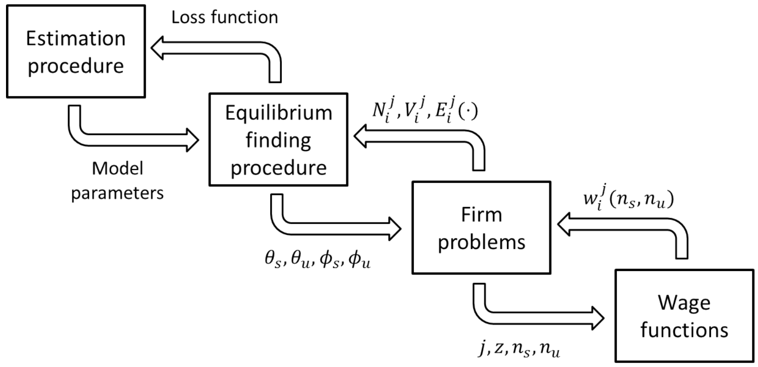


Figure B.3: Hierarchy for the numerical procedures.

term implied by the given equations to less than a tolerance value. Two observations make this computation easier. First, from 2.8, one can see that U_i is a function of only θ_i and parameters of the model when the minimum wage does not bind for workers of type i. This suggests a procedure to minimize the dimension of the problem: guess whether the minimum wage binds for each type of worker, and calculate U_i directly from θ_i if the minimum wage is not binding for type i, instead of using it as a choice variable.

In the estimation procedure described in the next subsection, since the target economy is one where the minimum wage binds only for unskilled workers, we impose this restriction in the equilibrium procedure to reduce computational time.[2] This approach is not problematic if there is a neighborhood of the parameter space around the optimal point where the minimum wage is always binding for the unskilled only. However, when running the counterfactual exercises, we allow for any combination of minimum wage status. To do so, we sequentially solve the model given one of potentially four assumptions about minimum wages, until a solution such the assumption holds is found. Checking if the assumption holds is simply a matter of checking whether freely bargained wages are above or below the minimum wage.

The second observation that helps with the computation of an equilibrium is that, when the minimum wage binds for workers i, using ϕ_i as a choice variable is easier than choosing U_i directly. This is because ϕ_i is a dimensionless ratio, bound by 0 and 1. The corresponding value of U_i is obtained from ϕ_i and θ_i, using equation 2.8.. So, if the minimum wage is guessed to bind only for unskilled workers, for instance, then the problem of finding an equilibrium is to choose θ_s, θ_u and ϕ_u that set three residual terms to zero.

To calculate the residual terms associated with a given choice of θ_i and ϕ_i, we solve the problem of all firms, aggregate all employment and vacancy decisions, and then calculate the relative differences between the choice variables and the corresponding values implied by the aggregates.[3] In the model, the distribution of firm productivities, $G(z)$, is continuous, but

[2] As discussed later in that section, it also makes the loss function less prone to discontinuities.

[3] We use relative measures of the difference between the LHS and the RHS of equations 2.7 and 2.8 as the residuals to be set to zero in the optimization procedure. The LHS is the value of θ_i or rU_i (implied by ϕ_i) that is the "guess" taken as given when solving the firm problem, and the RHS is calculated using aggregates obtained after solving the problem of the firm for all firms and interpolating the results. The procedure is not substantially affected by the specific functional form of the residual (e.g. log(RHS/LHS) or RHS/LHS-1).

since a closed form solution is not available we need to discretize it in the numerical procedure. To reduce computational time while still maintaining "smoothness" down to the desired tolerance levels, we use a interpolation procedure. Namely, we solve the model for a relatively small number of firm types (20) and then interpolate quantities like labor demand and profits over a much finer grid of productivity levels (100,000) using cubic splines. The quality of the approximation can be tested by solving the problem for in-between levels of productivity and comparing the solution to the interpolated value. We found the interpolation to be very reliable, which is expected given the smoothness of $G(z)$ and the continuity of the production function on z.[4] [5]

Given that $G(z)$ is a Generalized Pareto distribution in our quantitative exercises, we need to truncate it at the top in order for the interpolation procedure to work. We choose 30 as the upper bound (for reference, remember that the distribution is constrained to have mean 1). At the estimated value for the distribution in the quantitative exercises, the mass above that threshold for the non-truncated distribution is less than 0.001%. The results are not sensitive to changes in the threshold (conditional on re-estimating the model if the changes are relatively large).[6]

When solving the problem of an individual firm type, we use a standard optimization procedure to solve the first order conditions (or for the strategic solution described at the end of Appendix D) taking the chosen values of θ_i and U_i as given. This involves using a numerical integration procedure for the integral terms in the expressions for wages and their

Given that θ_i is a ratio and must be greater than zero, we use $z_i = \log(\theta_i)$ as the choice variable in the optimization procedure instead of θ_i itself, thus eliminating the need for constrained optimization.

