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

Daniel Haanwinckel†
and
Rodrigo R. Soares‡

October 2020

Abstract

We develop a search model of informal labor markets with realistic labor regulations, including minimum wage, and heterogeneous workers and firms. Smaller firms and lower wages in the informal sector emerge endogenously as firms and workers decide whether to comply with regulations. Because skilled and unskilled workers are imperfect substitutes in production, the model uniquely captures the informality consequences of shocks that affect returns to skill, such as rising educational levels. The model also reproduces empirical patterns incompatible with other frameworks: the presence of skilled and unskilled workers in the formal and informal sectors, the rising share of skilled workers by firm size, and formal and firm-size wage premiums that vary by skill level. We estimate the model using 2003 data from Brazil and show that it successfully predicts labor market changes observed between 2003 and 2012. Under a range of different assumptions, changes in workforce composition appear as the main drivers of the reduction in informality over this period. Policy simulations using the estimated model suggest that progressive payroll taxes are a cost-effective way to reduce informality.

Keywords: informality, labor market, search, minimum wage, compensating differentials, Brazil

JEL Codes: J24, J31, J46, J64, O17

*Previous versions circulated under the title “A Compensating Differentials Theory of Informal Labor Markets: Quantitative Model and Implications for a Developing Country.” This paper benefited from comments from Jérôme Adda (its editor), David Card, Pedro Carneiro, Aureo de Paula, Gustavo Gonzaga, Patrick Kline, Nicholas Li, Marc Muendler, Renata Narita, Gabriel Ulyssea, Thaís Vilela, Eduardo Zilberman, three anonymous referees, and seminar participants at Berkeley, IPEA-Rio, PUC-Rio, UCB, the 8th IZA/World Bank Conference on Employment and Development, the 2014 PACDEV Conference, the 2nd LACEA Labor Network Meeting, the 1st Bay Area Labor and Public Finance Graduate Conference, the 37th SBE Annual Meeting, and the 20th and 24th LACEA Annual Meetings. We are thankful to Data Zoom, developed by the Department of Economics at PUC-Rio, for providing the codes for accessing IBGE microdata. Haanwinckel gratefully acknowledges financial support from FAPERJ during the early stages of this project.

†University of California, Los Angeles, haanwinckel@econ.ucla.edu.
‡Insper and Columbia University, r.soares@insper.edu.br.
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 policy challenges are exacerbated in developing countries, where the governments’ enforcement abilities are limited and large fractions of the workforce are informal (often 30% or more). Specific programs and institutional efforts targeted at reducing labor informality have typically met with limited success (Perry et al., 2007).

Surprisingly, this pattern of persistently high informality was sharply reversed in most of Latin America in the early 2000s. In half a dozen countries, informality rates among salaried workers were reduced by one-fifth or more over a period of roughly 10 years (Tornarolli et al., 2012). In Brazil, the focus of our quantitative exercises, informality rates among salaried workers fell by more than a third, from 28.1% in 2003 to 17.3% in 2012. The Brazilian case is particularly puzzling because, during the same period, the minimum wage increased by 61% in real terms—at least twice the growth rate of GDP per capita—and changes in other labor regulations were negligible. Traditionally, high minimum wages and distortionary labor market regulations are seen as main drivers of increases in informality.

Brazil also experienced other relevant economic transformations, such as substantial increases in average schooling and productivity, during this period. 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. These could have counteracted the impact of the increases in the minimum wage, possibly providing an answer to the puzzle. But modern informality literature lacks an adequate theoretical framework with which to analyze this possibility, given its reliance on traditional search models that assume either a one-to-one match between workers and firms or, alternatively, constant marginal productivity of labor. These models immediately rule out complementarities across different types of labor and, therefore, equilibrium responses to changes in the relative supply of and demand for 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 and imperfect substitutability across different types of labor within the firm, a realistic set of labor regulations (including a minimum wage), and explicit compliance decisions by workers and firms. We estimate the model using data from Brazil and show that it reproduces various margins of change in labor market outcomes observed during the 2000s. Our quantitative exercises also highlight that the interaction between skill composition of the labor force and productivity can have first-order implications for the evolution of informality. 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 be impossible either under the frameworks commonly used in previous literature or 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 on the intra-firm bargaining theory proposed by Cahuc, Marque and Wasmer (2008), who build on Stole and Zwiebel (1996a), and extend it in four directions. First, we consider firms with different productivity levels, as opposed to a single representative firm. Second, we characterize an equilibrium where labor can move between the formal and informal sectors. Third, we incorporate a more realistic set of labor regulations, including a minimum wage, which adds a non-trivial degree of complexity to the characterization of the solution. And fourth, 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 wage in the bargaining problem (the minimum wage is modeled as an exogenous price for low-skill workers), and
we include convex vacancy-posting costs, as in Acemoglu and Hawkins (2014), which generate empirically relevant wage differentials across firms.

In the model, workers can be either skilled or unskilled and simultaneously search for formal and informal jobs when unemployed. Firms are heterogeneous in a skill-biased productivity parameter so that more productive firms are also more skill-intensive. Firms decide first whether to comply with labor regulations and then, at each moment, how many skilled and unskilled vacancies to post. By not complying with regulations, firms avoid payroll taxes and are not subject to the minimum wage. Instead, they face an informality penalty that increases with firm size (representing the probability of being audited and the associated fine) and is heterogeneous across firms (representing differences in enforcement intensity by location or sector of economic activity). Labor regulations also include mandated benefits, which, from the perspective of employees, make formal jobs more valuable than informal jobs at a given wage. Finally, wages are set by intra-firm bargaining under nonbinding contracts.

In equilibrium, firms and workers self-select into the formal and informal sectors. Complying with labor regulations requires spending a significant amount of resources. But noncompliance is also costly, especially for large firms, due to the probability of being detected and the associated fine. Workers would rather receive employment benefits, but they may be willing to accept informal jobs 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. Thus, contrary to the classic labor market duality hypothesis (see Cain, 1976), there is no sense in which firms and workers allocated to different sectors are intrinsically different.

Average cross-sector differences in productivity, firm size, and wages result from a combination of four elements: self-selection, compensating differentials, minimum wages, and firm-size wage premiums. In a steady-state equilibrium, firms with lower productivity employ fewer workers and are more likely to operate informally. These firms also employ a lower fraction of skilled workers, which is the main reason why wages are lower in the informal sector. In a particular case of the model, with linear vacancy-posting costs and no minimum wage, informal wages are higher than formal wages conditional on skill, fully compensating workers for the lack of mandated benefits and shorter employment spells. In the general case, however, workers may strictly prefer formal jobs for two reasons. First, the minimum wage elevates formal wages above the point of full compensation for unskilled workers. Second, convex vacancy-posting costs generate firm-size wage premiums that make jobs in larger firms preferable.

We estimate the model using data from the Brazilian labor market in 2003. Our estimation uses a minimum distance procedure where most targets are taken from the Brazilian Monthly Employment Survey (PME, after Pesquisa Mensal de Emprego). The other targets come from aggregate sources and from a quasi-experimental estimate of the effect of enforcement of regulation on informality rates (following Almeida and Carneiro, 2012). We treat each of the six metropolitan regions covered by the PME as an independent labor market, allowing some parameters associated with local institutional capability and productivity to vary across regions. Regional heterogeneity enriches the estimation procedure and allows for stronger validation tests of the theory.

The model quantitatively reproduces a wealth of facts from the cross-sectional distribution of workers across firms and compliance statuses. In particular, it generates an equilibrium where skilled and unskilled workers are employed in the

---

does not have different sectors nor firm-size wage premiums.
formal and informal sectors, larger firms tend to be formal, the share of skilled workers is higher in the formal sector, the minimum wage binds for unskilled workers only, the formal wage premium falls with skill, and the firm-size wage premium rises with skill.

We then look at the model’s predictions for the changes in informality between 2003 and 2012 as an out-of-sample validation strategy. In this exercise, we plug into the estimated model the observed changes in several parameters—tax rates, mandated benefits, enforcement of labor regulations, minimum wages, workforce composition, and aggregate productivity/demand—and calculate the equilibrium response of informality. We show that the model is able to reproduce the qualitative pattern of labor market changes observed in Brazil along various margins, including informality, unemployment, formal and firm-size wage premiums, and most of the distribution of workers across firm sizes. Quantitatively, it accounts for 57% of the decline in informality and 58% of the reduction in unemployment observed during the 2003–2012 period. The model also performs well in predicting the pattern of changes across the six metropolitan regions included in the PME dataset.

Subsequently, we assess the contribution of each of the factors listed above in generating the observed decline in informality. We simulate various counterfactuals, changing some exogenous variables while holding others constant at their 2003 levels. These exercises play a dual role. First, from a positive perspective, they shed light on the main driving forces behind the reduction in informality observed in Brazil during the 2000s. Second, from a methodological perspective, they help validate the model by discussing whether our comparative statics results are in line with the evidence available from well-identified reduced-form estimates. We find that changes in workforce composition are the most important factor in the reduction in informality during this period: without increases in skill levels, informality and unemployment rates would have gone up instead of declining. At the same time, we show that the interaction between skill composition, productivity/demand factors, and minimum wages is essential in order to generate the pattern of changes observed across different labor market margins. This highlights the importance of an equilibrium model that is able to consider all of these forces simultaneously.

Though focused on informality, our model also has immediate implications for wage inequality. We use these same simulations to briefly discuss the causes behind the recent decline of wage inequality in Brazil. Our conclusions are broadly in line with the academic debate on this topic, as reviewed by Firpo and Portella (2019): workforce composition and demand shocks are fundamental drivers of the reduction in returns to skills. We add to this literature by showing that the equilibrium effects of these shocks also helped reduce wage inequality through two other channels, by compressing cross-firm wage dispersion for similar workers and by counteracting potentially adverse effects of the minimum wage on informality and unemployment.

Our final quantitative exercise illustrates the use of the model for normative 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 (i.e., 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 option can reduce informality while minimizing impacts on the government budget. The second one is much less 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 much weaker incentives under the second policy since employers do not benefit from the government transfer when the minimum wage is binding. This nuance would be missed by existing theories since
none of them combines informality, unemployment, skill heterogeneity, and minimum wages.

In addition to the theoretical points and quantitative exercises mentioned above, the paper makes two conceptual contributions to 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, relying only on regulatory distortions and 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 (sometimes 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 (e.g., Fields, 1975, 1990, and Rauch, 1991). For the more productive workers in the top tier, informality is a matter of opportunity, which is reflected in 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. This interpretation is supported by the pattern of declining formal wage premiums that has been documented in various developing countries (see, e.g., Lehman and Pignatti, 2007, Bargain and Kwenda, 2011, Botelho and Ponczek, 2011, and Gunther and Launov, 2012). In our model, the two tiers are clearly identified by the two skill levels, and the pattern of a decreasing wage gap results from the binding minimum wage for unskilled workers.  

Our model builds upon many search models of informality but differs from them in key aspects. Boeri and Garibaldi (2007) and Boeri, Garibaldi and Ribeiro (2011) present models with worker heterogeneity but without complementarity between different types of labor and with simplified institutional features. In both papers, the equilibrium displays complete segregation of workers by skill level across the formal and informal sectors. Albrecht, Navarro and Vroman (2009) introduce uncertainty about workers’ productivity in the formal sector and a richer institutional setting, but maintain 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. On the institutional side, Ulyssea (2010) incorporates unemployment insurance and severance payments, and Meghir, Narita and Robin (2015) account for both of these dimensions as well as minimum wages.

---

2This interpretation differs from the original concept of segmented labor markets, as described in Cain (1976) or Dickens and Lang (1985), according to which there are intrinsic structural differences across sectors that reduce transition probabilities and create two almost independent markets. The significant flow of workers in and out of the informal sector documented in various distinct contexts undermines the hypothesis of strong noneconomic barriers of entry to the so-called primary sector.

3To our knowledge, Araujo and Ponczek (2011) present the only alternative model that delivers a decreasing wage gap among salaried workers, but it does so in an entirely different setting and through a very different mechanism (a one-to-one random matching model with asymmetric information, where workers can take employers to court). Bargain et al. (2012) analyze heterogeneity in income gaps between formal and informal self-employed workers.

4This perspective is supported by the experiment in de Mel, McKenzie and Woodruff (2013) and by other empirical evidence showing that firms change their compliance decision in response to changes in tax rates (Monteiro and Assunção, 2012 and Fajnzylber, Maloney and Montes-Rojas, 2011) or in the intensity of enforcement of labor regulation (Almeida and Carneiro, 2012).

