

WORKING PAPER · NO. 2020-45

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

Daniel Haanwinckel and Rodrigo R. Soaresz

APRIL 2020

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

Daniel Haanwinckel[†]

and

Rodrigo R. Soares[‡]

April 2020

Abstract

We develop a search model of informal labor markets with worker and firm heterogeneity, intra-firm bargaining with imperfect substitutability across types of workers, and a comprehensive set of labor regulations, including minimum wage. Stylized facts associated with the informal sector, such as smaller firms and lower wages, emerge endogenously as firms and workers decide whether to comply with regulations. Imperfect substitutability across types of workers, decreasing returns to scale, and convex vacancy-posting costs enable the model to reproduce empirical patterns incompatible with existing frameworks in the literature: the presence of skilled and unskilled workers in the formal and informal sectors, the rising share of skilled workers by firm size, the declining formal wage premium by skill, and the rising firm-size wage premium by skill. These features also allow us to analyze the equilibrium responses to changes in the demand and supply of different types of labor. We estimate the model using Brazilian data and show that it reproduces various margins of labor market changes observed between 2003 and 2012. The change in the composition of the labor force appears as the main driving force behind the reduction in informality. We illustrate the use of the model for policy analysis by assessing the effectiveness of a progressive payroll tax in reducing 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, David Card, Pedro Carneiro, Gustavo Gonzaga, Patrick Kline, Nicholas Li, Marc Muendler, Renata Narita, Gabriel Ulyssea, Thaís Vilela, Eduardo Zilberman, 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 Chicago, *haanwinckel* at *uchicago.edu*.

[‡]Columbia University, *r.soares* at *columbia.edu*.

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, 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 (1996*a*), 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.¹ And fourth, 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.

¹Carbonnier (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 does not have different sectors nor firm-size wage premiums.

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 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).² 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.³

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.⁴ On

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

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.⁵

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 (*Instituto Brasileiro de Geografia e Estatística*, IBGE). 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.

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 formal and informal sectors following a compensating differentials logic. But they have a single, homogeneous, labor input (workers are heterogeneous in their endowment of this input) and, given the competitive labor markets assumption, cannot account for unemployment. Marrufo (2001) develops a similar competitive model where firms use a single type of labor and workers choose which sector to work in, but she models workers’ choices as a Roy model—therefore implicitly assuming structural differences across the formal and informal sectors—and does not allow for endogenous compliance decisions on the side of the firms. The competitive model in Amaral and Quintin (2006) has labor heterogeneity and firms hiring both types of workers. However, it focuses on firm—rather than labor—informality, does not have labor market regulations, and, since it features a competitive labor market, cannot account for wage differentials across sectors or unemployment.

Table 1 – Labor Market Outcomes, Brazil, 2003-2012

Sample	Wages in 2003 (R\$)		Wage growth (%)		Formal wage gap (%)		Informality (%)		Unemployment (%)	
	Formal	Informal	Formal	Informal	2003	2012	2003	2012	2003	2012
All	2.54	1.93	28.5	48.5	31.4	13.7	28.1	17.3	13.1	5.7
<i>By schooling:</i>										
Less than 8 years	1.87	1.37	29.8	44.8	36.2	22.1	33.8	23.9	12.7	4.8
8 to 10 years	2.04	1.48	24.1	44.6	38.0	18.4	31.7	22.8	17.2	7.4
H.S., college drop.	2.69	2.17	13.4	25.8	23.9	11.7	24.0	14.1	13.7	6.4
College or more	8.23	6.36	-9.8	11.6	29.4	4.6	16.8	11.9	4.4	2.8

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.

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 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 considerably faster than the growth in income per capita. The minimum wage, in particular, accumulated real gains of 61% between 2003 and 2012.⁶

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).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

⁶The 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.

⁷In 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.

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). 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.⁸ 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, θ_s and θ_u . These variables are important for firms and workers because they determine the probability that vacancies posted by firms are filled, and, accordingly, the probability that unemployed workers find a job. The other two variables are the values of unemployment for skilled and unskilled workers, U_s and U_u . These are the outside options of workers when bargaining and so are important determinants of wages. The bargained wage is, for each firm, a function of the number of workers currently employed, as firm size affects the marginal productivities of the different types of labor. The problem of the firm is then to choose a vacancy-posting strategy—or, equivalently, a firm size—conditional on

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.

its specific wage function and on its compliance decision, made at the beginning of time. Workers accept or reject the offers they receive from firms and bargain over wages. An equilibrium is found by determining the values of θ_s , θ_u , U_s and U_u that are consistent with the aggregate behavior of firms and workers.

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 worker is $q(\theta_i)dt$. θ_i is the labor market tightness in market i , that is, the ratio of vacancies to unemployed workers: $\theta_i = \frac{V_i}{U_i}$, $i \in \{s, u\}$. The probability that an unemployed worker finds a job in a 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 $\Xi_i^j(v)$ for posting v vacancies for workers of skill level i . We assume that the marginal vacancy-posting cost $\xi_i^j(v) = \frac{\partial \Xi_i^j(v)}{\partial v}$ is increasing, as in Bertola and Caballero (1994) and Acemoglu and Hawkins (2014), and that $\Xi_i^j(0) = \xi_i^j(0) = 0$. 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 an important component of wage differences between sectors.

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

⁹This 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.

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 $\frac{d^2 F^z(n_s, n_u)}{dn_s dz} > \frac{d^2 F^z(n_s, n_u)}{dn_u dz}$. 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 τ over its total payroll. If a firm chooses instead to hire workers informally, it avoids payroll taxes but incurs in an informality penalty $\rho^k(n_s + n_u)$. 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), 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.