[4]The model provides a direct test for the quality of the interpolation. First, calculate U_i using individual wages for all 100,000 interpolated firms and the first expression in 2.8 . Then, compare this value to the results found assuming that the FOC holds for all firms (the other expressions). To the extent that the interpolated values do not necessarily solve the FOC, there might be a discrepancy between these two ways to calculate U_i. In our baseline calibration, the relative difference is at most 0.14%.

[5]Even though the discrete nature of the firm distribution makes the problem non-smooth, derivative-based methods usually work well if the initial guess is close enough to the solution. It is important, though, to use a relatively large change in the choice variables when calculating the numerical derivatives, compared to the number of atoms in the firm productivity distribution. If the change in parameters is too small, then it's unlikely that any of the "marginal atoms" of the distribution will shift its compliance decision, even when the choice variable is relevant for that decision. In this case, the effect of changing the choice variables might be biased by not taking into account extensive margin effects. See the "A note on tolerance levels" subsection below.

[6]Additional details on handling the productivity distribution:

We first obtain a vector of 100,001 values uniformly distributed from 0 through 30. This leads to 100,000 intervals whose bounds are the elements in that vector. Given the shape parameter for $G(z)$, we can calculate the CDF at each of the 100,001 points, and thus the probability mass associated with each of the 100,000 intervals (remembering to normalize so that the probabilities add up to 1). Finally, we calculate the mean of the continuous distribution $G(z)$ conditional on lying within each interval. We use these conditional means as the value of z associated with the interval, for the sake of increased precision (instead of, for instance, using midpoints). Thus, while the bounds of the intervals remain fixed, both the probability mass function and the values of z used by the model change as the shape parameter of the distribution changes. Of course, given the large number of intervals, the changes in the z vector tend to be minor.

derivatives. We used a trapezoidal rule with 1,000 trapezoids in a uniform grid. The firm's compliance decision is determined by comparing formal and informal profits for each level of productivity z, after the interpolation is done.

Finally, by integrating vacancies and employment along the discrete distribution of productivity, we can calculate what are the implied tightnesses, θ_i, and reservation wage, rU_i, using equations 2.7 and 2.8. Note that this computation is not possible if the initial guess for θ_i is too low, since it can lead to levels of employment greater than the measure of the workforce. In this case, a larger initial value for θ_i should be provided. After we find the θ_i and ϕ_i that solve the equilibrium equations, we can verify whether the initial guess for which minimum wages bind is correct. If so, an equilibrium has been found. If not, a different guess must be tried.[7]

Estimation Procedure

The numerical procedure implemented in the previous subsection can be seen as a function mapping from the space of parameters to the space of moments and quantities implied by the model. In the estimation procedure, we focus on 7 parameters, listed in Table 2.5, and 8 moments, listed in Table 2.6. Let us denote these parameters as a vector $x \in X \subset \mathbb{R}^7$, the moments calculated from the model as a function $h : X \to \mathbb{R}^8$, and the value of these moments in the population of interest (in terms of logarithms) as π. Under the assumption that there is a unique vector of parameters x_0 such that minimizes $[\pi - h(x)]' W [\pi - h(x)]$, we can obtain an estimate of x_0 by solving the following minimization problem:

$$\hat{x} = \underset{x \in X}{\operatorname{argmax}} \, [\hat{\pi} - h(x)]' W [\hat{\pi} - h(x)] \tag{B.10}$$

where $\hat{\pi}$ is an estimate for π and W is a symmetric weighting matrix.

Under some assumptions that include consistency of $\hat{\pi}$, differentiability of $h(\cdot)$, uniform convergence of the minimand, and that x_0 is the unique solution to $E\{H(x)'W[\pi - h(x)]\} = 0$ (where $H(x) = \nabla h(x)$ is the Jacobian matrix of h at x), $\hat{x}$ converges to x_0 as the sample from which $\hat{\pi}$ is calculated increases in size. Further, the asymptotic variance of $\hat{x}$ is given by:

$$AVAR\left[\sqrt{N}(\hat{x} - x_0)\right] = \left[H(x_0)'WH(x_0)\right]^{-1} H(x_0)'WVWH(x_0) \left[H(x_0)'WH(x_0)\right]^{-1} \tag{B.11}$$

where V is the covariance matrix of the estimates $\hat{\pi}$. This matrix can be estimated by replacing $H(\hat{x})$ for $H(x_0)$ and $\hat{V}$ for V, where $\hat{V}$ is a consistent estimate for V.