5Galiani and Weinschelbaum (2012) model a competitive labor market with heterogeneous firms and workers and self-selection of both into for-
The critical features that set our model apart from 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 within a rich institutional setting, our model reproduces a range of empirical patterns incompatible with previous theoretical models of informality. In addition, it allows us to study the equilibrium effects of 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 Background

The term “informality” is used to describe many different aspects of noncompliance with regulations. In this paper, we focus on the decision by firms and workers not to comply with labor laws 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 comply with minimum wage legislation and 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, the Instituto Brasileiro de Geografia e Estatística, IBGE (IBGE, 2007). The PME collects information on workers and their employment status in the six largest metropolitan regions in Brazil. We concentrate on the period between 2003 and 2012 due to the availability of data collected 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 top row in Table 1 shows that while the average formal hourly wage was 2.54 Brazilian Reais in 2003, the average informal wage was 1.93, or 31% lower. At the same time, there were also substantial differences in the characteristics of these workers and the firms where they worked. Roughly 40% of informal employees had less than 8 years of schooling, while the analogous number was less than 28% for formal employees. And only a small minority
Table 1 – Labor Market Outcomes, Brazil, 2003-2012

<table>
<thead>
<tr>
<th>Sample</th>
<th>Wages in 2003 (R$)</th>
<th>Wage growth (%)</th>
<th>Formal wage gap (%)</th>
<th>Informality (%)</th>
<th>Unemployment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2.54</td>
<td>1.93</td>
<td>28.5</td>
<td>48.5</td>
<td>31.4</td>
</tr>
<tr>
<td>By schooling:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 8 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.87</td>
<td>1.37</td>
<td>29.8</td>
<td>44.8</td>
<td>36.2</td>
<td>22.1</td>
</tr>
<tr>
<td>8 to 10 years</td>
<td>2.04</td>
<td>1.48</td>
<td>24.1</td>
<td>44.6</td>
<td>38.0</td>
</tr>
<tr>
<td>H.S., college drop.</td>
<td>2.69</td>
<td>2.17</td>
<td>13.4</td>
<td>25.8</td>
<td>23.9</td>
</tr>
<tr>
<td>College or more</td>
<td>8.23</td>
<td>6.36</td>
<td>-9.8</td>
<td>11.6</td>
<td>29.4</td>
</tr>
</tbody>
</table>

Source: IBGE/PME, author’s calculations.

Notes: Data are presented as averages for April-December in 2003 and 2012. Informality is the fraction of salaried workers in the private sector with a signed work card. Wage levels, growth, and gaps are calculated by taking the mean log wage in the respective subsample and exponentiating the result (or difference in results across subsamples). Wage levels are shown as multiples of the minimum wage as of 2003. Wage growth is real, corrected using the IPCA price index.

of formal employees worked in firms with 5 workers or less (1 out of 16), while 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 interpretation is not entirely accurate. Table 1 shows that the informality rate among workers with a college degree is 16.8%, not dramatically lower than the overall rate of 28.1%. Moreover, informal workers with a college degree earn almost three times as much as the average formal employee, indicating that at least some informal jobs are clearly much better than the average formal job. Note that since we have restricted our sample to wage earners, these individuals are not self-employed professionals defaulting on taxes or social security contributions. 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.

It remains true, nevertheless, that larger firms are more likely to have a higher fraction of educated workers, irrespective of sector, and that they are also more likely to be formal. While over 52% of formal employees in 2003 had at least a high-school degree, this number was below 40% for informal employees; and while over 85% of formal workers were employed in firms with eleven or more employees, this number was around 50% for informal employees. But the important takeaway is that the relationship between firm size and composition of workers is present in both sectors, suggesting again that at the margin, the technology used by formal and informal firms is not substantially different.

The first row in Table 1 also shows that Brazil experienced a 10.8 p.p. reduction in informality between 2003 and 2012, starting from a level of 28.1%. This reduction of 38% in the informality rate in less than 10 years has no historical precedent. At the same time, there were substantial reductions in unemployment and in the unconditional wage gap between formal and informal workers. In Appendix B, we show that the decline in informality was widespread in the economy and not driven by workforce reallocation towards specific sectors of economic activity (i.e., not due to a movement of employment to industries that are intrinsically more formal). What makes this pattern particularly intriguing is the observation that the cost of formal job offers rose dramatically throughout the period—at a rate con-
siderably faster than the growth in income per capita. The minimum wage, in particular, accumulated real gains of 61% between 2003 and 2012.\textsuperscript{6}

There is some evidence that the enforcement of labor regulations in Brazil became marginally more efficient during this period, which could have helped bring down informality rates. But the change in enforcement was not so substantial and, in addition, seems unable to 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 rise as a consequence of increased enforcement, which is opposite to the large reduction observed in the data. Additionally, while the effect of increased enforcement on unemployment in most models is ambiguous, there was a substantial reduction in unemployment in the data (see, for example, the opposing results in Boeri and Garibaldi, 2007 and Ulyssea, 2010 when compared to Bosch and Esteban-Pretel, 2012 and Meghir, Narita and Robin, 2015).

The focus of this paper, from both the methodological and positive perspectives, is on the changing composition of the labor force. Brazil witnessed, for example, an increase in the proportion of workers with completed high-school education of 16 p.p. (34%) between 2003 and 2012. This change may have contributed to the patterns described here, despite rarely appearing in the literature as an important determinant of informality. Three intuitive arguments hint at its potentially important role. First, since informality is much lower among the 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).\textsuperscript{7} Second, the increase 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. Third, if skilled and unskilled workers are complementary, it should also increase the wage of unskilled workers, making the minimum wage less binding, which should further increase formalization and reduce the formal wage gap. 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 driving force behind the evolution of informality.

In the next section, we develop a model that is able to incorporate the dimensions discussed here and use it, among other things, to help rationalize both the cross-sectional patterns and the changes in informality observed in Brazil during the 2000s.

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 that 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 (2018).

\textsuperscript{6}The rise in informality in Brazil during the 1990s is traditionally attributed to the increase in formal labor costs brought about by the 1988 constitution. Barros and Corseuil (2001) explain how the new constitution significantly raised employment costs (payroll taxes and contribution, firing costs, and mandated benefits). Bosch, Goni-Pacchioni and Maloney (2012) discuss how these changes impacted informality during the 1990s. We present a brief discussion of changes in labor legislation and tax rates after 2003 in Appendix A. Apart from the increase in the minimum wage, changes in labor regulations after 2003 were negligible.

\textsuperscript{7}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 informality rates in Brazil from 2002 to 2007.
Instead, firms make one single formality decision encompassing all of their job relations. From now on, we use the terms “informal firm” and “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. In the model, 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 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. Firms differ in a productivity parameter z that increases overall productivity and also the relative productivity of skilled workers. Firms decide whether their job relations will be formal or informal by weighting the relative costs of labor regulations against an informality penalty that increases in firm size and can vary across firms (according to a measure of enforcement intensity, k). Skill-biased productivity and the informality penalty are the main determinants of 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, but there are skilled workers employed in the informal sector as well.

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, θs and θu. 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, Us and Uu. 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 labor. The problem of the firm is then to choose a vacancy-posting strategy—or, equivalently, a 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 θs, θu, Us and Uu that are consistent with the aggregate behavior of firms and workers.

3.1 Labor Markets

There are two separate labor markets, one for each skill level. Firms need to post vacancies in order to find workers. 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, is 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

---

8For analytical convenience, we do not include on-the-job search. Instead, we use convex vacancy-posting costs to generate cross-firm wage dispersion for similar workers, which is an important component of wage differentials by skill and by sector. Ruling out on-the-job search prevents us from fully matching transitional dynamics since there are no job-to-job flows. With on-the-job search, each firm’s decisions depend on endogenous distributions of wages and employment across firms. Without it, the state of the labor market is fully summarized by only two scalars: the value of unemployment and the labor market tightness. This tractability allows us to add complexity along many other dimensions while still being able to estimate the model structurally. These dimensions are worker heterogeneity, imperfect substitution and decreasing returns to scale at the firm level, realistic institutional features, and regional variation. Even without on-the-job search, our full estimation procedure still requires many days when using 100 cores in a modern compute cluster.
worker is \( q(\theta_i)dt \). \( \theta_i \) is the labor market tightness in market \( i \), that is, the ratio of vacancies to unemployed workers: \( \theta_i = \frac{V_i}{s_i}, i \in \{s,u\} \). The probability that an unemployed worker finds a job in a short time interval \( dt \) is given by \( \theta_i q(\theta_i)dt \).

We make no distinction between the visibility of formal and informal job posts in the search process.\(^9\) 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 with skill level \( i \) 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 \( i \).

Firms operating in sector \( j \in \{for, inf\} \) for formal and informal, respectively, pay an instantaneous cost \( \Xi_j'(v) \) for posting \( v \) vacancies for workers of skill level \( i \). We assume that the marginal vacancy-posting cost \( \xi_j'(v) = \frac{d\Xi_j'(v)}{dv} \) is increasing, as in Bertola and Caballero (1994) and Acemoglu and Hawkins (2014), and that \( \Xi_j'(0) = \xi_j'(0) = 0 \). Small firms posting few vacancies often rely on networks—relatives, friends, etc.—to find workers at low cost (Chandrasekhar, Morten and Peter, 2020). Larger firms need to access the broader labor market and pay costs associated with advertisement and selection. The largest firms face even higher costs as they exhaust suitable candidates in their local area. We also allow for differences in vacancy-posting costs between sectors and skill levels. We show later on that this formulation generates a firm-size wage premium that is, at the same time, an important component of wage differences between sectors and the reason for wage dispersion within sector and skill level.

### 3.2 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 a productivity parameter distributed across firms according to a distribution function \( G_z(z) \). We assume that \( F^z(\cdot) \) strictly concave in \( (n_s,n_u) \) for any \( z \) in the support of \( G_z(z) \), and increasing in \( z \). We further assume that \( \frac{d^2F^z(n_s,n_u)}{dndz} > 0 \). That is, firms with higher \( z \) have a relatively higher productivity of skilled labor. 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, \( v_s \) and \( v_u \). Firms also decide on whether to comply with labor regulations or not. We assume that this decision is made at the beginning of time and cannot be changed thereafter. If a firm complies, it must pay taxes \( \tau \) over its total payroll. If a firm chooses instead to hire workers informally, it avoids payroll taxes but incurs in an informality penalty represented by the function \( p^k(n_s + n_u) \), with \( \frac{dp^k(n_s + n_u)}{dn_s} \cdot \frac{dp^k(n_s + n_u)}{dn_u} > 0 \). In this expression, \( n_i \) is the total number of workers of skill level \( i \) hired by the firm, and \( k \) is a firm-specific enforcement parameter with cumulative distribution \( G_k(k) \). As in Meghir, Narita and Robin (2015),

\(^9\)This assumption is of little consequence in our quantitative exercise, given that we allow for differences in the vacancy-posting costs between sectors (see below). In the data, we do not observe numbers of vacancies, only employment patterns. A scenario where formal vacancies cost twice as much as informal ones, but both are equally visible, is observationally similar to another scenario where vacancies cost the same in both sectors, but informal ones are twice as visible. Bosch and Esteban-Pretel (2012) also assume equal visibility of formal and informal vacancies. Different visibility by sector, as in Meghir, Narita and Robin (2015), could lead to differences in congestion externalities by sector. Assuming different matching functions, as in Ulyssea (2010), would rule out congestion externalities across sectors, conditional on the unemployment rate. We believe that measuring the wedge between within-sector and cross-sector congestion externalities would be difficult and beyond the scope of this paper. Without these measurements, the functional form differences between our approach and those of the papers cited above are largely arbitrary.
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, such as courts, available to formal firms. The assumptions on $\rho^j(\cdot)$ capture the idea that informality costs increase more than proportionally with size, as firms become more visible and it becomes harder to evade government monitoring.

We interpret firm heterogeneity in enforcement $k$ as coming from differences—in location or sector of economic activity—that make it harder for some firms to evade labor regulation or that increase the value of being formal. We assume $k$ is independent of $z$; any relationship between productivity and intensity of enforcement is driven endogenously by the dependence of $\rho^j(n_s + n_u)$ on firm sizes.

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

$$\psi^z_{j,k}(n_s, n_u, v_s, v_u) = \begin{cases} F^z(n_s, n_u) - \sum_{i=s,u} [(1+\tau)n_i w^z_{i,for}(n_s, n_u) + \Xi^z_{i,for}(v_i)], & \text{if } j = \text{for}, \\
 F^z(n_s, n_u) - \rho^k(n_s + n_u) - \sum_{i=s,u} [n_i w^z_{i,inf}(n_s, n_u) + \Xi^z_{i,inf}(v_i)], & \text{if } j = \text{inf}, \end{cases}$$

where $w^z_{i,for}(n_s, n_u)$ and $w^z_{i,inf}(n_s, n_u)$ are wages that the firm $(z, k)$ pays to workers of type $i$, according to its compliance status $j$ and the current number of employees, $n_s$ and $n_u$. We describe how the wage functions are determined in the next subsection. Instantaneous profits are given by total revenue minus total payroll, vacancy-posting costs, and either payroll taxes (in the case of formal firms) or the informality penalty (for informal firms).

Job relations are destroyed at exogenous separation rates $\lambda^j_i$, which depend on skill level and compliance decision. This allows the model to capture the empirical pattern of higher labor turnover among unskilled and informal workers.\(^{10}\) For simplicity, we ignore granularity issues when analyzing the dynamics of employment at the firm level, following other papers with search frictions, multi-worker firms, and concave production functions (e.g., Cahuc, Marque and Wäsmmer, 2008 and Acemoglu and Hawkins, 2014). The law of motion of labor quantities inside each firm is

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

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 on in the paper, we show that all job offers are accepted in equilibrium.