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 $\rho^k(n_s + n_u)$ 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

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

where $w_i^{z,for}(n_s, n_u)$ and $w_i^{z,k,inf}(n_s, n_u)$ 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 λ_i^j , 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

¹⁰See 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.

level, following other papers with search frictions, multi-worker firms, and concave production functions (e.g., Cahuc, Marque and Wasmer, 2008 and Acemoglu and Hawkins, 2014). The law of motion of labor quantities inside each firm is

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

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

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

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

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

$$(r + \lambda_i^j) \pi_i^{z,k,j}(n_s, n_u) = \begin{cases} F_i^z(n_s, n_u) - (1 + \tau) \left[w_i^{z,for}(n_s, n_u) + \sum_{l=s,u} n_l \frac{\partial w_l^{z,for}(\cdot)}{\partial n_i} \right] & , \text{ for } j = for \\ F_i^z(n_s, n_u) - w_i^{z,k,inf}(n_s, n_u) - \sum_{l=s,u} n_l \frac{\partial w_l^{z,k,inf}(\cdot)}{\partial n_i} - \rho^k(n_s + n_u) & , \text{ for } j = inf, \text{ and} \end{cases} \quad (2)$$

$$\pi_i^{z,k,j}(n_s, n_u) = \frac{\xi_i^j(v_i)}{q(\theta_i)}, \quad (3)$$

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

Equation 2 is an intuitive description of the marginal value of a worker as the discounted sum of expected rents, taking into account the discount rate r and the separation hazard rate λ_i^j . The instantaneous rent is given not only by the difference between marginal product and wage but also by the effect of this additional employee, due to changes in marginal productivities, on the wages of all other workers currently employed by the firm. At the time

¹¹To be entirely precise, we assume that firms are born at the beginning of time with stocks of workers $(n_s^{z,k,j,initial}, n_u^{z,k,j,initial})$ 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.

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 $\xi_i^j(v_i)$ 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:

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

We denote the optimal labor choices in firm z as $n_s^{z,for}$ and $n_u^{z,for}$ (as opposed to arbitrary choices n_s and n_u). The argument of $\xi_i^{for}(\cdot)$ is the steady-state number of posted vacancies as a function of employment. The case for informal firms is analogous, just omitting the payroll tax τ , adding the marginal effect of n_i on the informality penalty $\rho^k(\cdot)$, and writing the dependency on k .¹²

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 σ , 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.¹³ We follow the solution developed by Cahuc, Marque and Wasmer (2008), who analyze this type of problem in a context with search frictions.

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 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 $J_i^j(w)$ as the value that workers of type $i \in \{s, u\}$ place on holding a job position of type $j \in \{for, inf\}$ that pays wage w .¹⁴ Call U_i the opportunity cost of

¹²In a slight abuse of notation, we write $n_s^{z,for}$ and $n_s^{z,k,for}$ 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.

¹³Since 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.

¹⁴Firm heterogeneity z, k does not directly affect the valuation of the job conditional on wages and the formality choice (which

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 $J_i^{for}(w)$, 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:

$$rJ_i^{for}(w) = a_i w + b_i + \lambda_i^{for} \left[U_i - J_i^{for}(w) \right], \text{ and} \quad (4)$$

$$rJ_i^{inf}(w) = w + \lambda_i^{inf} \left[U_i - J_i^{inf}(w) \right], \quad (5)$$

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

The value $J_i^j(w) - U_i$ is the rent earned by workers of type i when they accept a job offer in sector j . For firms, the marginal value of a worker of type i is given by $\pi_i^{z,k,j}(n_s, n_u)$, which was discussed in the previous subsection. So, the Nash sharing rule imposes that the wage function $w_i^{z,k,j}(n_s, n_u)$ must satisfy¹⁶

$$(1 - \sigma) \left[J_i^j \left(w_i^{z,k,j}(n_s, n_u) \right) - U_i \right] = \sigma \pi_i^{z,j}(n_s, n_u), \text{ where } i \in \{s, u\}, \text{ and } j \in \{for, inf\}, \forall z, n_s, \text{ and } n_u. \quad (6)$$

Due to the derivative terms in expression 2 (for $\pi_i^{z,j}$), the set of Nash bargaining equations results in a system of nonlinear differential equations. In Appendix D, we adapt the solution in Cahuc, Marque and Wasmer (2008) to account for two sectors, heterogeneous firms, mandated benefits, and payroll taxes. The resulting wage functions are

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

$$w_i^{z,k,inf}(n_s, n_u) = (1 - \sigma) rU_i + \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma}} \frac{\partial H^{z,k}(\epsilon n_s, \epsilon n_u)}{\partial n_i} d\epsilon,$$

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.

¹⁵This choice is not without loss of generality. We show in this section that including unemployment benefits in the asset value of formal employment implies that all else being equal, higher unemployment benefits reduce formal wages (in partial equilibrium). If, instead, we explicitly modeled unemployment benefits as lump-sum payments received when the formal job relationship is destroyed, they might lead to an increase in wages (in partial equilibrium). This is because workers' fallback positions would be better because of unemployment insurance. Still, note that this would not necessarily be true in equilibrium, since the utility of unemployment is endogenous. In any case, this alternative interpretation 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. So, it is not clear to what extent unemployment benefits can be used as leverage during a bargaining process. In addition, workers can only claim these benefits if they have been employed for a certain number of months in the recent past (six to 12, depending on how many times the worker has claimed the benefit before), ruling out most workers with short tenures. 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 interpretation adopted in our paper, and in Ulyssea (2010), seems therefore more appropriate for our context. Finally, the quantitative implications of unemployment insurance in our exercise are trivial, so our main points are not sensitive to this choice.

¹⁶Our bargaining expression in the presence of payroll taxes differs from that in Mortensen and Pissarides (2001) because we define σ as the effective bargaining share of workers, while they define their bargaining parameter β as the exponent in the generalized Nash bargaining solution. To convert from their notation to ours, one should use the expression $\sigma = a_i \beta / [a_i - (1 + t_i - a_i)(1 - \beta)]$. 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).

with $c_i = [(1 - \sigma)a_i + \sigma(1 + \tau)]$ and $H^{z,k}(n_s, n_u) = F^z(n_s, n_u) - \rho^k(n_s + n_u)$.