[7] Additional details on solving the equilibrium set of equations:

Before starting the derivative-based method, we use a simple heuristics to approximate the solution given the size of the residuals, increasing or decreasing θ_s or θ_u if there is excess demand or supply for that kind of workers, respectively. After the residuals are relatively small, the derivative-based method is called. It is possible that the discreteness of the productivity distribution implies non-existence of an equilibrium for a given tolerance level. The choice of the granularity of the discrete productivity distribution must take this problem into account.

The assumption that $h(x)$ is differentiable does not hold strictly, given that the model's equilibrium is solved numerically by discretizing the distribution of firms. In particular, as marginal firms change discretely into and out of informality, the model outcomes also change discretely. However, given the fine granularity of the firm's distribution (100,000 atoms), we expect our numerical implementation to be a very good approximation of the continuous case. In the standard error estimations, we do not explicitly account for this additional source of imprecision. However, we do verify that results are not sensitive to choosing a larger number of atoms in that distribution.

Another potential source of non-smoothness are transitions into and out of different minimum wage regimes. As explained in the model description, changes in minimum wage can lead to discrete changes in other wages and in employment decisions. We avoid this problem by focusing in the case in which the minimum wage binds for unskilled workers, and disregarding all other cases in the estimation procedure. We also make sure that the initial points satisfy this constraint.

Given this econometric framework, we must complete five tasks in order to estimate x_0: choosing the functional form of the moments, using the data to obtain $\hat{\pi}$ and $\hat{V}$, choosing the weighting matrix, solving the optimization procedure, and calculating the covariance matrix. Below we lay some additional details on each step.

1) Choosing Moments and Functional Forms

The rationale for the specific choice of the 8 target moments is explained in the main text. However, instead of targeting them directly, we define π as the log of these moments. This choice avoids problems related to the scaling of the moments: we focus on relative gaps between the model outputs and the targets, rather than on simple differences.

2) Obtaining $\hat{\pi}$ and $\hat{V}$

The first six of the eight moments listed in Table 2.6 are calculated directly from the PME survey dataset. To obtain $\hat{\pi}$, we calculate the weighted mean of the desired moments in the sample (with the weights given by survey sampling weights), and then take the log of these means. To obtain the covariance matrix of the moments (in levels), we first estimate each variance or covariance without using weights. Then, we multiply these variances and covariances by a factor $K = n\frac{\sum_i w_i^2}{\left(\sum_i w_i\right)^2}$, where w_i is an individual weight, to account for the weighting. This factor needs to be calculated separately for each specific variance or covariance, taking into account only the relevant sample for those variables (e.g. only informal workers when assessing informal wage).

It is also important to note that, given the panel structure of the PME, the observations cannot be assumed to be independent. We take the most conservative approach possible and bundle all observations for the same worker as one, taking a weighted mean of outcomes within individual and adding up the weights (separately for each statistic). This procedure ensures that, while the aggregate means match what a researcher would find by using the sampling weights and pooled data, each worker counts as only one observation for the purposes of calculating the covariance matrix $\hat{V}$.

After the matrix of covariances for the moments is calculated, we pre- and post-multiply it by a diagonal matrix where terms in the diagonal are the inverse of the corresponding $\hat{\pi}$ term. This is just a delta-method adjustment to obtain the covariance matrix for the moments in logs. The result is the 6×6 top-left component of the matrix $\hat{V}$.

The seventh moment is the relative wage difference between the minimum wage and informal wages for unskilled workers, which we proxy by a quantile regression result from Bargain and Kwenda (2011). Specifically, we take the result from Table 3, for informal salaried, in specification 4: Panel, fixed effects quantile regression. The log wage penalty is -0.078 and its standard error is 0.004. We consider this estimate to be uncorrelated with the other moments, so that the non-diagonal terms of the seventh row and column are zero.