\(^{10}\)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 unskilled employees should have stronger incentives to maintain job relations. 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. Therefore, we take this difference as exogenous in our empirical analysis.
The problem of the firm in its recursive Bellman formulation is given by

\[ \Pi^{z,k} = \max_{j \in \{for,inf\}} \Pi^{z,k,j}(n^z_s,n^z_k,\bar{n}^z_{s,j,initial},\bar{n}^z_{k,j,initial}), \] with

\[ \Pi^{z,k,j}(n_s,n_u) = \max_{\{\nu,u\}} \left( \frac{1}{1+rdt} \right) \left\{ \psi^{z,k,j}(n_s,n_u,v_s,v_u) \right\} \]

s.t. \( n_j^+ = n_j + \hat{n}_j dt = (1 - \lambda^j dt) n_j(t) + v_i q(\theta_i) dt, \quad i = s,u. \)

(1)

For a firm with productivity \( z \) and enforcement intensity \( k \), given a compliance decision \( j \), the total present value of profits is the sum of instantaneous profits earned at the end of the short 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 the present discounted value of profits. We restrict attention to steady-state solutions where the numbers of workers of different types are constant in each firm.\(^{11}\)

Denote by \( \pi^{z,k,j}_i(n_s,n_u) \) the marginal value of an additional worker of type \( i \) in a firm of type \( z,k \), with compliance status \( j \): \( \pi^{z,k,j}_i(n_s,n_u) = \frac{\partial F^{z,k,j}(n_s,n_u)}{\partial n_i} \). We derive the first-order conditions for the firm’s problem in Appendix C. By imposing \( n_i = 0 \) in the FOCs, the expressions are simplified to:

\[ (r + \lambda^j)\pi^{z,k,j}_i(n_s,n_u) = \begin{cases} F^z_i(n_s,n_u) - (1 + \tau) \left[ w^{z,for}_i(n_s,n_u) + \sum_{l=x,M} \frac{n_l}{\partial n_j} w^{z,inf}_j(n_s,n_u) \right] & \text{for } j = for \\ F^z_i(n_s,n_u) - w^{z,inf}_i(n_s,n_u) + \sum_{l=x,k} \frac{n_l}{\partial n_i} k_l(n_s + n_u) & \text{for } j = inf, \end{cases} \]

(2)

\[ \pi^{z,k,j}_i(n_s,n_u) = \frac{\xi^j_i(v_i)}{q(\theta_i)}, \]

(3)

with \( F^z_i(n_s,n_u) = \frac{\partial F^z(n_s,n_u)}{\partial n_i} \). The instantaneous rent is given not only by the difference between marginal product and wage but also by the effect of this additional employee, due to changes in marginal productivities, on the wages of all other workers currently employed by the firm. At the time of the hiring decision or bargaining, previous vacancy costs are sunk and thus do not appear in this expression.

Equation 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 marginal cost \( \xi^j_i(v_i) \) 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 formal firm analogous to the one in Cahuc, Marque and Wasmer (2008), in which marginal product equals a generalized notion of marginal cost:

\(^{11}\)To be entirely precise, we assume that firms are born at the beginning of time with stocks of workers \( n^z_{s,j,initial},n^z_{k,j,initial} \) equal to the steady-state employment values, given their type and sectoral choice. This choice, along with the assumptions that the formality decisions are made at the beginning of time and that firm parameters \( z \) and \( k \) are constant over time, eliminates the need for analyzing out-of-steady-state firm dynamics. The remainder of the paper therefore restricts its attention to steady-state solutions. See Acemoglu and Hawkins (2014) for an interesting study of firm dynamics in a context with multi-worker firms and convex vacancy-posting costs, as in this paper, but without informality and worker heterogeneity.
We denote the optimal labor choices in firm $z$ mandated benefits and job destruction rates). That would not be the case if the model had on-the-job search and offer matching by incumbent firms.

real wages.

informality penalty is irrelevant. We use the same abuse of notation with wage functions and the optimal wage in equilibrium.

formal workers is analogous, just omitting the payroll tax $\tau$, and adding the marginal effect of $n_i$ on the informality penalty $\rho^k(\cdot)$, and writing the dependency on $k$.\footnote{In a slight abuse of notation, we write $n_i^{for}$ and $n_i^{inf}$ interchangeably. This is because, conditional on being in the formal sector, the informality penalty is irrelevant. We use the same abuse of notation with wage functions and the optimal wage in equilibrium.}

### 3.3 Wage Determination

Wages are 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 $\sigma$, which corresponds to the bargaining power of workers. Unlike the standard model in Pissarides (2000), we do not assume homogeneous labor nor constant returns to scale in the production function, and we allow workers and firms to engage in renegotiation after the initial match. As discussed in Stole and Zwiebel (1996a), 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.\footnote{Since we look at steady states, we interpret this assumption as representing the fact that in a steady state, a change in the composition of workers in the firm affects, through bargaining, the wages of all workers. This logic is appealing when thinking about the long-run determination of real wages.} We follow the solution developed by Cahuc, Marque and Wasmer (2008), who analyze this type of problem in a context with search frictions.

This framework incorporates complementarities across different types of labor and, therefore, equilibrium responses to changes in the relative supply of and demand for different types of workers, features that are absent from traditional search models. In contexts of rapidly changing schooling, as seen in many developing countries, these are potentially important forces in explaining the evolution of labor market outcomes, in particular informality. Also unlike 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. Adding this degree of freedom can be a straightforward way to create differences across sectors. In our model, labor market regulations (such as the minimum wage) and differences in observable variables (such as worker heterogeneity, turnover rates, and firm sizes) play this role while also allowing for a richer pattern of wage dispersion. See Subsection 3.5 below for a detailed discussion. We consider the implications of lower bargaining power in the informal sector in a robustness exercise (Appendix I.2).

We first describe how wages are determined in the absence of a binding minimum wage. Then, we explain how the introduction of a binding minimum wage changes the results. Define $J_i(w)$ 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 $w$.\footnote{Firm heterogeneity $z$ does not directly affect the valuation of the job conditional on wages and the formality choice (which determines mandated benefits and job destruction rates). That would not be the case if the model had on-the-job search and offer matching by incumbent firms.} 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 benefit, possibly including 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 $J_i^{\text{for}}(w)$, instead of in $U_i$, as done by Ulyssea (2010). Since workers are assumed to be risk-neutral, this greatly simplifies the solution.15

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

$$\begin{align*}
\rho_i^{\text{for}}(w) &= a_i w + b_i + \lambda_i^{\text{for}} \left[ U_i - J_i^{\text{for}}(w) \right], \quad \text{and} \\
\rho_i^{\text{inf}}(w) &= w + \lambda_i^{\text{inf}} \left[ U_i - J_i^{\text{inf}}(w) \right],
\end{align*}$$

(4) (5)

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

The value $J_i(w) - U_i$ 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 $\pi_i^{\text{z,k}}(n_s, n_u)$, which was discussed in the previous subsection. So, the Nash sharing rule imposes that the wage function $w_i^{\text{z,k}}(n_s, n_u)$ must satisfy

$$\begin{align*}
(1 - \sigma) \left[ J_i^{\text{z,k}}(n_s, n_u) - U_i \right] &= \sigma \pi_i^{\text{z,k}}(n_s, n_u), \quad \text{where } i \in \{s, u\}, \quad \text{and } j \in \{\text{for, inf}\}, \quad \forall \, z, \, n_s, \, \text{and } n_u.
\end{align*}$$

(6)

Due to the derivative terms in expression 2 (for $\pi_i^{\text{z,k}}$), 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 Wasmier (2008) to account for two sectors, heterogeneous firms, mandated benefits, and payroll taxes. The resulting wage functions are

$$\begin{align*}
w_i^{\text{z,k}}(n_s, n_u) &= \frac{1 - \sigma}{c_i} \left( rU_i - b_i \right) + \frac{1}{1 + \tau} \int_0^1 \varepsilon \frac{\partial F(c_i \varepsilon, n_s, n_u)}{\partial n_i} \, d\varepsilon, \quad \text{and} \\
w_i^{\text{z,k}}(n_s, n_u) &= (1 - \sigma) rU_i + \int_0^1 \varepsilon \frac{\partial H(c_i \varepsilon, n_s, n_u)}{\partial n_i} \, d\varepsilon,
\end{align*}$$

(7) (8)

with $c_i = [(1 - \sigma) a_i + \sigma (1 + \tau)]$ and $H(c_i \varepsilon, n_s, n_u) = F(c_i \varepsilon, n_s, n_u) - \rho(c_i \varepsilon, n_s, n_u)$.

As in the solution to 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 equations reduce to $w_i^{\text{z,k}}(n_s, n_u) = \frac{1 - \sigma}{c_i} \left( rU_i - b_i \right) + \sigma \frac{\partial F}{\partial n_i}$ and $w_i^{\text{z,k}}(n_s, n_u) = (1 - \sigma) (rU_i - b_i) + \sigma \frac{\partial H}{\partial n_i}$. However, with decreasing returns to scale, heterogeneous

15This is different from the formulation seen in some search models, where unemployment insurance is part of the value of unemployment. In our model, including $\rho_i^{\text{B}}$ in the valuation of $U_i$ would not be correct because only formal workers are eligible for unemployment benefits. This choice has implications for wage setting in partial equilibrium. In our formulation, unemployment benefits decrease bargained wages in the formal sector because workers value unemployment benefits in expectation. In the alternative specification sometimes seen in the literature, unemployment benefits add to workers’ threat point in the bargain, and thus increase bargained wages. But this second formulation is only valid if formal workers are always eligible for unemployment benefits and can use them as credible threats when bargaining. In Brazil, workers are only eligible for unemployment benefits if they are not fired for just cause and if they have been employed for at least six months in the same firm. So, it is not clear to what extent unemployment benefits can be used as leverage during a bargaining process. This is particularly important because of the historically short duration of employment spells in the Brazilian formal labor market (see, for example, Gonzaga, 2003). The formulation adopted in our paper, and in Ulyssea (2010), seems therefore more appropriate for our context. Nevertheless, we explore the alternative formulation in Appendix I.1 and show that it does not change the qualitative results of the model and has only modest implications for the quantitative results. Appendix I.1 also discusses this issue in further detail.

16Our bargaining expression in the presence of payroll taxes differs from that in Mortensen and Pissarides (2001) because we define $\sigma$ as the effective bargaining share of workers, while they define their bargaining parameter $\beta$ as the exponent in the generalized Nash bargaining solution. To convert from their notation to ours, one should use the expression $\sigma = a_i \beta / (a_i - (1 + \tau - a_i)(1 - \beta))$. Thus, in our comparative statics exercises below, the share of rents accruing to workers is always constant, whereas it could vary under Mortensen and Pissarides (2001).
labor, and intra-firm bargaining, the second term in these expressions is not simply the marginal productivity of the input considered, but instead a weighted average of inframarginal productivities, with weights $\epsilon^{\frac{1+\sigma}{\sigma-1}}$ higher for points closer to the margin. We refer the reader to Stole and Zwiebel (1996b, Stole and Zwiebel (1996a), and Cahuc, Marque and Wasmer (2008) for detailed discussions of the characterization of this type of solution. In Appendix D, we derive our results (using a more general formulation with skill-specific payroll taxes) 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. In addition, we believe that distortions introduced by labor market regulations such as the minimum wage are key features of informal labor markets, which constitute the focus of this paper.

If the bargained wage in a formal firm for one type of worker—typically, unskilled workers—is lower than the minimum wage, then the minimum wage constraint is binding. The Nash bargaining equation is no longer satisfied for unskilled workers; indeed, in this situation, these workers receive a share of rents larger than $\sigma$. This also implies that the previous wage function for skilled workers is not valid anymore, since the term $\frac{\partial w_z}{\partial n_s}$ for unskilled in equation 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 when the minimum wage binds for unskilled workers, the wage equation for skilled workers in the formal sector is

$$w_{z, for}^*(n_s, n_u) = \frac{1 - \sigma}{c_s} (rU_s - b_s) + \frac{1}{1 + \tau} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma-1}} \frac{\partial F_z(\epsilon n_s, n_u)}{\partial n_i} d\epsilon.$$

From the perspective of a firm, whether the minimum wage binds is a function not only 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. 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, $\pi_{z, for}^{s,k}$ and $\pi_{z, for}^{u,k}$. 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 $\pi_{z, for}^{u,k}$ and a discontinuous decrease in $\pi_{z, for}^{s,k}$. The increase in $\pi_{z, for}^{s,k}$, 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. 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

---

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 A.2 can help to understand this discussion.
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.

### 3.4 Equilibrium

Now we turn to the determination of the endogenous equilibrium variables \( \theta_i \) and \( U_i \). The labor market tightness, as explained in subsection 3.1, is given by the ratio of vacancies to unemployed workers. Define the measure of workers of type \( i \) employed in sector \( j \) as

\[
N_i^j = m \int \int n_i^{z,k,j} \mathbf{1} \left( \text{Firm } z, k \text{ chooses compliance } j \right) dG_z(z) dG_k(k).
\]

Since, in equilibrium, \( n_i = 0 \) for all firms, \( V_i^{z,k,j} = \lambda_i n_i^{z,k,j} / q(\theta_i) \implies V_i^j = \lambda_i N_i^j / q(\theta_i) \). We can therefore find the expressions that pin down \( \theta_i \),

\[
\theta_s = \frac{\lambda_s^{for} N_s^{for} + \lambda_s^{inf} N_s^{inf}}{q(\theta_s) \left( \eta - N_s^{for} + N_s^{inf} \right)} \quad \text{and} \quad \theta_u = \frac{\lambda_u^{for} N_u^{for} + \lambda_u^{inf} N_u^{inf}}{q(\theta_u) \left( 1 - \eta - N_u^{for} + N_u^{inf} \right)}.
\]

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

\[
r U_i = d_i + \theta_i q(\theta_i) \left[ \phi_i E[J_i^{for}(w_i^{for})] + (1 - \phi_i) E[J_i^{inf}(w_i^{inf})] - U_i \right]
= \frac{1}{\theta_i q(\theta_i) + \frac{\phi_i}{r + \lambda_i^{for}} + \frac{(1 - \phi_i)}{r + \lambda_i^{inf}}} \left[ \frac{1}{\theta_i q(\theta_i)} d_i + \frac{\phi_i}{r + \lambda_i^{for}} \left[ a_i E[w_i^{for}] + b_i \right] + \frac{(1 - \phi_i)}{r + \lambda_i^{inf}} E[w_i^{inf}] \right].
\]

The instantaneous utility of being unemployed is composed of a utility flow \( d_i \), which may be negative, plus 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 \( \phi_i = V_i^{for} / V_i^{for} + V_i^{inf} = \frac{\lambda_i^{for} N_i^{for}}{\lambda_i^{for} N_i^{for} + \lambda_i^{inf} N_i^{inf}} \) that the match is with a formal firm. The value of finding a job in a given sector is given by the expected value of the utility associated with the average job in that sector. The second line in the equation above shows that the instantaneous utility of unemployment can be expressed as a weighted average of the utility flow \( d_i \) and the average wages in the formal and informal sectors, where the weights are functions of the transition probabilities between states and expected durations.