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 $w_i^{z,for}(n_s, n_u) = \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{\sigma}{c_i} \frac{\partial F^z}{\partial n_i}$ and $w_i^{z,k,inf}(n_s, n_u) = (1 - \sigma)(rU_i - b_i) + \sigma \frac{\partial H^{z,k}}{\partial n_i}$. However, with decreasing returns to scale, heterogeneous 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 $\epsilon^{\frac{1-\sigma}{\sigma} \frac{\alpha_i}{1+\tau}}$ 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 σ . This also implies that the previous wage function for skilled workers is not valid anymore, since the term $\frac{\partial w_u^{z,for}}{\partial n_s}$ in equation 2 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

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

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, $\pi_s^{z,k,for}$ and $\pi_u^{z,k,for}$. However, when the minimum wage becomes binding for unskilled workers, the effect of firm size on unskilled wages disappears, leading to a discontinuous increase in $\pi_s^{z,k,for}$ and a discontinuous decrease in $\pi_u^{z,k,for}$. The increase in $\pi_s^{z,k,for}$, in turn, causes a discrete increase in skilled wages, which might give an incentive for firms to strategically reduce the number of unskilled workers or increase the number of skilled workers—just enough so that bargained unskilled wages are slightly above the minimum wage.

In Appendix D, we show that because of this discontinuity, there might not be a solution to the first-order conditions when the unconstrained (freely bargained) unskilled wage is slightly lower than the minimum wage. In

these cases, firms engage in the strategic manipulation of firm size described above.¹⁷ In our quantitative exercises, we deal explicitly with this issue by assuming that firms in this situation choose employment figures that (i) satisfy the first-order condition for skilled workers, and (ii) lie immediately to the “left” (in terms of n_u) of the region of the (n_s, n_u) space where the minimum wage binds for unskilled workers. Details are laid out in Appendix D.

3.4 Equilibrium

Now we turn to the determination of the endogenous equilibrium variables θ_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

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

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

$$\theta_s = \frac{\lambda_s^{for} N_s^{for} + \lambda_s^{inf} N_s^{inf}}{q(\theta_s) (\eta - N_s^{for} + N_s^{inf})} \quad \text{and} \quad \theta_u = \frac{\lambda_u^{for} N_u^{for} + \lambda_u^{inf} N_u^{inf}}{q(\theta_u) (1 - \eta - N_u^{for} + N_u^{inf})}. \quad (7)$$

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

$$rU_i = d_i + \theta_i q(\theta_i) [\phi_i E[J_i^{for}(w_i^{for})] + (1 - \phi_i) E[J_i^{inf}(w_i^{inf})]] - U_i \quad (8)$$

$$= \frac{1}{\frac{1}{\theta_i q(\theta_i)} + \frac{\phi_i}{r + \lambda_i^{for}} + \frac{(1 - \phi_i)}{r + \lambda_i^{inf}}} \left[\frac{1}{\theta_i q(\theta_i)} d_i + \frac{\phi_i}{r + \lambda_i^{for}} [a_i E[w_i^{for}] + b_i] + \frac{(1 - \phi_i)}{r + \lambda_i^{inf}} E[w_i^{inf}] \right]. \quad (9)$$

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 $\phi_i = \frac{V_i^{for}}{V_i^{for} + V_i^{inf}} = \frac{\lambda_i^{for} N_i^{for}}{\lambda_i^{for} N_i^{for} + \lambda_i^{inf} N_i^{inf}}$ 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 θ_i , and unemployment values U_i , such that

1. the wage functions solve the system of differential equations given by expressions 2 and 6;

¹⁷It 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.

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
5. 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 θ_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 $w_i^{z,k,j}$ vary across firms by plugging in optimal firm sizes $n_i^{z,k,j}$ in the bargaining equations 6 and using the definition of the values of employment from equations 4 and 5:

$$\begin{aligned}
 w_i^{z,for} &= \max \left\{ \frac{1}{a_i} \left[rU_i - b_i + (r + \lambda_i^{for}) \frac{\sigma}{1 - \sigma} \frac{1}{q(\theta_i)} \xi_i^{for} \left(\frac{\lambda_i^{for} n_i^{z,for}}{q(\theta_i)} \right) \right], \bar{w}_i \right\}, \text{ and} \\
 w_i^{z,k,inf} &= rU_i + (r + \lambda_i^{inf}) \frac{\sigma}{1 - \sigma} \frac{1}{q(\theta_i)} \xi_i^{inf} \left(\frac{\lambda_i^{inf} n_i^{z,k,inf}}{q(\theta_i)} \right).
 \end{aligned}$$

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 ($\xi_i^j(\cdot) = \bar{\xi}_i$). 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:

$$w_i^{inf} = a_i w_i^{for} + b_i + \left(\lambda_i^{inf} - \lambda_i^{for} \right) \frac{\sigma}{1 - \sigma} \frac{\bar{\xi}_i}{q(\theta_i)}. \tag{10}$$

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.¹⁸

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 (*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

¹⁸When 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).

$B(z) : (0, \infty) \rightarrow (0, 1)$ 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.¹⁹

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:

$$\begin{aligned} \rho^{k,r}(n_s, n_u) &= p_{k,r} \times (n_s + n_u)^{P^E}, \text{ where} \\ p_{k,r} &= \exp(P_r + (k - 3)P^D), \end{aligned}$$

and k has a discrete uniform distribution with support $\{1, 2, \dots, 5\}$. 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:

$$\xi_i^j(v) = \xi^j \xi_i [1 - \exp(-\xi^S v)],$$

where ξ_i and ξ^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 ξ^S . Without loss of generality, we normalize $\xi_u = 1$. The total vacancy-posting cost corresponding to this marginal cost is

$$\Xi_i^j(v) = \xi^j \xi_i \left[v + \frac{\exp(-\xi^S v) - 1}{\xi^S} \right].$$

¹⁹See Appendix Figure A.8 for plots of $B(z)$ that illustrate the role of these two parameters (using the actual estimated model and selected counterfactuals). It is also worth mentioning that in section 3, we assume that $F^z(\cdot)$ 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 , $F^z(\cdot)$ can be decreasing in z if $\gamma > 0$. However, $F^z(\cdot)$ 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, $\lambda_s = \lambda^{for} = 1$.