The eight moment is the labor share of income. The National Accounts System does not provide standard errors for this estimate. We circumvent this problem by using time-series variation in this number. More precisely, we assume that the labor share changes smoothly over time, and deviations from that smooth pattern should reflect sampling and aggregation errors within the calculation procedure. Then, we fit the labor share numbers from 1995 through 2008 in a polynomial in time. Specifically, we define the time variable as $t = year - 2003$, regress the labor shares in a polynomial in t, and use the standard error of the intercept as the standard error of the labor share estimate. The time series presents a clear convex pattern, and both the quadratic and cubic polynomials provide a tight fit. The standard errors are similar under both specifications, so we choose the largest of them (associated with the quadratic specification). Finally, we also consider non-diagonal terms in the eight row of the covariance matrix to be zero.

3) Choosing the Weighting Matrix

The results shown in the paper use the identity matrix as the weighting matrix W. We consider this to be a conservative and intuitive approach, while also sidestepping the need to calculate an additional component to the covariance if the model is mispecified (see Chamberlain (1994)). To verify the sensitivity to this choice, we re-ran the optimization procedure starting from the estimated results but using the optimal weighting matrix $\hat{V}^{-1}$. The resulting estimates were very similar to the ones shown in the paper, with no parameter changing by more than 5%.

4) Solving the Minimization Problem

Once the $h(x)$ function is defined and both $\hat{\pi}$ and W are available, estimating $\hat{x}$ requires solving the problem defined in B.10 up to the desired precision. We transform the input parameters to circumvent the need for constrained optimization. We use a logarithmic transformation for parameters that should be positive but unbounded, and a logit transformation for parameters that must lie in the $(0, 1)$ interval.

A different set of starting points should be used, since it cannot be guaranteed that any local minimum is the global minimum. We automate this process by using the genetic algorithm optimization tool from Matlab(R), with a randomly drawn initial population of 50 points. The initial points are drawn uniformly over a specific range of parameters values,[8] designed to contrain the initial points to have (i) a binding minimum wage for unskilled workers, (ii) non-binding minimum wage for unskilled workers, and (iii) both informal and informal firms in equilibrium. After finding the equilibrium for the randomly drawn points, we discard those that do not satisfy any of these three criteria. Even though the initial population is constrained to these bounds, the genetic algorithm can "escape" it through mutations. We use the best point after 10 generations as the starting point (after noting that all points with low values for the loss function appear to be near, suggesting a global minimum).

After the initial point was defined, we used a standard nonlinear minimization procedure (Matlab(R)'s fmincon) to estimate $\hat{x}$.

It is worth noting that the layered structure of the equilibrium calculation makes it computationally demanding. Each evaluation of $h(x)$ requires numerically solving a set of three equations. In turn, each evaluation of this set of equations requires solving the problem of the firm for 20 different types, leading to 20 separate optimization problems. Each firm problem is itself the numerical solution to a system of equations whose computation include solving a number of numerical integrations. All in all, each evaluation of $h(x)$ may easily take a few minutes on a relatively fast computer.[9] The whole estimation procedure can take days, even when using parallel computing in the outer optimization problem.

5) Obtaining the Variance of the Estimates

Once $\hat{x}$ is available, one can use the numerical implementation of function $h(x)$ to obtain the numerical Jacobian matrix $H(\hat{x})$. Then, it is a simple matter to compute the covariance matrix of the estimated parameters using equation B.11. The finite differences must be taken regarding the original parameter, not the transformed variables from the optimization

[8]The specific range for the transformed parameters, along with the corresponding values for the actual parameter values, is as follows:

Parameters	Range			
	Transformed		Not transformed	
A (productivity)	2	2.75	7.39	15.6426
B (technology bias)	0.3	0.9	0.5744	0.7109
α (dec. returns)	-0.5	0.5	0.3775	0.6225
γ (CES param.)	-1.2	-0.5	0.2315	0.3775
C (informality cost)	-2.75	-2	0.0639	0.1353
ξ (search cost)	-0.5	0.5	0.6065	1.6487
T (firm dist. shape)	-2.2	-1.4	0.0998	0.1978

[9]For these tests, we used a computer with an Intel(R) Core(TM) i7-4710HQ processor, large enough RAM, not running in parallel, and using standard optimization procedures in Matlab(R).

procedure (alternatively, one can estimate the covariance matrix for the transformed variables and obtain the desired covariance matrix using the delta method).

A Note on Tolerance Levels

Given the layered structure of the procedure, it is important to have a hierarchy of tolerances in the nested optimization problems. The "inner" procedures must use stricter tolerance values than the "outer" ones. Otherwise, numerical approximation errors in the former will lead to systematic errors in the latter. In a similar note, for procedures that use finite differences to calculate numerical derivatives, the finite difference must be of a substantially higher order than the tolerance of "inner" procedures.