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

1. the wage functions solve the system of differential equations given by expressions \( 2 \) and \( 6 \);
2. the labor schedules \( n_i^{z,k,j} \) solve equation \( 3 \) 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 \( 1 \);
4. the labor market tightnesses are consistent with equation \( 7 \); and
the unemployment values are consistent with equation 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 some exercises in the next sections, we show how this surplus varies with certain policy changes.

Solving the model numerically is not trivial and can be computationally consuming. The problem of each firm \( z, k \) requires finding optimal employment under the formal and informal regimes and then comparing profits. In each of these optimization problems, the objective function requires numerical integration. In turn, each evaluation of the equilibrium-finding procedure (that is, each guess of \( \theta_i \) and \( U_i \)) requires solving the problem of some firm types, interpolating the results for all others, and aggregating the results using another numerical integration. In Appendix C, we show how to rewrite the firm’s first-order conditions such that fewer numerical integrations are required. Appendix E provides the details of the numerical procedures. Using these optimized procedures, equilibria can usually be found in less than one minute using a modern processor.

3.5 Discussion: Compensating Differentials, the Formal Wage Premium, and the Firm-Size Wage Premium

We can find an expression for how equilibrium wages \( w_{z,k,j}^{i} \) vary across firms by plugging in optimal firm sizes \( n_{z,k,j}^{i} \) in the bargaining equations 6 and using the definition of the values of employment from equations 4 and 5:

\[
w_{z,for}^{i} = \max \left\{ \frac{1}{a_i} \left[ rU_i - b_i + (r + \lambda_{for}^i) \frac{1}{1 - \sigma} \right] \bar{e}_i \left( \frac{\lambda_{for}^i n_{z,for}^{i}}{q(\theta_i)} \right), \bar{w}_i \right\}, \text{and}
\]

\[
w_{z,inf}^{i} = rU_i + (r + \lambda_{inf}^i) \frac{1}{1 - \sigma} \bar{e}_i \left( \frac{\lambda_{inf}^i n_{z,inf}^{i}}{q(\theta_i)} \right).
\]

The model has three sources of cross-firm wage dispersion among similar workers: compensating differentials, minimum wages, and firm-size wage premiums. We can isolate the first component by assuming that there is no minimum wage (\( \bar{w}_i = 0 \)) and that vacancy-posting costs are linear and identical across sectors (\( \bar{e}_i(\cdot) = \bar{e}_j \)). These assumptions eliminate the dependency of equilibrium wages on \( z, k \) in the expressions above, meaning that in this case, there are only four wage values in the economy (one for each combination of skill \( i \) and sector \( j \)). From equation 6, this scenario also implies that workers are indifferent when it comes to formal and informal employment. We can then find the following compensating differentials relationship:

\[
w_i^{inf} = a_i w_i^{for} + b_i + \left( \lambda_{i,inf} - \lambda_{i,for} \right) \frac{\bar{e}_i}{1 - \sigma} \frac{1}{q(\theta_j)}. \tag{10}
\]

In this setting, if jobs in both sectors have the same expected duration (\( \lambda_{i,for} = \lambda_{i,inf} \)), then informal wages should be higher than formal ones, with the difference equal to the value that workers ascribe to mandated benefits (the \( a_i \) and \( b_i \) parameters). If the expected duration in the formal sector is longer, as we see in the data, then informal wages should
be even higher.

Now we introduce the minimum wage. When the minimum wage binds for a given skill level, the bargaining equation does not hold for formal firms; workers in the formal sector have higher de facto bargaining power. In this situation, informal wages do not fully compensate for benefits and duration, and workers strictly prefer formal jobs. However, if matched with an informal firm, they still accept its employment offer since it is too costly to remain unemployed and wait for a better job offer to arrive.

Finally, in the full model, differences in the replacement costs of workers lead to firm-size wage premiums conditional on sector. More productive firms hire more workers, post more vacancies, and thus face higher marginal vacancy-posting costs. Rents associated with the match are higher in these firms, and part of these rents are seized by workers through bargaining. Thus, larger firms typically offer higher wages, and their job offers are strictly preferred from the point of view of workers. Because more productive firms tend to be formal, this channel also works against the compensating differentials logic summarized by equation 10, increasing average formal wages relative to average informal wages.

In the Brazilian data, the formal wage premium is higher for unskilled workers, while the firm-size premium is higher for skilled workers. The elements discussed above can explain this pattern. Because the minimum wage does not bind for skilled workers, the formal wage premium is generated by the compensating differentials relationship on one end and the firm-size wage premium on the other. For unskilled workers, a binding minimum wage flattens the formal wage curve on the left of the firm-size distribution, reducing the firm-size wage premium. And because the minimum wage boosts formal wages in small firms, it increases the formal wage premium relative to skilled workers. We show in the next section that the estimated model successfully replicates these patterns. In addition, since firm-size premiums tend to be irrelevant among smaller firms in the estimated model, equation 10 indeed holds (as an approximation) for skilled jobs in firms at the margin of informality.

At the firm level, differences in mean wages between sectors reflect selection and incentives. On average, formal firms are more productive than informal firms, which, as discussed above, might lead to wage differences due to the size of match rents. But at the margin (defined as $z(k)$ such that formal profits equal informal profits), formal and informal firms are structurally identical. Even for these firms, employment decisions and wages may differ substantially across sectors due to regulatory distortions and differences in job destruction rates. It remains true, though, that these marginal firms are virtually indifferent to whether they operate in the formal or informal sectors and could have made different compliance choices if parameter values were slightly different.

### 4 Fitting the Model

We fit the model to the Brazilian labor market in 2003, estimating most parameters and imputing others from labor legislation and past literature. We choose 2003 as our reference point for two reasons. First, it is close to the beginning of the reversal of the informality trend. Second, one of the datasets that we use, the Informal Urban Economy Survey (Economia Informal Urbana, ECINF), is only available for 1997 and 2003.

Our main data source is the Brazilian Monthly Employment Survey (Pesquisa Mensal de Emprego, PME), a household panel survey conducted by the Brazilian Census Bureau (Instituto Brasileiro de Geografia e Estatística, IBGE), similar
in design to the US Current Population Survey. The PME interviews households up to eight times over a period covering 16 months, collecting information on employment, wages, occupational choice, formality status and other characteristics, such as educational attainment. We restrict the sample to the months of April through December of 2003, when the minimum wage was constant, and to individuals between 16 and 59 years of age who were either salaried workers (in the private sector) or unemployed.\footnote{When validating the model and simulating counterfactuals, we analyze changes from 2003 to 2012. Using the months of April through December in both years ensures that the minimum wage is constant within each year analyzed and that we are comparing the same months in both years (so that seasonality is not a concern).}

In addition to the PME, we also use three other datasets generated by the IBGE: (i) the ECINF, which targets small urban firms—most of which are unregistered—and thus provides an estimate of the number of informal firms in the economy; (ii) the Central Registry of Firms (\textit{Cadastro Central de Empresas}, CEMPRE), a registry of formal firms; and (iii) annual projections of the size of the workforce. We use a few additional data sources to obtain some national-level information, such as the labor share of national income. These sources are listed in subsection 4.4.2.

The PME covers the six main metropolitan regions of Brazil, which correspond to the state capitals Belo Horizonte, Porto Alegre, Recife, Rio de Janeiro, Salvador, and São Paulo. We treat each of these metropolitan regions as a different labor market. We allow some parameters to vary across regions to capture differences in productivity and institutions. Region-specific parameters are marked with a subscript $r$. Exploring regional variation, in addition, allows us to validate the model by testing whether the predicted changes across regions are similar to what is seen in the data. This validation exercise, along with others, is presented in the next section.

The remainder of this section describes how parameters are estimated and shows that the model fits the data well. Several of our estimation targets—such as formal wage premium, firm-size wage premiums, and distribution of workers by firm size in the formal and informal sectors, all calculated by skill level—would be impossible to match using existing models of informality. We close the section by illustrating the properties of the equilibrium generated by the model using the example of the metropolitan region of Recife.

### 4.1 Functional Forms

Unless otherwise noted, all parameters are assumed to be strictly positive.

Firms use a CES production function

$$F^{z,r}(n_s, n_u) = A_r z [B(z)n_s^\gamma + (1 - B(z))n_u^\gamma]^\frac{\alpha}{\gamma}, \text{ with } B(z) = \frac{1}{1 + \left(\frac{z}{B_1}\right)^{-B_2}}.$$  

$A_r$ represents total factor productivity for the smallest firms in each region (those with $z = 1$), while $z$ captures productivity heterogeneity across firms. The exponent $\alpha \in (0, 1)$ determines returns to scale, while $\gamma \in (-\infty, 1)$ determines the micro-level elasticity of substitution between skilled and unskilled labor: $1/(1 - \gamma)$. The function $B(z) : (0, \infty) \to (0, 1)$ determines the relative weight of skilled labor in production, representing the idea discussed in section 3 that more productive firms are also more intensive in skilled labor. It depends on $B_1$ and $B_2$, which determine, respectively, the level and slope of the function.\footnote{See Appendix Figure A.8 for plots of $B(z)$ that illustrate the role of these two parameters (using the actual estimated model and selected parameters).}
Firm-specific productivity \( z \) follows a bounded Pareto distribution with support \([1, 10,000]\):

\[
G_z(z) = \frac{1 - z^{-T_r}}{1 - 10,000^{-T_r}},
\]

where \( T_r \in (1, \infty) \) is a region-specific distribution parameter. We choose a Pareto distribution to account for the fact that while the majority of firms are small, a large part of the workforce is employed by large firms (see IBGE, 2005). It is bounded at 10,000 for computational convenience; the support is wide enough to allow us to capture the cross-sectional distribution of employment by firm size in the Brazilian economy. The lower the region-specific parameter \( T_r \), the fatter the tail of the productivity distribution. Together, regional heterogeneity in \( A_r \) and \( T_r \) capture potential differences in infrastructure, access to capital markets, supply of skilled entrepreneurs, industrial composition, etc.

The informality penalty is given by:

\[
\rho^{k,r}(n_s + n_u) = p_{k,r} \times (n_s + n_u)^{p_r}, \quad \text{where} \quad p_{k,r} = \exp(P_r + (k - 3)P_D),
\]

and \( k \) has a discrete uniform distribution with support \( \{1, 2, \ldots, 5\} \). In this formulation, \( P_r \in \mathbb{R} \) is a region-specific measure of enforcement of regulations (corresponding to the log of the median enforcement intensity) and \( P_D \) is a measure of dispersion (assumed to be constant across regions). As mentioned in section 3, heterogeneity in enforcement intends to capture differences across industries, local enforcement capabilities, and unobserved factors that make firms more or less visible to labor inspectors.

In the specification of the matching technology, we follow the literature and use a Cobb–Douglas function, yielding \( q(\theta) = D\theta^{-E} \), where \( D \) is the matching scale and \( E \) is the matching elasticity.

The marginal vacancy-posting cost is assumed to take the following form:

\[
\xi^j_i(v) = \xi^j \xi^i [1 - \exp(-\xi^S v)],
\]

where \( \xi^i \) and \( \xi^j \) are skill- and sector-specific cost shifters, respectively. This function is an S-shaped curve with intercept zero that eventually converges to \( \xi^j \xi^i \). Its steepness is determined by \( \xi^S \). Without loss of generality, we normalize \( \xi^i = 1 \). The total vacancy-posting cost corresponding to this marginal cost is

\[
\Xi^j_i(v) = \xi^j \xi^i \left[ v + \frac{\exp(-\xi^S v) - 1}{\xi^S} \right].
\]

counterfactuals). It is also worth mentioning that in section 3, we assume that \( F^i(\cdot) \) is strictly increasing in \( z \), but this is not always true for all possible parameters and input values under the functional form \( B(z) \) above. As an example, in the limit case where all workers are unskilled and \( B_2 \) is close to infinity, such that \( B(z) \) goes quickly from 0 to 1 as \( z \) crosses \( B_1 \), \( F^i(\cdot) \) can be decreasing in \( z \) if \( \gamma > 0 \). However, \( F^i(\cdot) \) is increasing in \( z \) for all “reasonable” parameter values and skill shares.
Job destruction rates are a composition of region, skill, and sector factors:

$$\lambda_{i,r}^j = \lambda_r \lambda_i \lambda^j,$$

where we normalize the skilled and formal components, $$\lambda_i = \lambda_i^{for} = 1$$.

The valuation of fixed benefits by workers is given simply by

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

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).

Finally, we let two additional sets of parameters vary by region: the shares of skilled workers $$\eta_r$$ and the utility flow from unemployment $$d_i,r$$.

### 4.2 Empirical Specification of Skill Levels

To add empirical content to the model, we need to link skill levels to observables in the data. In 2003, many workers in Brazil had low levels of education, were very likely to work in the informal sector, and were very likely to earn close to the minimum wage when employed in the formal sector (especially if they lived in a relatively poor region). Older and more experienced workers, and those with at least a high school diploma, were much more likely to be formally employed. This is the margin we attempt to capture with our definition of skills in the model.