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 η_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 η_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.²⁰

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 CEMPRE 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 α . 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).²¹

Finally, we impute γ 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 γ 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.²²

²⁰The 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 take the average log wage of the worker in the panel and subtract from it the mean log hourly wage among workers in his region-age-education group. We do the same for age. Next, we run a Cox proportional hazards model where the main independent variables are region fixed effects and age-education group fixed effects. To proxy for heterogeneity in ability within groups, we control for a cubic polynomial in residual log wage and another in residual age. Job destruction rates at the region-age-education-sector level are constructed from the estimated relative hazards. We normalize them such that the mean job destruction rate in the region matches the weighted average of the corresponding age-education-sector components, and then we calculate the rates by skill level. Finally, the region-age-education-sector estimates are transformed into region-skill-sector-level moments along with the other PME variables. See Appendix F for additional details.

²¹Since vacancies are not observed in the data, there is no way for the model to distinguish whether a firm is posting lots of vacancies with a low marginal cost or few vacancies with a high marginal cost, so the matching function and the vacancy-posting cost can be scaled without loss of generality.

²²In other words, we are able to reproduce cross-sectional moments very closely even for substantially different values of γ . Using changes in labor supply across regions to identify that parameter would not help, because we allow for flexible, region-specific shifts in labor demand in our counterfactual exercises (see below). Moreover, we refrain from using 2012 data in the estimation, so that 2003-2012 changes can be used as out-of-sample validation exercises. We could have imposed stronger restrictions on how labor demand varies across states (e.g., rule out regional differences in the distribution of firm productivity) and identify γ from cross-sectional variation in

Table 2 – Imputed Parameters

Parameter	Value	Source
Panel A: Estimated from Data		
$\bar{\eta}_r$ (share of skilled workers, avg across regions)	0.5267 (0.0117)	Estimated from PME
$\bar{\lambda}_{u,r}^{inf}$ (inf. unskilled job destruction)	0.0850 (0.0028)	Estimated from PME
$\bar{\lambda}_{s,r}^{inf}$ (inf. skilled job destruction)	0.0103 (0.0055)	Estimated from PME
$\bar{\lambda}_{u,r}^{for}$ (formal unskilled job destruction)	0.0316 (0.0022)	Estimated from PME
$\bar{\lambda}_{s,r}^{for}$ (formal skilled job destruction)	0.0038 (0.0020)	Estimated from PME
m (measure of firms)	0.0905	Calculated from PME, CEMPRE, and ENCINF
Panel B: Calculated from Labor Legislation		
τ (payroll tax rate)	0.7206	Appendix A
a_s, a_u (variable benefits)	0.266, 0.318	Appendix A
b_s^F, b_u^F (fixed benefits)	0.02, 0.05	Appendix A
b_s^D, b_u^D (unemp. insurance)	7.48, 4.00	Appendix A
Panel C: Taken from the Literature and Others		
r (discount rate)	0.008	Real interest rate
σ (worker bargaining power)	0.5	Assumed symmetric
D (matching scale)	0.3	Normalization, value from Ulyssea (2010)
E (matching elasticity)	0.5	Ulyssea (2010)
γ (elasticity of substitution = $\frac{1}{1-\gamma}$)	0.436	Fernandez and Messina (2018); other values in robustness tests

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

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 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.²³

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.

labor supply. However, we only have six regions in the PME data and do not think this approach is appealing in the context we study.

²³The 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).

4.4.1 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, α . 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.

This quasi-experimental estimate helps us identify a key parameter and lends credibility to the counterfactual exercises. The exponent P^E from the $\rho(\cdot)$ 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 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 I.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 I.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,

Table 3 – Estimated Parameters

Parameter	Value	s.e.
<i>Production Function</i>		
\bar{A}_r (productivity shifter, avg across regions)	5.0272	0.3433
B_1 (skill bias shifter)	0.2817	0.1509
B_2 (relative skill bias in productive firms)	0.2900	0.0417
α (decreasing returns)	0.5356	0.0190
<i>Distribution of Firms</i>		
\bar{T}_r (tail parameter, avg across regions)	1.8853	0.0530
<i>Informality Cost</i>		
\bar{P}_r (enforcement shifter, avg across regions)	0.7548	0.2647
P^D (enforcement penalty dispersion)	3.6169	0.1528
P^E (firm size exponent)	2.5287	0.2023
<i>Vacancy-Posting Cost</i>		
ξ^{for} (formal sector component)	548.65	100.11
ξ^{inf} (informal sector component)	0.0054	0.0002
ξ_s (skilled worker shifter)	3.5625	5.8927
ξ^S (steepness)	0.0007	0.0001
<i>Utility Flow while Unemployed</i>		
\bar{d}_{sr} (skilled workers)	-8.2053	0.0862
\bar{d}_{ur} (unskilled workers)	0.1073	0.0934

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 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.²⁴ 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.

²⁴A binding minimum wage does not necessarily imply that the formal wage premium is positive, as illustrated in equivalent figures for some other regions.