In our application, the Classical Minimum Distance minimization problem has a tolerance of 1×10^{-5}, the equilibrium finding procedure has a tolerance of 1×10^{-8}, and the problem of the firm has a tolerance of 1×10^{-13}. The minimum size of the finite difference is set to be equal to the tolerance level of the problem.

B.6 Some Preliminary Evidence on Educational Composition and Labor Market Outcomes

This appendix provides some tentative empirical evidence on the relationship between the educational composition of the population and labor market equilibrium outcomes. Since we could not find any empirical study focusing on this relationship and providing this type of evidence, we thought it would be useful to generate some preliminary results in this direction.

We use data from the 1991, 2000, and 2010 Brazilian censuses and consider micro-regions as the relevant definition of local labor markets. Micro-regions are sets of contiguous municipalities sharing similar geographic and socioeconomic conditions defined by the Brazilian Census Bureau (IBGE). This geographic unit has been repeatedly used in the previous literature as the relevant definition of local labor markets in Brazil (see, for example, Kovak, 2013). In order to minimize heterogeneity, we focus on a sample of men between ages 20 and 50, not in school, and living in urban areas.

Our goal is to analyze the relationship between educational composition and labor market equilibrium outcomes at the level of local labor markets. Therefore, the independent variable of interest is always the share of individuals in the micro-region with at least 8 years of schooling. The dependent variables are micro-region formality or employment rates netted out of compositional effects. Specifically, the dependent variables are micro-region fixed-effects in individual level regressions, run separately for each year, where the dependent variable is either formality status (among salaried workers) or an indicator of employed (among the entire sample). The individual level regressions control for a quartic polynomial on age, dummies for race, and, in some specifications (as indicate in the table), dummies for educational levels.

The micro-region regressions include as demographic controls the shares of the sample in two age categories (30-39 and 40-50) and the log of population (all calculated based on the sample used in the individual-level regressions explained above). To allow for differential trends across local labor markets with different initial conditions, we also control for an interaction of the initial (1991) formality rate (from the individual-level regressions that do not control for schooling) with year dummies. In some specifications, we also control for the shares of employment in 8 broadly defined sectors (agribusiness and extractive industries, excluding mining; mining; manufacture; construction; utilities; retail; services; and government), and for interactions of year fixed-effects with a set of initial (1991) socioeconomic characteristics (schooling, which is the independent variable of interest; average earnings; and employment, which is one of the dependent variables considered). All regressions include micro-region and year fixed-effects and are weighted by the inverse of the standard error of the dependent variable (obtained from the individual level regressions). Standard errors are clustered at the micro-region level.

The results from these regressions are presented in Table B.4. In the table, each coefficient corresponds to a different regression, with the rows indicating different specifications and dependent variables (not controlling and controlling for education in the first stage, and using informality and employment as dependent variables). The columns correspond to different sets of controls, as indicated at the bottom of the table. All coefficients in the table refer to the same independent variable: the fraction of the population in the micro-region with at least 8 years of schooling. We present results not controlling and controlling for education in our "first-stage" to distinguish the individual level association between schooling and formality from the equilibrium effect of the composition of the population on the incidence of formality, conditional on individual schooling.

The first two rows show that there is a robust correlation between the share of the population with at least 8 years of schooling and the formality rate in the data. As expected, the coefficients are reduced in magnitude as we include micro-region fixed effects and move from column 1 to 2, but remain roughly stable across the various specifications between columns 2 and 7. So the correlation between the fraction of skilled individuals and formality is not related to differential trends across states or across micro-regions with different initial characteristics, nor to overall patterns of development and growth (as reflected on demographic patterns, average earnings, or sectoral composition of employment).