Based on this idea, we match skill levels to age–education groups in the following way. Workers are classified into 12 groups, corresponding to combinations of three schooling groups (less than 8 years, 8–10 years, and 11 or more years) and four age groups (16–19, 20–24, 25–29, and 30–59). We assume that all workers with less than 8 years of schooling and less than 20 years of age are unskilled and that all workers with more than 11 years of schooling and at least 30 years of age are skilled. The other 10 groups are interpreted as combinations of these two types, having varying shares of skilled workers. This matching procedure, detailed in Appendix F (Step 2), generates as a final outcome the share of skilled individuals in each of the 10 intermediary age–education groups discussed above. These shares are used to calculate all skill-specific variables—parameters and moments—mentioned elsewhere in the paper. The skill composition of each age–education cell generated by this procedure is shown in Appendix Table A.4.

### 4.3 Parameters Imputed from the Data and Literature

Table 2 presents a first subset of the parameter values we use.

The share of skilled workers $$\eta_r$$ and the job destruction rates $$\lambda_{i,r}^j$$ are estimated from the PME using the definition of skills discussed in the previous subsection (for conciseness, we present national averages in the table). When we estimate formal and informal job destruction rates conditional on skill, unobserved heterogeneity in worker ability might introduce a bias. To sidestep this issue, we estimate job destruction rates using a Cox proportional hazards model with controls for residual worker ability within worker group.\(^\text{20}\)

\(^{20}\)The bias might arise because unobserved ability could have a direct effect on the job destruction rate, and more able workers within age–education group are more likely to be formally employed. Given this concern, the job destruction rates are estimated in the following way. First,
We impute the measure of firms $m$ using the total number of salaried workers and firms, both formal and informal, for Brazil as a whole, and we assume that it applies to all regions in our dataset. The PME asks unemployed workers what the nature of their last employment was. We use this information to proxy for the fraction of unemployed workers who are looking for salaried jobs. We estimate that salaried workers and unemployed workers looking for salaried jobs account for 73% of the workforce. We obtain the number of formal firms from CEMPRES and the number of informal firms from ECINF, excluding self-employed workers. Since the PME covers the six main metropolitan regions, while ECINF is representative only at the national level, we multiply the fraction of salaried workers in the workforce, calculated from the PME, by the total size of the workforce, calculated by IBGE, in order to get the total number of salaried workers in the country. The measure $m$ is the ratio of firms to salaried workers at the national level.

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 the real interest rate. We use the matching function elasticity $E$ from Ulyssea (2010) and assume symmetric bargaining, meaning that the bargaining power of workers is set to 0.5. The matching elasticity is borrowed from the literature because we think it would not be identified without strong assumptions, since we match transitions into employment very well without this additional degree of freedom. We also do not believe it is crucial for our results, and the range of values used in the literature is narrow. As for the bargaining power, it is not strongly identified separately from the returns to scale parameter $\alpha$. Both have implications for rent sharing and the pass-through of productivity shocks. We have a cross-sectional moment that can be used to identify one of them given the other: the labor share of income. But we lack firm-level data from which we could identify pass-through, and we also lack exogenous aggregate shocks to productivity that vary across regions. We choose symmetric bargaining because it is a common choice in the literature. Setting the value of the matching efficiency $D$ is a normalization since it is not separately identified from the vacancy-posting cost (which we estimate).

Finally, we impute $\gamma$ from the literature because we do not have a credible identification design for it. We lack an exogenous, strong shifter of either labor supply or relative demand for skills. Trying to estimate $\gamma$ by minimum distance using the procedure described in subsection 4.4.3 would lead to weak identification, since the model is very flexible in how the demand for skills can vary across regions.

Fernandez and Messina (2018) estimate elasticities of substitution using data from Brazil, Argentina, and Chile, using a model with three levels of skill. They estimate the elasticity of substitution between college-educated and unskilled...
Table 2 – Imputed Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Estimated from Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta$ (share of skilled workers, avg across regions)</td>
<td>0.5267</td>
<td>Estimated from PME</td>
</tr>
<tr>
<td>($\bar{\lambda}_{in}^{inf}$ (inf. unskilled job destruction))</td>
<td>0.0850</td>
<td>Estimated from PME</td>
</tr>
<tr>
<td>($\bar{\lambda}_{in}^{inf}$ (inf. skilled job destruction))</td>
<td>0.0103</td>
<td>Estimated from PME</td>
</tr>
<tr>
<td>($\bar{\lambda}_{for}^{inf}$ (formal unskilled job destruction))</td>
<td>0.0316</td>
<td>Estimated from PME</td>
</tr>
<tr>
<td>($\bar{\lambda}_{for}^{inf}$ (formal skilled job destruction))</td>
<td>0.0038</td>
<td>Estimated from PME</td>
</tr>
<tr>
<td>$m$ (measure of firms)</td>
<td>0.0905</td>
<td>Calculated from PME, CEMPRE, and ENCINF</td>
</tr>
<tr>
<td><strong>Panel B: Calculated from Labor Legislation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$ (payroll tax rate)</td>
<td>0.7206</td>
<td>Appendix A</td>
</tr>
<tr>
<td>$a_s, a_u$ (variable benefits)</td>
<td>0.266, 0.318</td>
<td>Appendix A</td>
</tr>
<tr>
<td>$b_F^s, b_F^u$ (fixed benefits)</td>
<td>0.02, 0.05</td>
<td>Appendix A</td>
</tr>
<tr>
<td>$b_D^s, b_D^u$ (unemp. insurance)</td>
<td>7.48, 4.00</td>
<td>Appendix A</td>
</tr>
<tr>
<td><strong>Panel C: Taken from the Literature and Others</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$ (discount rate)</td>
<td>0.008</td>
<td>Real interest rate</td>
</tr>
<tr>
<td>$\sigma$ (worker bargaining power)</td>
<td>0.5</td>
<td>Assumed symmetric</td>
</tr>
<tr>
<td>$D$ (matching scale)</td>
<td>0.3</td>
<td>Normalization, value from Ulyssea (2010)</td>
</tr>
<tr>
<td>$E$ (matching elasticity)</td>
<td>0.5</td>
<td>Ulyssea (2010)</td>
</tr>
<tr>
<td>$\gamma$ (elasticity of substitution = $\frac{1}{1+\gamma}$)</td>
<td>0.436</td>
<td>Fernandez and Messina (2018); other values in robustness tests</td>
</tr>
</tbody>
</table>

Note: For the parameters estimated with data from the PME, presented in the first five rows, we show standard errors in parentheses.

workers to be 1.25, which corresponds to $\gamma = 0.200$, and between high school- and primary-educated workers to be 2.3, which corresponds to $\gamma = 0.566$. Our definition of skilled workers includes both high school- and college-educated workers, and thus we use the average between these two values. This corresponds to an elasticity of 1.78, and $\gamma = 0.437$. As a robustness test, we re-estimate the model and generate the main counterfactuals using the two values estimated by Fernandez and Messina (2018), encompassing a broad range of elasticities of substitution. As discussed below, with the exception of the quantitative implications of one policy experiment, our main conclusions are not sensitive to this choice.\footnote{The intermediate value used in our benchmark leads to a marginally better fit than the other two values. Our benchmark value is also close to values typically used for the elasticity of substitution between high school- and college-educated workers in the US, as discussed in Acemoglu and Autor (2011).}

### 4.4 Estimation Strategy

This section provides an overview of our estimation strategy. It has three steps. First, we extract moments at the region-skill level from the PME. Second, we extract three national-level moments from other data sets. And third, we estimate the model with a minimum distance procedure, imputing the parameters from Table 2 and using moments obtained in steps 1 and 2 as targets.

Below, we present an overview of each of these steps and explain our use of bootstrap for inference. For further details, see Appendix F.
4.4.1 PME Moments at the Region-Skill Level

In our first step, we obtain the following moments at the region-skill level: informality rates, mean log hourly wages, formal wage premium, firm-size wage premiums (for firms with 6-10 and 11 or more workers, relative to the excluded category with 1-5 workers), job finding rates (transitions from unemployment to any kind of employment), and distribution of workers across firm sizes in the formal and informal sectors (shares of formal and informal workers in firms with 6 to 10 employees, and 11 or more employees). Most of them are calculated by taking averages of variables at the individual level (with survey weights) and using the composition of skills by age-education groups discussed in subsection 4.2.

The wage premiums are obtained from panel regressions run separately for each region-age-education group. The dependent variable is the log hourly wage and the controls are individual fixed effects, a linear term in age, and time fixed effects (calendar months in 2003). For the formal wage premium, the main independent variable is formality status. Firm-size wage premiums are obtained from similar regressions that control for firm size categories. In these regressions, the individual fixed effects capture unobserved heterogeneity in ability within age-education group, which might otherwise bias the estimates. In the text, we refer to these wage premiums as unconditional, meaning that, for example, the moments related to firm-size wage premiums are not conditional on the formal wage premium, and vice-versa. Wage premiums are aggregated from age-education groups to skills using the definition of skills discussed before.

4.4.2 Additional National Moments

We use three additional moments, obtained from other data sets, as targets. The first one is the labor share in national income, calculated from the National Accounts System (applying the corrections proposed by Gollin, 2002). The labor share is helpful for identifying the returns to scale parameter of the production function, $\alpha$. In the model, we define the labor share as the fraction of total production (net of search costs and informality penalties) that is not firm profits nor government surplus.

The second moment is the causal effect of enforcement of labor regulations on the informality rate. To obtain this estimate, we use insights and data from Almeida and Carneiro (2012). Municipality-level enforcement is measured as the share of the local workforce that is targeted by inspectors. Almeida and Carneiro (2012) show that enforcement is strongly predicted by the interaction of two variables: the number of labor inspectors based in the closest office of the Ministry of Labor and the travel distance (in hours) between that office and the municipality. This is true even conditionally on a broad set of controls. Based on their discussion, we run an instrumental variables regression to find the effect of enforcement on the municipality-level informality rate, using the interaction as the instrument and the same set of controls that they use (the data they used are available from Almeida and Carneiro, 2019).

This quasi-experimental estimate helps us identify a key parameter and lends credibility to the counterfactual exercises. The exponent $P^E$ from the $\rho(.)$ function determines how the informality penalty grows with firm size and, hence, the relative slope of formal profits versus informal profits as functions of $z$. It thus governs how many firms change their compliance decisions following shocks that change relative profits in the informal sector. This is why it can be identified from the effect of enforcement on informality rates. This is also the reason why $P^E$ is crucial for other comparative static exercises, such as changes in the minimum wage.
The third national-level target is the ratio of the number of workers employed in formal firms with more than 500 employees to those in formal firms with more than 100 employees (calculated from the Ministry of Labor’s registry of formal workers, the Relação Anual de Informações Sociais, or RAIS). Because the categories of firm size in the PME are censored at 10 employees, this additional target provides helpful information about the right tail of the firm size distribution, imposing restrictions on the distribution of firm productivity in the model.

### 4.4.3 Minimum Distance Estimation

The remaining parameters of the model are estimated by minimizing the distance between the 123 moments collected in steps 1 and 2 and their corresponding values implied by the model. We define a function $h : X \rightarrow \mathbb{R}^{123}$ which takes a set of parameters $x \in X \subset \mathbb{R}^{39}$, finds the equilibrium of the model for each of the six regions, and simulates the corresponding moments by integrating over the distribution of firms. The three national moments are weighted averages of the values for the six regions (using region workforce as weight).

The estimation procedure finds the parameters $x$ that minimize a weighted sum of squared relative errors, using a gradient descent algorithm and parallel processing to analyze a large number of starting points. Weights are proportional to the share of the workforce from which each moment was calculated. See Appendix F for a detailed description of the loss function, including the weights. The appendix also discusses the computational implementation of the estimator, including optimization choices.

Estimated parameters are shown in Table 3. For conciseness, the table presents national averages of region-specific parameters weighted by workforce size. The flow utility of unemployment is small for unskilled workers and strongly negative for skilled workers. Small or negative values of leisure are necessary to match realistic levels of wage dispersion for similar workers in matching models, as discussed in Hornstein, Krusell and Violante (2011). Vacancy-posting costs are high in the skilled labor market, but negligible for informal firms. Thus, match rents are very small in the informal sector, and workers are close to indifferent between unemployment and informal employment from the point of view of the model. While informality provides a higher flow utility, being unemployed is valuable because it offers the possibility of finding a formal job.

Most other parameters are difficult to interpret directly. Instead, we prefer to discuss the properties of the implied equilibrium. Before that, though, we close the estimation subsection by discussing quality of fit and our bootstrap procedure.

Table 4 shows that the estimated model matches all target variables with considerable accuracy. Here again, for conciseness, we show national averages weighted by population. We also show averages across skilled and unskilled workers for all variables.

The model fits the data well across most margins: informality, unemployment, wages, distribution of workers in the formal and informal sectors by firm size, and the aggregate variables related to the elasticity of the enforcement function, the labor share in production, and the distribution of workers across very large firms. The formal wage premium for unskilled workers and the firm-size wage premium for firms with more than 10 employees are also matched with some precision.

The main difficulty of the model is the formal wage premium for skilled workers and the firm-size wage premium for
firms with 6 to 10 employees. These two dimensions are interrelated. The model is unable to reproduce sufficiently large firm size premiums for firms in the middle range of the distribution, probably due to parametric restrictions on the distributions of $z$, $k$, and vacancy-posting costs. These restrictions, in turn, affect its capacity to reproduce the unconditional formal wage premium for skilled workers. Nevertheless, even along these dimensions, the model reproduces the qualitative patterns observed in the data: the formal wage premium is decreasing in skill level, and the firm-size wage premium is increasing in firm size and skill level.