Table 4 – Model Fit and Validation, Country Averages, 2003 and 2012

Outcomes	Fit in 2003			Validation: Δ 2003-2012		
	Data ₂₀₀₃	Model ₂₀₀₃	s.e.	Δ Data ₂₀₀₃₋₂₀₁₂	Δ Model ₂₀₀₃₋₂₀₁₂	s.e.
Informality Rate	0.2927	0.2916	0.0296	-0.1053	-0.0603	0.0091
Skilled	0.1492	0.1530	0.0206	-0.0374	-0.0195	0.0043
Unskilled	0.5089	0.4989	0.0465	-0.1526	-0.0435	0.0137
Unemployment (implied)	0.1344	0.1340	0.0098	-0.0750	-0.0434	0.0073
Skilled	0.0311	0.0315	0.0141	-0.0160	0.0027	0.0013
Unskilled	0.2514	0.2500	0.0126	-0.1049	-0.0536	0.0154
Wages (\ln)	0.7860	0.8009	0.0437	0.3382	0.3437	0.0106
Skilled	1.3454	1.3401	0.0661	0.0773	0.0774	0.0093
Unskilled	-0.0268	0.0269	0.0314	0.5151	0.5152	0.0152
Formal wage premium	0.0617	0.0101	0.0351	-0.0214	-0.0353	0.0107
Skilled	0.0429	-0.0431	0.0258	-0.0215	-0.0014	0.0011
Unskilled	0.0892	0.0946	0.0560	-0.0047	-0.0723	0.0249
Firm-size premium 6-10 workers	0.0494	0.0078	0.0048	-0.0121	-0.0065	0.0020
Skilled	0.0592	0.0130	0.0081	-0.0173	-0.0106	0.0029
Unskilled	0.0330	-0.0023	0.0037	-0.0074	-0.0003	0.0037
Firm-size premium 11 or more workers	0.0732	0.1030	0.0224	-0.0108	-0.0017	0.0041
Skilled	0.0837	0.1240	0.0361	-0.0128	-0.0079	0.0031
Unskilled	0.0611	0.0734	0.0117	-0.0190	-0.0073	0.0078
% formal workers firms 6-10 employees	0.0775	0.0386	0.0040	-0.0271	-0.0064	0.0015
Skilled	0.0579	0.0242	0.0032	-0.0154	-0.0029	0.0006
Unskilled	0.1280	0.0705	0.0069	-0.0490	-0.0045	0.0035
% formal workers workers firms \geq 11 employees	0.8552	0.8594	0.0104	0.0492	0.0235	0.0044
Skilled	0.8968	0.9226	0.0098	0.0240	0.0084	0.0020
Unskilled	0.7431	0.7119	0.0255	0.1057	0.0141	0.0116
% informal workers firms 6-10 employees	0.1234	0.1585	0.0123	-0.0116	-0.0189	0.0035
Skilled	0.1145	0.1157	0.0156	-0.0051	-0.0001	0.0023
Unskilled	0.1267	0.1769	0.0122	-0.0128	-0.0224	0.0042
% informal workers workers firms \geq 11 employees	0.5133	0.5207	0.0466	0.0345	-0.0227	0.0048
Skilled	0.6797	0.6712	0.0527	-0.0203	-0.0461	0.0052
Unskilled	0.4383	0.4534	0.0438	0.0299	-0.0392	0.0079
Labor share	0.528	0.5157	0.0027			
Enforcement elasticity of informality	-0.0534	-0.0529	0.016			
formal workers in firms \geq 500						
formal workers in firms \geq 100	0.7224	0.7595	0.0194			

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.

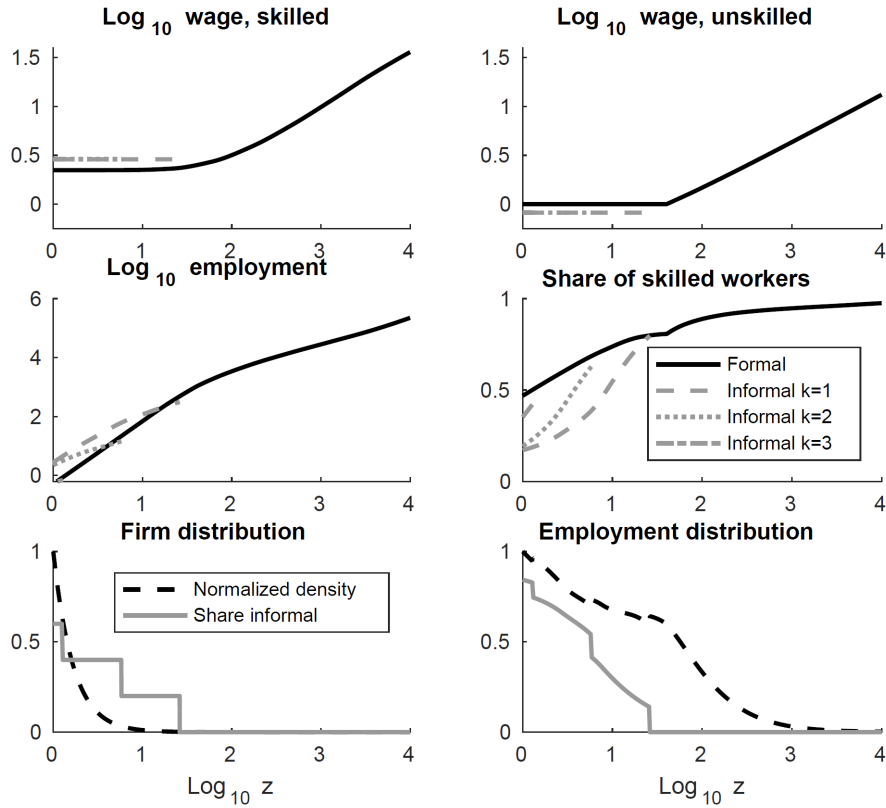


Figure 1 – Cross-Sectional Features of the Equilibrium, Recife Metropolitan Region, 2003

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.

Interestingly, low z firms that are formal tend to be smaller and, at the same time, more skill intensive than similar 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.²⁵ 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

²⁵The share of children 7–14 years old attending school increased from 80.9% in 1980 to 97% in 1999 (de Oliveira, 2007).

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%.

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.²⁶

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).²⁷

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

²⁶This 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).

²⁷See Appendix Figure A.8 for a plot of the $B(z)$ function in both periods.

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 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

Table 5 – Individual Contributions of Each Factor to Changes in the Brazilian Labor Market from 2003 to 2012

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

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).

(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 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.13. 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.13, 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.²⁸

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

²⁸Appendix 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.

declines in informality rates. We use census data for the whole country from 1991, 2000, and 2010. 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 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

Table 6 – Hypothetical Policy Experiments

Outcomes	(1)	(2)		(3)	(4)
	1 p.p. reduction in payroll tax	Progressive payroll tax		Doubling transfer to unskilled	
	$\Delta\tau = -0.01$	$\Delta\tau_s = 0$		$\Delta b_u^F = 0.05$	
		$\Delta\tau_u = -0.01$	$\Delta\tau_u = -0.10$		
Informality	-0.0027	-0.0013	-0.0139	-0.0013	
Skilled	-0.0017	-0.0002	-0.0025	-0.0005	
Unskilled	-0.0051	-0.0039	-0.0414	-0.0006	
Unemployment	-0.04	-0.0003	-0.0038	0.0027	
Skilled	0.0000	0.0000	-0.0001	0.0000	
Unskilled	-0.1000	-0.0009	-0.0103	0.0079	
Wages (log)	0.0042	0.0007	0.0065	0.0021	
Skilled	0.0051	0.0001	0.0013	0.0003	
Unskilled	0.0030	0.0025	0.0251	0.0004	
Govt Surplus (%)	-1.75	-0.17	-1.83	-1.16	
Net Output (%)	0.06	0.01	0.08	-0.14	
Labor Share	0.0005	0.0001	0.0016	-0.0008	

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 $b_u^F = 0.05$). 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.