The estimates in the first row do not control for individual schooling when calculating the conditional informality rate in the "first-stage." They therefore capture both the individual relationship between schooling and formality and the potential aggregate effect of the composition of the population on individual level formality probabilities (through equilibrium labor market outcomes). The second row, in turn, controls for schooling in the "first-stage," so its results reflect the equilibrium response to changes in the educational composition of the population, conditional on individual level schooling. The fact that the results from the second row are consistently significant indicates that the aggregate effects of the composition of the labor force on labor market equilibrium outcomes are indeed relevant. The relative magnitude of the coefficients across the two rows would suggest that more than 60% of the

Table B.4: Effect of share of population with at least eight years of schooling on formality and employment.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.: Formality							
No control for indiv. school.	1.151***	0.306***	0.407***	0.283***	0.367***	0.389***	0.356***
	(0.0603)	(0.0938)	(0.0890)	(0.0928)	(0.0754)	(0.0882)	(0.0903)
Control for indiv. school.	0.955***	0.203**	0.312***	0.184**	0.285***	0.294***	0.222**
	(0.0592)	(0.0947)	(0.0895)	(0.0916)	(0.0751)	(0.0889)	(0.0913)
Dep. Var.: Employment							
No control for indiv. school.	0.280***	0.0983*	0.0381	-0.0254	0.141***	0.0254	0.0585
	(0.0269)	(0.0530)	(0.0455)	(0.0573)	(0.0485)	(0.0455)	(0.0455)
Control. for indiv. school.	0.149***	0.0268	-0.0247	-0.0791	0.0759*	-0.0367	-0.0198
	(0.0247)	(0.0496)	(0.0430)	(0.0540)	(0.0453)	(0.0430)	(0.0430)
Fixed effects:							
Micro-region	No	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	No	Yes	Yes	Yes
State-Year	No	No	No	Yes	No	No	No
Controls:							
Demographic	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1991 Form.×Year	No	No	Yes	Yes	Yes	Yes	Yes
Sectoral Shares	No	No	No	No	Yes	No	No
Avg. Earnings	No	No	No	No	No	Yes	No
1991 Other×Year	No	No	No	No	No	No	Yes

Obs.: Standard errors (in parentheses) clustered at the micro-region level; *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Data from the Brazilian censuses (1991, 2000, and 2010). Sample composed of males between ages 20 and 50, not in school, living in urban areas. Each number is the coefficient on the the share of individuals with at least 8 years of schooling from a different micro-region level regression (509 micro-regions, 1,527 observations). Dependent variables are micro-region formality and employment rates, netted out of compositional effects (micro-region fixed effects from individual level regressions, run separately for each year, where the dependent variable is either an indicator of formality or employment, and independent variables are a quartic polynomial on age, dummies for race, and dummies for educational category, as indicated in the table). Demographic controls are the shares of the population in two age categories (30-39 and 40-50) and the log of population (both calculated with the sample used in the individual-level). 1991 Formality×Year is the 1991 formality dependent variable (taken from the first-stage regression without individual schooling) interacted with year dummies. Sectoral shares are shares of the employed population in each of 8 broadly defined sectors (agribusiness and extractive industries, ex-mining; mining; manufacture; construction; utilities; retail; services; and government). 1991 Other×Year include interactions of year fixed effects with 1991 levels of three other variables: the independent variable (schooling), average earnings, and the employment dependent variable. Regressions are weighted by the inverse of the standard error of the dependent variable (obtained from the individual level regressions).

aggregate correlation between educational composition of the population and informality may be due to these equilibrium effects, while less than 40% would be due to the direct relationship between schooling and informality at the individual level.

The magnitude of the estimated effects are between 30% and 40% of the quantitative effects from the comparative statics exercise in column 6 of Table 2.8. This difference may be due to the lack of a truly exogenous source of identification in our empirical results from this section, to limitations in our definition of skilled workers in this empirical setting, or to different sample (most importantly, the exercise here uses all micro-regions in Brazil, while the calibration was conducted using data from the PME, which includes only the 6 main metropolitan areas in the country).[10]

The results related to employment, shown in the 3^{rd} and 4^{th} rows, are much less robust. Some specifications point to a positive and statistically significant relationship between educational composition and employment, but most results are small in magnitude and not statistically significant. Overall, we do not find a systematic relationship between educational composition of the population and employment rates. One potential explanation is that the utility from unemployment may be different across skilled and unskilled workers – possibly higher for skilled workers, due to higher wealth and savings –, something not considered in the model. This might weaken the correlation between educational composition and employment in the data. In addition, the problems alluded to in the previous paragraph could also be interfering with these employment results.

[10] As an illustration of the "measurement error" issue related to the mapping of skill in the model to the data, consider the following example. If 20% of individuals with less than 8 years of schooling correspond to the skilled in the model instead of unskilled, and 10% of those with at least 8 years of schooling are unskilled, there would be no change in the baseline shared of skilled workers in the workforce (leading to the same calibration). However, the change in the share of skilled workers implied by the increase in schooling would be 70% of what was used in the quantitative exercises, leading to a roughly similar reduction in the effects on informality and unemployment.