Appendix Figure A.9 shows the quality of fit across regions. The model performs exceptionally well in the informality and unemployment dimensions, and reasonably well in log wages and wage premiums. Shares of formal and informal workers by firm size categories are not particularly well matched; on average, the model does no better than simply guessing the national means. But given that the model is severely over-identified (more than three times as many targets as parameters), we believe the overall fit is satisfactory, with particularly good fits along the dimensions that constitute our main interest and that are most important for the counterfactuals.

4.4.4 Bootstrap

Our bootstrap fully accounts for the multi-step nature of the estimation procedure. In each bootstrap replication, we start by resampling the PME data (drawing individuals within region, keeping their individual histories in the panel intact) and repeating all steps from the beginning (including the definition of skill levels). When generating the aggregate moments, we also re-sample the Almeida and Carneiro (2012) data (drawing municipalities within state) and use a parametric assumption to resample the labor share, based on an estimate of its standard error. We do not have an estimate of the standard error of the tail parameter associated with formal workers in large firms. We believe the sampling variation in this estimate is negligible compared to our other targets, since this information comes from administrative data on the universe of formal firms. Nevertheless, in order to be conservative, we assume a standard error of 0.02 for this variable in the bootstrap simulations. Finally, we reestimate the remaining parameters using the new set of moments. The only difference is that we use a starting point close to the point estimates, instead of attempting several starting points. We use 100 bootstrap repetitions.

Table 3 shows the standard errors of the estimated parameters. Most of them are small when compared to the point estimates. However, some are relatively large, particularly the standard errors for the search cost parameters. To assess whether this uncertainty limits the meaningful use of the model, we also provide standard errors for our main counterfactual exercises, by recalculating them with the reestimated model in each bootstrap replication (see Appendix J.1). As discussed in detail in the next section, we find that the counterfactuals are precisely estimated.

4.5 Properties of the Estimated Equilibrium

In order to illustrate the richness of the equilibrium generated by the model, we present in Figure 1 the cross-sectional profiles of wages, firm size, and employment composition across the distribution of firm productivity ($z$). We present the case of the Recife metropolitan region to highlight some properties of the equilibrium, and the same figures for the other five metropolitan regions are available in Appendix J.4. The solid, darker line in the figures corresponds to formal firms. The dashed gray lines correspond to informal firms with different costs of informality, indicated by $k = 1$, $k = 2$, and $k = 3$. Firms of enforcement type 4 and 5 (higher informality costs) are formal for all levels of $z$ in all regions.

The figure summarizes the main features of the model. Wages of skilled workers are systematically higher than wages
Figure 1 – Cross-Sectional Features of the Equilibrium, Recife Metropolitan Region, 2003

Note: The top four panels show firm choices in equilibrium for different values of firm productivity $z$ and informality cost shifter $k$. Conditional on being formal, $k$ is irrelevant, and thus there is a single curve for formal firms in each panel. Different curves for informal firms reflect different values of $k$. These curves do not cover the whole range of productivities because, when $z$ is high enough, firms choose to comply with regulations. Because firms with $k = 4$ and $k = 5$ are formal for all values of $z$, there are only three such curves. The bottom two panels show distributions of firms and total employment by firm productivity $z$. They also show shares of firms of a given $z$, or shares of workers employed in those firms, that are informal.
Table 3 – Estimated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Production Function</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \bar{A} ) (productivity shifter, avg across regions)</td>
<td>5.0272</td>
<td>0.3433</td>
</tr>
<tr>
<td>( B_1 ) (skill bias shifter)</td>
<td>0.2817</td>
<td>0.1509</td>
</tr>
<tr>
<td>( B_2 ) (relative skill bias in productive firms)</td>
<td>0.2900</td>
<td>0.0417</td>
</tr>
<tr>
<td>( \alpha ) (decreasing returns)</td>
<td>0.5356</td>
<td>0.0190</td>
</tr>
<tr>
<td><strong>Distribution of Firms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \bar{T} ) (tail parameter, avg across regions)</td>
<td>1.8853</td>
<td>0.0530</td>
</tr>
<tr>
<td><strong>Informality Cost</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \bar{P} ) (enforcement shifter, avg across regions)</td>
<td>0.7548</td>
<td>0.2647</td>
</tr>
<tr>
<td>( P^D ) (enforcement penalty dispersion)</td>
<td>3.6169</td>
<td>0.1528</td>
</tr>
<tr>
<td>( P^E ) (firm size exponent)</td>
<td>2.5287</td>
<td>0.2023</td>
</tr>
<tr>
<td><strong>Vacancy-Posting Cost</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \xi^f ) (formal sector component)</td>
<td>548.65</td>
<td>100.11</td>
</tr>
<tr>
<td>( \xi^{inf} ) (informal sector component)</td>
<td>0.0054</td>
<td>0.0002</td>
</tr>
<tr>
<td>( \xi^{sk} ) (skilled worker shifter)</td>
<td>3.5625</td>
<td>5.8927</td>
</tr>
<tr>
<td>( \xi^{st} ) (steepness)</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Utility Flow while Unemployed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d_{s} ) (skilled workers)</td>
<td>-8.2053</td>
<td>0.0862</td>
</tr>
<tr>
<td>( d_{u} ) (unskilled workers)</td>
<td>0.1073</td>
<td>0.0934</td>
</tr>
</tbody>
</table>

...of unskilled workers. Nevertheless, one can find skilled and unskilled workers in both the formal and informal sectors, and some unskilled workers earn more than some skilled workers. There is some overlap in the distribution of firms across formal and informal sectors as well, due to heterogeneity in the cost of informality.

The wage panels illustrate the discussion in subsection 3.5. For skilled workers in small firms, wages are higher in the informal sector; the compensating differentials logic dominates because the minimum wage is not binding and match rents are close to zero for small firms. For unskilled workers, wages in small firms are higher in the formal sector because of the binding minimum wage.\(^{25}\) As the productivity parameter \( z \) grows, the firm-size wage premium becomes more relevant, and wages within the formal sector rise for both types of workers. For unskilled workers, there is a kink in the wage profile in the formal sector, corresponding to the point at which the minimum wage ceases to bind. The cross-sectional profile of wages across skill levels and sectors reflects, therefore, three fundamental forces: compensating differentials, the firm-size wage premium, and labor regulations.

Both firm size and the share of skilled workers are rising in the productivity parameter \( z \). More productive firms tend to hire more workers and to choose a higher share of skilled workers. Informal firms with higher values of \( z \) are also larger and relatively more intensive in skilled workers. Given the cost of informality \( k \), there is a level of productivity \( z \) (and a corresponding firm size) after which it is no longer worth it to remain informal. Sufficiently large firms are formal independently of their type, so while there is an overlap between the formal and informal sectors among smaller firms, very large firms are always formal.

Interestingly, low \( z \) firms that are formal tend to be smaller and, at the same time, more skill intensive than similar

\(^{25}\)A binding minimum wage does not necessarily imply that the formal wage premium is positive, as illustrated in equivalent figures for some other regions.
<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Fit in 2003</th>
<th>Validation: Δ2003-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data_{2003}</td>
<td>Model_{2003}</td>
</tr>
<tr>
<td>Informality Rate</td>
<td>0.2927</td>
<td>0.2916</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.1492</td>
<td>0.1530</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.0889</td>
<td>0.0909</td>
</tr>
<tr>
<td>Unemployment (implied)</td>
<td>0.1344</td>
<td>0.1340</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.0311</td>
<td>0.0315</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.2514</td>
<td>0.2500</td>
</tr>
<tr>
<td>Wages (ln)</td>
<td>0.7860</td>
<td>0.8009</td>
</tr>
<tr>
<td>Skilled</td>
<td>1.5484</td>
<td>1.5401</td>
</tr>
<tr>
<td>Unskilled</td>
<td>-0.0286</td>
<td>0.0269</td>
</tr>
<tr>
<td>Formal wage premium</td>
<td>0.0617</td>
<td>0.0101</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.0429</td>
<td>-0.0431</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.0892</td>
<td>0.0096</td>
</tr>
<tr>
<td>Firm-size premium 6-10 workers</td>
<td>0.0494</td>
<td>0.0078</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.0092</td>
<td>0.0030</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.0030</td>
<td>0.0023</td>
</tr>
<tr>
<td>Firm-size premium 11 or more workers</td>
<td>0.0732</td>
<td>0.1030</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.0037</td>
<td>0.1240</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.0011</td>
<td>0.0734</td>
</tr>
<tr>
<td>% formal workers firms 6-10 employees</td>
<td>0.0775</td>
<td>0.0386</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.0079</td>
<td>0.0042</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.1280</td>
<td>0.0705</td>
</tr>
<tr>
<td>% formal workers workers ≥11 employees</td>
<td>0.8552</td>
<td>0.8594</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.8998</td>
<td>0.9226</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.7431</td>
<td>0.7119</td>
</tr>
<tr>
<td>% informal workers firms 6-10 employees</td>
<td>0.1234</td>
<td>0.1585</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.1145</td>
<td>0.1157</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.1287</td>
<td>0.1769</td>
</tr>
<tr>
<td>% informal workers workers ≥11 employees</td>
<td>0.5133</td>
<td>0.5207</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.6797</td>
<td>0.6712</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.4383</td>
<td>0.4334</td>
</tr>
<tr>
<td>Labor share</td>
<td>0.528</td>
<td>0.5157</td>
</tr>
<tr>
<td>Enforcement elasticity of informality</td>
<td>-0.0534</td>
<td>-0.0529</td>
</tr>
<tr>
<td>formal workers in firms ≥ 500</td>
<td>-0.0534</td>
<td>-0.0529</td>
</tr>
<tr>
<td>formal workers in firms ≥ 100</td>
<td>0.7224</td>
<td>0.7595</td>
</tr>
</tbody>
</table>

Note: Wages are in multiples of the minimum wage in 2003, the numeraire in the model. The columns with headers “s.e.” show bootstrap standard errors for the moments in the data. For illustrative purposes, we present unemployment rates implied by the model rather than job-finding rates, since the former is more intuitive. The unemployment rate is a monotonic transformation of job finding rates, holding constant job destruction and informality rates.
informal firms. These are firms with very high informality costs that decide to formalize despite being unproductive. Being formal, they have to pay the binding minimum wage for unskilled workers, which tends to reduce their size and increase their share of skilled workers. Finally, there is a kink in the distribution of the share of skilled workers in formal firms, corresponding as well to the firm size at which the minimum wage ceases to bind. After this point, changes in the number of unskilled workers end up affecting the within-firm bargaining process, so firms switch to skilled labor in a non-smooth fashion at this margin.

The bottom two diagrams in the figure plot the distributions of firms and employment, and their formality statuses, on the productivity parameter $z$ (with the density normalized to 1 for the least productive firms). Consistent with patterns widely documented in the empirical literature, the estimated model displays (i) a very skewed distribution of firms, with lots of very small and unproductive firms and a few very large and productive firms, and (ii) a much less skewed distribution of employment, since the few very large firms employ a substantial fraction of the labor force. The shares of informal firms and employment decrease with firm size and productivity, with kinks due to the discreteness of the distribution of informality costs.

5 Quantitative Exercises

5.1 Validating the Model

Our first quantitative exercise is to use the changes in labor market outcomes between 2003 and 2012 to validate the predictions of the model. We plug into the estimated model the changes observed in the composition of the labor force, minimum wages, mandated benefits, payroll taxes, enforcement of regulations, and aggregate productivity/demand conditions, and we analyze how the predictions of the model, along many dimensions, compare to the changes observed in the data.

In the period we study, the informality rate dropped by 10.5 percentage points, and unemployment fell by 7.5 points (these numbers are slightly different from those in Table 1 because now, to be consistent with the estimation of the model, we are weighting regional averages by workforce size). Average wages increased by 34%, but the gains were not homogeneous: unskilled wages increased by 52%, as compared to 7.7% for skilled workers. The unconditional formal wage premium and the firm-size wage premiums declined, while employment became increasingly concentrated in larger firms, particularly formal ones.

As for the exogenous variables in the model, some also experienced significant changes during this period. There has been a consistent increase in school attendance among Brazilian school-aged children over the last decades, with near universalization of primary schooling achieved by the late 1990s. At the same time, demographic changes associated with historical reductions in fertility and population aging have translated into an older, and therefore more experienced, workforce. This combination manifested, during the 2000s, as a substantial improvement in the skill composition of the workforce. As an example, the fraction of the workforce with 8 or more years of schooling increased from 66% to 78.9% between 2003 and 2012. Given our definition of skills in the model and the changes in the schooling and age distributions observed in the data, the share of skilled workers increased by 12.1 percentage points, from 52.7% to 64.8%.

\[26\text{The share of children 7–14 years old attending school increased from 80.9% in 1980 to 97% in 1999 (de Oliveira, 2007).}\]
Minimum wages increased by 61.2% in real terms, while changes in payroll taxes and benefits were minor. An additional contribution to the worker’s severance payment fund (Fundo de Garantia por Tempo de Serviço, FGTS) was phased out, which we calculate as having reduced the final payroll tax only slightly, from 72.06% of the nominal wage to 71.43%. There were some small changes to the income tax and social security contribution schedules, which also had negligible real effects (see Appendix A for a detailed discussion of these changes and our calculations).