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.²⁹ The reduction in informality is roughly 50% of that in column 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 1.³⁰ 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.

²⁹In 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 close to the minimum wage (perhaps up to two times the minimum). In the model, that would correspond to the vast majority of unskilled workers and none of the skilled ones. In our calibration, high-wage unskilled workers are always employed at high-productivity, non-marginal formal firms. Thus, the simplifying assumption is conservative. It overstates the fall in revenues associated with the progressive taxation but does not overstate the impact on informality

³⁰This quantitative result is the only main result in the paper that is substantially affected by our choice of the parameter γ . The lower value of γ , corresponding to a higher elasticity of substitution, leads to a reduction in informality of only 0.78 percentage points, while the high γ , 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.

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_u^F 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 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.³¹ 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.³²

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

³¹We 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.

³²We 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.

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.³³

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 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

³³Engbom 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.

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.

References

- Acemoglu, Daron, and David Autor.** 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. Vol. 4 of *Handbook of Labor Economics*, , ed. David Card and Orley Ashenfelter, 1043 – 1171. Elsevier.
- Acemoglu, Daron, and William Hawkins.** 2014. "Search with multi-worker firms." *Theoretical Economics*, , (9): 583–628.
- Albrecht, James, Lucas Navarro, and Susan Vroman.** 2009. "The Effects of Labour Market Policies in an Economy with an Informal Sector." *The Economic Journal*, 119(539): 1105–1129.
- Almeida, Rita, and Pedro Carneiro.** 2012. "Enforcement of Labor Regulation and Informality." *American Economic Journal: Applied Economics*, 4(3): 64–89.
- Amaral, Pedro S., and Erwan Quintin.** 2006. "A competitive model of the informal sector." *Journal of Monetary Economics*, 53(7): 1541–1553.
- Araujo, Luis, and Vladimir P. Ponczek.** 2011. "Informal wages in an economy with active labor courts." *Textos para discussãoo, Escola de Economia de Sãoo Paulo, Getulio Vargas Foundation (Brazil)*, 294.
- Barbosa Filho, Fernando Holanda, and Rodrigo Leandro Moura.** 2015. "Evoluçoo Recente da Informalidade do Emprego no Brasil: Uma Análise Segundo Características da Oferta e Demanda de Trabalho." *Pesquisa e Planejamento Econõmico*, 45: 101–123.
- Bargain, Olivier, and Prudence Kwenda.** 2011. "Earnings Structures, Informal Employment, and Self-Employment: New Evidence From Brazil, Mexico and South Africa." *Review of Income and Wealth*, 57: S100–S122.
- Bargain, Olivier, Eliane El Badaoui, Prudence Kwenda, Eric Strobl, and Frank Walsh.** 2012. "The Formal Sector Wage Premium and Firm Size for Self-employed Workers." *IZA Discussion Paper*, 6604.
- Barros, Ricardo, and Carlos Henrique Corseuil.** 2001. "The Impact of Regulations on Brazilian Labor Market Performance." *Inter-American Development Bank Research Network Working paper*, R-427.
- Bertola, Giuseppe, and Ricardo J. Caballero.** 1994. "Cross-Sectional Efficiency and Labour Hoarding in a Matching Model of Unemployment." *The Review of Economic Studies*, 61(3): 435–456.
- Boeri, Tito, and Pietro Garibaldi.** 2007. "Shadow Sorting." *NBER International Seminar on Macroeconomics*, 2005: 125–163.

- Boeri, Tito, Pietro Garibaldi, and Marta Ribeiro.** 2011. "The Lighthouse Effect and Beyond." *Review of Income and Wealth*, 57: S54–S78.
- Bosch, Mariano, and Julen Esteban-Pretel.** 2012. "Job creation and job destruction in the presence of informal markets." *Journal of Development Economics*, 98(2): 270 – 286.
- Bosch, Mariano, and William F. Maloney.** 2010. "Comparative analysis of labor market dynamics using Markov processes: An application to informality." *Labour Economics*, 17(4): 621–631.
- Bosch, Mariano, Edwin Goni-Pacchioni, and William Maloney.** 2012. "Trade liberalization, labor reforms and formal-informal employment dynamics." *Labour Economics*, 19(5): 653 – 667.
- Botelho, Fernando, and Vladimir Ponczek.** 2011. "Segmentation in the Brazilian Labor Market." *Economic Development and Cultural Change*, 59(2): 437–463.
- Cahuc, Pierre, Francois Marque, and Etienne Wasmer.** 2008. "A Theory of Wages and Labor Demand with Intra-firm Bargaining and Matching Frictions." *International Economic Review*, 49(3): 943–972.
- Cain, Glen G.** 1976. "The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey." *Journal of Economic Literature*, 14(4): 1215–1257.
- Carbonnier, Clément.** 2015. "Payroll Taxation, qualifications, wages and unemployment rates in a frictional labor market with productive interactions between segments." Unpublished manuscript.
- Cardoso, Adalberto, and Telma Lage.** 2005. "A Inspeção do Trabalho no Brasil." *Revista de Ciências Sociais*, 48(3): 451–490.
- Chandrasekhar, Arun G, Melanie Morten, and Alessandra Peter.** 2020. "Network-Based Hiring: Local Benefits; Global Costs." National Bureau of Economic Research Working Paper 26806.
- Cheron, Arnaud, Jean-Olivier Hairault, and François Langot.** 2008. "A quantitative evaluation of payroll tax subsidies for low-wage workers: An equilibrium search approach." *Journal of Public Economics*, 92(3-4): 817–843.
- de Mel, Suresh, David McKenzie, and Christopher Woodruff.** 2013. "The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka." *American Economic Journal: Applied Economics*, 5(2): 122–50.
- de Oliveira, Romualdo Portela.** 2007. "Da universalização do ensino fundamental ao desafio da qualidade: uma análise histórica." *Educação & Sociedade*, 28(100): 661–690.
- Dickens, William T., and Kevin Lang.** 1985. "A Test of Dual Labor Market Theory." *The American Economic Review*, 75(4): pp. 792–805.
- Engbom, Niklas, and Christian Moser.** 2018. "Earnings Inequality and the Minimum Wage: Evidence from Brazil." Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute 7.