We estimate the change in enforcement of labor regulations using inspections data from the Ministry of Labor. MTE (2013) shows that the share of workers targeted by inspections increased by 33.9% from 2003 to 2012. We use this number as a proxy for enhanced enforcement efforts in the model, interpreting it as a change in the per firm probability of inspection.\(^{27}\)

Brazil also experienced shocks to productivity and prices, such as changes in the terms of trade, that might have affected wage levels and the wage gap between skilled and unskilled workers (see, for example, discussion in Ferreira and Veloso, 2013 and Messina and Silva, 2018). To incorporate these forces in our counterfactual exercise, we infer region-specific changes in productivity and skill bias from observed changes in average log wages by skill level. Specifically, we find values of the parameters \(A\) and \(B_1\) in each region, as of 2012, such that the model-predicted change in average log wages for each skill level matches the change observed in the data (taking into account all other changes occurring during the period). We find that the TFP parameter \(A\) in the model increased, on average, by 24% between 2003 and 2012. This number falls within the range of estimates for the cumulative growth in TFP obtained by Ferreira and Veloso (2013). In turn, \(B_1\) increased in all regions, such that these unmodeled demand shocks helped reduce wage inequality. This also seems plausible, given the discussion in Messina and Silva (2018).\(^{28}\)

The results of our counterfactual exercise, alongside the changes observed in the data, are presented in the last two columns of Table 4. The qualitative implications of the model match the patterns observed in the data along most dimensions considered: reductions in informality, unemployment, formal wage premium, and firm-size wage premiums, reductions in the share of formal and informal workers in firms with 6 to 10 employees, and increases in the share of formal workers in firms with 11 or more employees. Relative changes across skilled and unskilled workers are typically also in line with what we see in the data: reductions in informality, unemployment, the share of workers in firms with 6 to 10 employees, and increases in the share of workers in formal firms with more than 11 employees are all stronger for unskilled workers. Meanwhile, the reduction in the firm-size premium for firms with 6 to 10 workers is larger for skilled workers.

Overall, the model is able to accommodate, across various margins, most of the qualitative changes seen in the data. But some dimensions are missed. In the aggregate, the model is unable to reproduce the increase in the share of workers in informal firms with 11 or more employees. By skill level, it does not predict the relatively larger decline in the unconditional formal wage premium for skilled workers, nor that of the firm-size premium for unskilled workers in firms with 11 or more employees. More detailed data on the distribution of workers across firms of different sizes might help the model better predict some of these margins, but we are constrained by the information available from the PME. Notice that as mentioned before, since we consider unconditional moments, our ability to match the firm-size wage premium for large firms may also have implications for our ability to match the formal wage premium, so these

\(^{27}\)This definition is consistent with the one we used when estimating the effect of enforcement on the informality rate, which was one of the targets in the estimation of the model. Other indicators, such as total revenues from fines, also increased during the period. For a thorough discussion of enforcement of labor regulations in Brazil, see Cardoso and Lage (2005).

\(^{28}\)See Appendix Figure A.8 for a plot of the \(B(z)\) function in both periods.
dimensions are likely to be interrelated.

As an even stronger out-of-sample test of the model, we present the regional variation of the counterfactual results in Appendix Figure A.10. The figure plots the changes predicted by the model against the data, summarized as averages and differences across skill levels, for the six metropolitan regions covered by the PME dataset. There is a clear positive correlation between the predictions of the model and the observed changes in informality—our main outcome of interest. The same is true for unemployment. The performance in terms of regional variation seems to fall short particularly in relation to the changes in the distribution of workers across firm sizes (which are also the targets with the worst fit in 2003). Since these targets are not the main focus of our analysis and are only auxiliary moments used to allow the model to accommodate wage dispersion within skill levels and sectors, we do not see this pattern as a significant limitation.

Quantitatively, the model explains 57% of the reduction in informality observed in the data and 59% of that in unemployment. In addition, it overpredicts the reduction in the formal wage premium and underpredicts the reductions in the firm-size wage premiums.

### 5.2 Understanding the 2003–2012 Reduction in Informality in Brazil

In this subsection, we illustrate the use of the model for positive analysis by discussing the main driving forces behind the reductions in informality observed in Brazil between 2003 and 2012. Table 5 addresses this issue by reproducing the counterfactual exercise conducted before but holding each of the different exogenous variables constant at their 2003 values, one at a time. Each column in the table presents the results for a different counterfactual exercise, and each row contains a particular labor market outcome. Since the impact of changes in benefits and taxes was negligible, we save space by omitting them from the table. We focus on the main outcomes of interest and, for purposes of comparison, reproduce in column 1 the results obtained in the complete counterfactual exercise from Table 4.

There are two particularly striking results in Table 5. First, the model predicts that without an increase in the share of skilled workers (column 2), informality would have gone up by 3.5 percentage points instead of declining by 6. Unemployment would have increased by 1.5 percentage points instead of declining by 4.4. In this scenario, the formal wage premium would also have increased by 13% instead of declining by 4%. In short, the model is unable to reproduce any reduction at all in informality and unemployment when changes in labor force composition are ignored.

Second, the declines in both informality and unemployment would have been considerably larger—by 2.3 and 2.9 percentage points, respectively—if the minimum wage had not increased (column 3). In this scenario, the wages of unskilled workers would have increased by 48% instead of 51%, and the formal wage premium would have declined by an additional 5.6%.

Enforcement, in turn, is less relevant: without increased enforcement, the decline in informality would have been 1.3 percentage points smaller, but other dimensions of the labor market would have remained largely unchanged (column 4).

Finally, the parameters capturing TFP and aggregate demand conditions also play an important role (columns 5 and 6). In particular, without increases in A, the model predicts a reduction of 19% in the wages of skilled workers, rather than an increase of 22%. The increase in TFP is necessary to counteract the increased relative supply of skilled workers so...
Table 5 – Counterfactuals Holding Constant Each Factor One at a Time, Changes in Brazilian Labor Market Outcomes from 2003 to 2012

<table>
<thead>
<tr>
<th>Outcomes:</th>
<th>(1) All changes</th>
<th>(2) Fraction skilled</th>
<th>(3) Minimum wage</th>
<th>(4) Enforcement</th>
<th>(5) Productivity/Demand</th>
<th>(6) $B_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informality</td>
<td>-0.0603</td>
<td>0.0342</td>
<td>-0.0840</td>
<td>-0.0473</td>
<td>-0.0154</td>
<td>-0.0261</td>
</tr>
<tr>
<td>Unemployment</td>
<td>-0.0434</td>
<td>0.0132</td>
<td>-0.0730</td>
<td>-0.0455</td>
<td>-0.0136</td>
<td>-0.0247</td>
</tr>
<tr>
<td>Wages (ln)</td>
<td>0.3437</td>
<td>0.2620</td>
<td>0.3112</td>
<td>0.3436</td>
<td>0.1013</td>
<td>0.3421</td>
</tr>
<tr>
<td>Skilled</td>
<td>0.0774</td>
<td>0.2242</td>
<td>0.0756</td>
<td>0.0768</td>
<td>-0.1931</td>
<td>0.1167</td>
</tr>
<tr>
<td>Unskilled</td>
<td>0.5152</td>
<td>0.2933</td>
<td>0.4778</td>
<td>0.5202</td>
<td>0.2751</td>
<td>0.3752</td>
</tr>
<tr>
<td>Formal Wage Premium</td>
<td>-0.0353</td>
<td>0.1270</td>
<td>-0.0947</td>
<td>-0.0348</td>
<td>0.0873</td>
<td>0.0322</td>
</tr>
<tr>
<td>Firm Size Premium (6 – 10)</td>
<td>-0.0065</td>
<td>-0.0091</td>
<td>-0.0167</td>
<td>-0.0057</td>
<td>-0.0042</td>
<td>-0.0027</td>
</tr>
<tr>
<td>Firm Size Premium (≥ 11)</td>
<td>-0.0017</td>
<td>0.0124</td>
<td>0.0096</td>
<td>-0.0025</td>
<td>0.0334</td>
<td>0.0107</td>
</tr>
</tbody>
</table>

Note: Column 1 shows 2003-2012 changes predicted by our validation exercise (i.e., the same values from Table 4, Column 5). The remaining columns show changes predicted by the model if all of the parameters changed in the same way as in the validation exercise, but for one parameter (which is held constant at the 2003 level).

that the model can generate a net positive change in wages for both skill groups. $B_1$ captures an additional increase in the demand for unskilled labor, most likely driven by changes in aggregate domestic demand conditions and terms of trade (see the discussion in Messina and Silva, 2018). Nevertheless, even ignoring the productivity/demand parameters, the model is still able to generate reductions in informality and unemployment, though of much smaller magnitudes.

Overall, the composition of the labor force and these "productivity" parameters are, together with the minimum wage, the main driving forces behind the changes observed in the Brazilian labor market during this period. "Productivity" parameters appear as key in explaining changes in wage levels, while skill composition is essential to rationalizing the reductions in informality and unemployment rates.

Given that skill composition does not feature prominently in previous informality literature, it is worth spending some time discussing why it plays such an important role in our framework. In the model, increases in the relative supply of skilled workers make their labor market less tight (with the reverse taking place for unskilled workers). Because firms hire more skilled workers in the new equilibrium, the productivity of unskilled workers increases due to complementarities in the production function.

These forces end up affecting informality 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 sector to the formal. 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 occurs. 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).

The increase in unskilled wages that accompanies this process makes the minimum wage less binding—or, in the case of Brazil from 2003 to 2012, it neutralizes much of the rise in the minimum wage. Along with increases in unskilled wages due to changes in $A$ and $B_1$, this explains why minimum wages did not lead to increased informality during this period.
To strengthen our argument and show that changes in workforce composition are strictly necessary to replicate the patterns observed in the data, we conduct three additional exercises. The first two examine the plausibility of alternative explanations that do not rely on changes in the composition of the labor force. The third is an empirical exercise using census data, a broader sample, and a longer time horizon to explore changes in the composition of the labor force in local labor markets in Brazil.

In the first exercise, we consider a scenario where the fall in informality was driven by productivity and enforcement of regulations, without any change in skill composition. We estimate changes in $A$, $B_1$, and $P_r$ by region, such that changes in informality and wages by skill level are the same as the ones in the baseline model. Results are available in Appendix Table A.19. This scenario is implausible for two reasons. First, required changes in $P_r$ imply increases, on average, of 127% in enforcement between 2003 and 2012. No estimate currently available suggests a change in enforcement of similar magnitude. Second, under this scenario, the reduction in unemployment is much smaller, corresponding to less than a third of that in column 1 in Table 5 and less than a fifth of that seen in the data. This is consistent with the common view that enforcement can represent a trade-off between employment and formalization.

The second exercise considers demand-side shocks that could have affected firm entry and exit. We simulate another counterfactual where the share of skilled workers remains the same, but $T_r$—the shape parameter of the distribution of firm productivity $z$—is allowed to change along with $A_r$ and $B_{1r}$. As above, the new parameters are selected to match the changes in mean log wages by skill level and informality in the baseline model. This exercise is meant to capture alternative explanations in which the fall in informality is driven not by skill composition, but by demand shocks that lead to firm exit in some parts of the productivity distribution while fostering the growth of other types of firms. One example is terms-of-trade shocks that benefit particular industries, as in Melitz (2003). In addition to impacting factor prices through a Heckscher–Ohlin mechanism (via $B_1$), such shocks could benefit larger firms capable of exporting and hurt others through import competition. We use $T_r$ to proxy for such effects because we do not explicitly model entry and exit decisions.

The results from this exercise, presented in Appendix Table A.19, are at odds with the data along many important dimensions. Informality falls homogeneously across skilled and unskilled workers, unemployment remains roughly constant, the formal wage premium increases slightly (and substantially for skilled workers), and the firm-size wage premium increases for larger firms. Additionally, in order to match the observed change in relative wages, this counterfactual exercise estimates an increase in $B_1$ of over 1,100%, which also seems implausible.29

The third exercise, discussed in detail in Appendix H, presents some preliminary evidence of the relationship between labor force composition and informality rates. Since there is no reduced-form evidence currently available on this point, we examine whether local labor markets that experience faster gains in schooling also see larger declines in informality rates. We use census data for the whole country from 1991, 2000, and 2010 (IBGE, 1992, 2000, 2011). Unsurprisingly, we find that average schooling at the local level is positively correlated with formality. But, most importantly, we find that this relationship holds even when conditioning on individual-level schooling and a large array of controls, including local labor market fixed effects, time fixed effects, other individual characteristics, and a battery of variables that proxy for location-specific demand and institutions. This result supports the model’s prediction that a more skilled workforce leads to higher formality rates, even conditioning on a worker’s skill level. We do not have a

29 Appendix Figure A.8 shows that, in this scenario, $B(z) < 0.5$ for the majority of firms. This means that if at least half of the workers in these firms are skilled, skilled workers are actually less productive than unskilled ones, which again seems implausible.
clear identification for variations in average schooling levels, so by itself, this exercise is not definitive evidence of the role of education. It is, however, broadly consistent with the results of the structural model, lending further support to our conclusions.

Overall, these additional exercises do not rule out the possibility that demand shocks could have affected informality through firm entry or exit, nor the chance that our proxy for enforcement underestimates the importance of this factor. Indeed, our model does not fully explain the increase in formalization observed from 2003 to 2012. But these exercises do show that explanations that do not prominently feature skill composition are either incomplete or implausible.

Appendix G discusses the comparative statics properties of the model, changing each of the main exogenous variables one at a time (starting from the 2003 values). This additional exercise sheds some more light on the role of each factor, and it validates the performance of the model by confronting the comparative statics results with the empirical evidence from reduced-form estimates. The results reinforce the general conclusion of this subsection: the change in the skill composition is the most important determinant of informality, and its interactions with the minimum wage and technological/demand parameters are essential to rationalizing the changes in unemployment and wages. The fact that no exogenous variable in isolation is able to generate all the patterns observed in the data is particularly important. It highlights the value of an equilibrium model that incorporates the interactions between multiple potential explanations.

5.3 Policy Experiments

A major policy challenge in developing countries is how to bring down informality without increasing unemployment. 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 normative policy analysis.

The first policy we consider is a reduction in payroll taxes for low-wage workers. Lower payroll tax rates can lead to a decline in informality with no adverse effect on unemployment, but they can also substantially reduce government revenues. However, informal firms are relatively more intensive in unskilled labor. In addition, only a fraction of government revenues come from payroll taxes on unskilled workers, since their wages are lower and they account for a smaller fraction of formal employment. Thus, an intermediate alternative in which governments subsidize the employment of low-wage formal workers through a progressive payroll tax may have a larger impact on informality while simultaneously minimizing the implications for government revenue. Proposals like this have been considered as ways to subsidize low-wage workers in developed countries for different theoretical reasons (see Cheron, Hairault and Langot, 2008; Lee and Saez, 2012; Robin and Roux, 2002), but they rarely feature in the informality discussion in the developing world.

In Table 6, we examine the progressive payroll tax policy, starting from the model as of 2012. In the first column, we show as a reference point the results of simply reducing the overall payroll tax rate by 1 percentage point (to 0.7043). This change leads only to a minor reduction in informality (0.27 p.p.) but has a nontrivial cost in terms of government revenue (a surplus reduction of 1.75%).

In columns 2 and 3, we implement progressive payroll taxes. In column 2, the cut in payroll taxes is the same as in column 1 (1 p.p.) but is restricted to unskilled workers.\(^{30}\) The reduction in informality is roughly 50% of that in column

\(^{30}\text{In the model, for simplicity, the payroll tax is based on skill rather than on wage. Our intention is to capture a policy targeting workers earning}\)
Table 6 – Hypothetical Policy Experiments

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>(1) 1 p.p. reduction in payroll tax</th>
<th>(2) Progressive payroll tax $\Delta \tau_s = 0$</th>
<th>(3) $\Delta \tau_u = -0.10$</th>
<th>(4) Doubling transfer to unskilled $\Delta b_F^u = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informality</td>
<td>$\Delta \tau = -0.01$</td>
<td>$\Delta \tau_u = -0.01$</td>
<td>$\Delta \tau_s = -0.01$</td>
<td>$-0.013$</td>
</tr>
<tr>
<td>Skilled</td>
<td>$-0.0017$</td>
<td>$-0.0025$</td>
<td>$-0.0003$</td>
<td>$-0.0003$</td>
</tr>
<tr>
<td>Unskilled</td>
<td>$-0.0051$</td>
<td>$-0.0144$</td>
<td>$-0.0060$</td>
<td>$-0.0060$</td>
</tr>
<tr>
<td>Unemployment</td>
<td>$-0.0004$</td>
<td>$-0.0038$</td>
<td>$0.0027$</td>
<td>$0.0027$</td>
</tr>
<tr>
<td>Skilled</td>
<td>$0.0000$</td>
<td>$0.0001$</td>
<td>$0.0000$</td>
<td>$0.0000$</td>
</tr>
<tr>
<td>Unskilled</td>
<td>$-0.0010$</td>
<td>$-0.0103$</td>
<td>$0.0079$</td>
<td>$0.0079$</td>
</tr>
<tr>
<td>Wages (log)</td>
<td>$0.0042$</td>
<td>$0.0065$</td>
<td>$0.0021$</td>
<td>$0.0021$</td>
</tr>
<tr>
<td>Skilled</td>
<td>$0.0051$</td>
<td>$0.0013$</td>
<td>$0.0003$</td>
<td>$0.0003$</td>
</tr>
<tr>
<td>Unskilled</td>
<td>$0.0030$</td>
<td>$0.0251$</td>
<td>$0.0040$</td>
<td>$0.0040$</td>
</tr>
<tr>
<td>Govt Surplus (%)</td>
<td>$-1.75$</td>
<td>$-1.83$</td>
<td>$-1.16$</td>
<td>$-1.16$</td>
</tr>
<tr>
<td>Net Output (%)</td>
<td>$0.06$</td>
<td>$0.08$</td>
<td>$-0.14$</td>
<td>$-0.14$</td>
</tr>
<tr>
<td>Labor Share</td>
<td>$0.0005$</td>
<td>$0.0016$</td>
<td>$-0.0008$</td>
<td>$-0.0008$</td>
</tr>
</tbody>
</table>

Notes: Each column shows predicted changes following the introduction of a specific labor market policy. The exercise uses model parameters as of 2012 (with $\tau = 0.7143$ and $b_F^u = 0.05$). Government surplus is total payroll tax payments minus benefit transfers. Net output is total production net of search costs and the informality penalty, or, equivalently, the sum of wages, firm profits, and government surplus. Labor share is the sum of wages divided by the sum of wages and firm profits.

1 and the reduction in government surplus is only 10%. In column 3, we scale up the policy to a 10-percentage-point reduction in payroll taxes for low-wage workers. The budgetary implications are now close to the policy from column 1. The reduction in informality becomes substantial: 1.39 percentage points, corresponding to five times that seen in column 1. Unemployment is reduced by very little, 0.38 percentage point, but still 10 times more than in column 1.

Progressive payroll taxes achieve unequivocally better results for unemployment and formalization while minimizing impacts on government revenue. Lower taxes among unskilled workers induce marginal firms to comply, enlarging the tax base. Taxes raised from skilled workers in firms that formalize help offset part of the foregone revenue from low-skill workers in infra-marginal firms. The policy is clearly beneficial for unskilled workers, lowering their unemployment and increasing average wages. Thus, progressive payroll taxes can also help alleviate poverty and income inequality.

Next, we consider an increase in transfers from the government to low-wage formal workers. This policy subsidizes formal employment and could, in principle, be cost-effective, since it also targets workers at risk of being informal. We model it as increasing the benefit parameter $b_F^u$ from 0.05 to 0.10. The corresponding policy we have in mind for Brazil is the doubling of Abono Salarial, a transfer to formal low-wage workers.

We find that such transfers are ineffective due to the minimum wage. Column 4 shows that the reduction in informality is minimal (0.13 p.p.) despite the sizable costs incurred by the government (a surplus reduction of 1.16%). If the

---

31This quantitative result is the only main result in the paper that is substantially affected by our choice of the parameter $\gamma$. The lower value of $\gamma$, corresponding to a higher elasticity of substitution, leads to a reduction in informality of only 0.78 percentage points, while the high $\gamma$, corresponding to a lower elasticity of substitution, leads to a 1.8-percentage-point reduction in informality. Nevertheless, the qualitative discussions from the text still hold under both these alternative scenarios.
minimum wage did not bind, formal unskilled wages would drop after the increase in benefits because of rent sharing between workers and firms. Lower labor costs, in turn, would incentivize the creation of formal unskilled vacancies. With a binding minimum wage, wages cannot adjust downward, and those incentives are eliminated. The impact of the policy on informality is tiny. It is driven by an increase in the reservation wage of unskilled workers and, hence, in labor costs for informal firms. Workers raise their reservation wage because they are willing to wait longer to get the improved formal jobs, a factor that also leads to higher unemployment.

We make two caveats regarding our progressive payroll tax results. First, we assume that a firm’s compliance decision applies to all of its workers. If firms are free to formalize only part of their workforce, then the policy could affect only low-wage workers in marginal firms, leaving their high-wage workers informal. Second, progressive taxation would increase incentives to underreport wages. This possibility is not taken into account in our exercises because we assume payroll taxes are based on skill levels, not wages.

We nevertheless believe that these concerns are not enough to compromise the qualitative implications of the analysis, though the quantitative results from Table 6 should not be taken at face value. On the first 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 employers to court. The data support the view that most firms hire all of their workers either formally or informally. Among firms in the ECINF dataset with five employees, 32% hire all workers informally while 46% hire all of them formally. Only 22% of firms have both formal and informal employees. This number is even lower for smaller firms.

On the second point, we argue that changes in incentives for underreporting introduced by progressive taxation would most likely be inconsequential. First, there are already strong incentives for firms to underreport wages under current labor law because several contributions and taxes are proportional to earnings (see Appendix A). Thus, changes in incentives would be far from dramatic. Second, the value of many mandated benefits is also indexed by the contractual wage, so workers have an incentive to enforce truthful reporting by firms.

The conclusions of this subsection are broadly supported by recent evidence from a natural experiment in Colombia. Kugler, Kugler and Prada (2017) exploit a tax reform introduced in 2012 that reduced payroll taxes by 13.5% for workers earning less than 10 times the minimum wage. Using a difference-in-differences strategy, they show that the reform led to increases in formal employment for the affected group. Their results suggest an elasticity of formal employment in relation to payroll taxes between -0.26 and -0.63. Our number in Table 6, column 3 corresponds to an elasticity of -0.41 for unskilled workers, falling well within the range of their estimates. They also find stronger effects for smaller firms that pay lower wages, giving support to the mechanism discussed above. But, since they rely on a difference-in-differences strategy, they cannot estimate the impact of the policy on, for example, aggregate informality or government revenue. This highlights once again the relevance of using an equilibrium model to complement the evidence obtained from reduced-form estimates.

5.4 Implications for the Debate on Wage Inequality in Brazil

Wage inequality fell substantially from the early 1990s to the early 2010s in Brazil. Papers studying this decline consider multiple potential causes, including the supply of skills, trade shocks, and the role of labor market frictions. In a recent review of the literature, Firpo and Portella (2019) conclude that the primary factor behind the decline in
inequality was a reduction in the returns to experience and education. In turn, they attribute most of the closing of these wage gaps between skill levels to changes in the supply of skills. Additional factors, such as trade shocks, also helped reduce the wage gaps. There were reductions in within-group inequality as well, driven by minimum wages and reduced cross-firm wage dispersion for similar workers. In this final subsection, we discuss what our framework has to say regarding the debate on the recent reduction in wage inequality in Brazil.

The results in Table 5 are in broad agreement with the conclusions of Firpo and Portella (2019). In our model, workforce composition is the main factor explaining the 44-log-point reduction in the mean log wage gap between skilled and unskilled workers. Shocks to skill-specific demand, modeled via region-specific changes in $B_1$, are also important. Other factors do not affect the mean log-wage gap by more than 5 log points.

Firpo and Portella (2019) list three questions that remain unanswered in the literature. The first is the cause for the reduction in returns to experience relative to returns to education. The second is the role of technical change. And the third is the reduction in cross-firm wage dispersion for similar workers, a phenomenon documented using matched employer-employee data for the formal sector.

While our model is silent about the first two questions, it is uniquely positioned to answer the third. This is so because the model includes a broad array of elements that are important in the Brazilian context: workforce composition, the minimum wage, changes in skill-specific demand, and informality. Crucially, it also includes convex vacancy-posting costs, which generate firm-size wage premiums among identical workers in the formal sector. In Table 4, we show that firm-size wage premiums declined in the data, in accordance with the reduction in cross-firm wage differentials documented in the literature. Even though we do not target this moment in our validation exercise, the model accounts for 60% of the decline in the firm-size wage premium for skilled workers, and 40% for unskilled workers.

We can turn to our counterfactuals to figure out which factors are more closely related to the changes in firm-size wage premiums. We find that the observed decline is due to a combination of factors, not always working in the same direction. For skilled workers, the primary causes are productivity and skill composition. For unskilled workers, the minimum wage is also important. In counterfactuals excluding changes in productivity or workforce composition, cross-firm wage dispersion for both worker types would have increased rather than declined. In another counterfactual without changes in the minimum wage, the firm size wage premium would have increased for unskilled workers, but it would have fallen even more than in the data for skilled workers.

Finally, our model shows the importance of taking into account the interactions between multiple factors in a structural model. These interactions are evident in the role of the minimum wage. If we consider all changes happening in Brazil from 2003 through 2012, the minimum wage helps reduce the mean log wage gap between skill levels by 4 log points. Instead, if all other parameters remain at their 2003 levels, then raising the minimum wage widens the wage gap by 12 log points. This is because, in the former scenario, improvements in productivity and workforce composition make the

---

32 We cannot address the first question because our definition of skill does not distinguish between experience and education. On the second, our modeling of demand changes is stylized via production function parameters; so, we cannot distinguish between technical change and other drivers of changes in skill-specific demand, such as trade.

33 We focus on the premium for firms with 11 or more workers because it is empirically more relevant (given that the share of workers in firms with 6 to 10 employees is small) and because it is the one the model fits best.

34 Engbom and Moser (2018) find that minimum wages contributed to the compression of cross-firm wage differentials measured using two-way fixed effects models. While we find that minimum wages have ambiguous effects on the firm-size wage premium depending on skill, our results are consistent with theirs. This is because, if we average across skill levels, the counterfactual without the minimum wage would have higher firm-size wage premiums.
minimum wage less binding. In the latter, the economy cannot easily accommodate the minimum wage hike, leading to severe effects on informality and unemployment. These effects lower mean log wages for unskilled workers because they reduce their share in formal firms and weaken their bargaining position in informal firms.

6 Concluding Remarks

This paper studies how the interplay between workforce composition, aggregate demand conditions, and labor market institutions—particularly minimum wages—affects informality and other labor market outcomes. 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 compliance decisions 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 along many different margins. Following this, we use changes in tax rates, benefits, the minimum wage, enforcement of regulation, workforce composition, and productivity to show that the model is able to replicate the evolution of labor market outcomes in Brazil across various margins. The improvement in the skill composition of the labor force is the most important driver of the sharp decline in informality among salaried workers observed during the period, though changes in productivity/demand and the minimum wage are also key for rationalizing other patterns observed in the data. The search and matching framework we develop is essential to simultaneously account for all of these factors.

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 minimizing budgetary losses at the same time. On the other hand, a subsidy to formal unskilled workers is much less 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 policy. 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.

The data underlying this article are available in Zenodo, at DOI: https://dx.doi.org/10.5281/zenodo.4463989.

References


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