- Fajnzylber, Pablo, William F. Maloney, and Gabriel V. Montes-Rojas.** 2011. “Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program.” *Journal of Development Economics*, 94(2): 262 – 276.
- Fernandez, Manuel, and Julian Messina.** 2018. “Skill premium, labor supply, and changes in the structure of wages in Latin America.” *Journal of Development Economics*, 135: 555 – 573.
- Ferreira, Pedro C., and Fernando Veloso.** 2013. “O Desenvolvimento Econômico Brasileiro no Píçes-Guerra.” In *Desenvolvimento Econômico: Uma Perspectiva Brasileira.*, ed. Fernando Veloso, Pedro C. Ferreira, Fabio Giambiagi and Samuel Pessíçea, 129–165. Rio de Janeiro:Campus.
- Fields, Gary S.** 1975. “Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs.” *Journal of Development Economics*, 2(2): 165 – 187.
- Fields, G. S.** 1990. “Labour market modelling and the urban informal sector: Theory and evidence.” In *The informal sector revisited.*, ed. D. Turnham, B. Salomíçce and A. Schwarz, 49–69. Organisation for Economic Cooperation and Development.
- Firpo, Sergio, and Alysson Portella.** 2019. “Decline in Wage Inequality in Brazil : A Survey.”
- Foguel, Miguel, Lauro Ramos, and Francisco Carneiro.** 2001. “The Impacts of the Minimum Wage on the Labor Market, Poverty and Fiscal Budget in Brazil.” *IPEA Textos para Discussao*, 839: 1–42.
- Galiani, Sebastian, and Frederico Weinschelbaum.** 2012. “Modelling Informality Formally: Households and Firms.” *Economic Inquiry*, 50: 821–838.
- Gollin, Douglas.** 2002. “Getting Income Shares Right.” *Journal of Political Economy*, 110(2): 458–474.
- Gonzaga, Gustavo.** 2003. “Labor Turnover and Labor Legislation in Brazil.” *Economíçæa*, 4(1): 165–222.
- Gunther, Isabel, and Andrey Launov.** 2012. “Informal employment in developing countries: Opportunity or last resort?” *Journal of Development Economics*, 97(1): 88–98.
- Hornstein, Andreas, Per Krusell, and Giovanni L. Violante.** 2011. “Frictional Wage Dispersion in Search Models: A Quantitative Assessment.” *American Economic Review*, 101(7): 2873–98.
- IBGE.** 2005. “Estatíçísticas do Cadastro Central de Empresas 2003.” Available at <http://www.ibge.gov.br/home/estatistica/economia/cadastroempresa/2003/default.shtm>.
- Kovak, Brian.** 2013. “Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?” *American Economic Review*, 103: 1960–1976.
- Kugler, Adriana, Maurice Kugler, and Luis Prada.** 2017. “Do Payroll Tax Breaks Stimulated Informality? Evidence from Colombia’s Reform.” *Economíçæa*, 18: 3–40.
- La Porta, Rafael, and Andrei Shleifer.** 2008. “The Unofficial Economy and Economic Development.” *Brookings Papers on Economic Activity*, 2008(2): 275–363.

- Lee, David, and Emmanuel Saez.** 2012. "Optimal minimum wage policy in competitive labor markets." *Journal of Public Economics*, 96: 739–749.
- Lehman, Hartmut, and Norberto Pignatti.** 2007. "Informal Employment Relationships and Labor Market Segmentation in Transition Economies: Evidence from Ukraine." *IZA Discussion Paper*, 3269.
- Lemos, Sara.** 2009. "Minimum Wage Effects in a Developing Country." *Labour Economics*, 16: 224–237.
- Lucas, Robert E. Jr.** 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics*, 9(2): 508–523.
- Maloney, William F.** 2004. "Informality Revisited." *World Development*, 32(7): 1159–1178.
- Marrufo, Grecia.** 2001. "The Incidence of Social Security Taxes in Economies with Partial Compliance: Evidence from the SS Reform in Mexico." PhD diss. University of Chicago.
- Meghir, Costas, Renata Narita, and Jean-Marc Robin.** 2015. "Wages and Informality in Developing Countries." *American Economic Review*, 105(4): 1509–46.
- Melitz, Marc J.** 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.
- Mello, Rafael F., and Daniel D. Santos.** 2009. "Aceleraç o educacional e a queda recente da informalidade." *Boletim Mercado de Trabalho IPEA*, 39: 27–33.
- Menezes Filho, Naercio Aquino, Marcos Mendes, and Eduardo Simoes. de Almeida.** 2004. "O Diferencial de Salarios Formal-Informal no Brasil: Segmentaç o ou V zes de Seleç o?" *Revista Brasileira de Economia*, 58(2): 235–48.
- Messina, Julian, and Joana Silva.** 2018. *Wage Inequality in Latin America: Understanding the Past to Prepare for the Future*. World Bank Group.
- Monteiro, Joana C.M., and Juliano J. Assunç o.** 2012. "Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises." *Journal of Development Economics*, 99(1): 105 – 115.
- Mortensen, Dale T., and Christopher Pissarides.** 2001. "Taxes, Subsidies and Equilibrium Labour Market Outcomes." *CEPR Discussion Paper 2989*.
- MTE.** 2013. "Resultados da Fiscalizaç o do Trabalho - N vel Brasil - 2003 a 2012." Available at http://portal.mte.gov.br/fisca_trab/estatisticas.htm.
- Perry, Guillermo, William F. Maloney, Omar Arias, Pablo Fajnzylber, Andrew Mason, and Jaime Saavedra.** 2007. *Informality: Exit and Exclusion*. World Bank.
- Pissarides, Christopher A.** 2000. *Equilibrium Unemployment Theory*. MIT Press.

- Rauch, James E.** 1991. "Modelling the informal sector formally." *Journal of Development Economics*, 35(1): 33 – 47.
- Robin, Jean-Marc, and Sébastien Roux.** 2002. "An equilibrium model of the labour market with endogenous capital and two-sided search." *Annales d'Economie et de Statistique*, (67): 257–307.
- Souza, André Portela, Sérgio P. Firpo, Vladimir Ponczek, Eduardo Zylberstajn, and Felipe Ribeiro.** 2012. *Custo do Trabalho no Brasil: Proposta de uma nova metodologia de mensuração*. FGV/EESP.
- Stole, Lars A., and Jeffrey Zwiebel.** 1996a. "Intra-Firm Bargaining under Non-Binding Contracts." *The Review of Economic Studies*, 63(3): 375–410.
- Stole, Lars A., and Jeffrey Zwiebel.** 1996b. "Organizational Design and Technology Choice under Intrafirm Bargaining." *The American Economic Review*, 86(1): 195–222.
- Tornarolli, Leopoldo, Diego Battistini, Leonardo Gasparini, and Pablo Gluzmann.** 2012. "Exploring trends in labor informality in Latin America, 1990-2010." Center for Distributive, Labor and Social Studies (CEDLAS), Universidad Nacional de La Plata.
- Ulyssea, Gabriel.** 2010. "Regulation of entry, labor market institutions and the informal sector." *Journal of Development Economics*, 91(1): 87 – 99.
- Ulyssea, Gabriel.** 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil." *American Economic Review*, 108(8): 2015–2047.

A Appendix: Costs of Formal Labor and Valuation of Benefits by the Formal Employee

In this appendix, we calculate the cost of formal employment and the valuation of mandated benefits by formal workers according to the methodology of Souza et al. (2012). We first show the results for the baseline year, October 2003. Then, we discuss the changes in regulations from 2003 to 2012 and calculate the parameters for October 2012.

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

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

A.1 Costs of Formal Labor

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

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

Table A.1 – Cost of Formal Employment in October 2003

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

cannot rely on them to do so.

Now we turn to the contributions that the employer is obliged to pay. These are levied over not only the nominal monthly wage but also the additional payments described above (thirteenth salary, vacation stipend, and advance notice). The first item is the monthly contribution of 8% of the wage to the worker’s severance payment fund (FGTS). In the following row, we state the expected balance of this fund after 33.24 months, which is the average duration of formal employment (for simplicity, this duration is assumed to be the same across workers). This information is used to calculate the severance payment, which is 50% of the total FGTS balance at the time of dismissal. Note that of this 50% payment, 40% go to the dismissed employee, and the remaining 10% are appropriated by the government. In 2003, there was an additional temporary contribution to the FGTS fund of 0.5%, which expired in December 2006.

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

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

We then proceed to the calculation of the cost of formal employment in October 2012. The only change in regulations that affected the cost paid by the employer was the phasing out of the temporary FGTS contribution. When we exclude that contribution, we find that the equivalent payroll tax rate in October 2012 was 71.43% of the nominal wage.

Table A.2 – Valuation of Variable Benefits

Item	Rationale	October 2003		October 2012	
		Low wage	High wage	Low wage	High wage
Nominal wage (A)		508.84	1181.36	1009.69	1948.70
13th salary (A.1)	1/12 of A	42.40	98.45	84.14	162.39
Vacation stipend (A.2)	0.33/12 of A	14.13	32.82	28.05	54.13
Advance notice	(A+A.1+A.2) x prob. dismissal	26.52	38.93	52.63	64.22
Raw total wage (B)		591.90	1351.56	1174.50	2229.44
INSS deduction	7.65%/7.93% (03) or 8.00%/8.27% (12) of B	-48.03	-114.74	-98.24	-195.32
Income tax (IRPF) deduction	0%/5.90% (03) or 0%/5.60% (12) of B	-0.87	-35.24	-2.23	-47.00
Valuation of FGTS fund	50% of employer contribution	23.68	54.06	46.98	89.18
Severance payment	40% of FGTS balance x prob. dismissal	18.94	43.25	37.58	71.34
Work accident insurance (SAT)	2% of B	11.84	27.03	23.49	44.59
Total with contributions (C)		597.47	1325.93	1182.08	2192.22
Vacation adjustment	Equal to the cost of vacation paid by employer	72.97	169.41	144.79	279.44
Total valuation		670.43	1495.34	1326.87	2471.67
Variable benefits parameter		0.318	0.266	0.314	0.268

A.2 Valuation of Mandated Benefits

In this subsection, we account for all characteristics of formal employment that can make it more or less attractive to workers when compared to informal employment. Unlike the previous section, some of the items we consider, such as the income tax, affect skilled and unskilled workers differently. Thus, we have separate valuations for different types of workers. So that these values are parametric and not dependent on our modeling choices, we use workers with at least a secondary education to calculate the numbers for skilled workers, and workers with less than 8 years of schooling are used to calculate the numbers for unskilled workers. Small changes in the definitions of the skilled and unskilled groups in these calculations make very little difference to the final numbers used in the model.

Table A.2 shows our numbers for benefits and contributions, calculated as fractions of the reference salary. These determine the variable benefits parameters in our *benefits* expressions: a_s and a_u . The first five rows are similar to those in Table A.1: formal workers receive not only the nominal monthly wage but also the thirteenth salary, the vacation stipend, and the advance notification in case of dismissal. Two items are then deducted from the gross wages: the social security contribution (INSS) and the income tax (IRPF). We calculate the deductions for each formal worker in the PME dataset using the corresponding tax brackets and rates in each year, and then we calculate the average rate per worker.

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

The remaining benefits are the severance payment, the compulsory work accident insurance (SAT), and vacations.

The first two items are calculated in a manner similar to that of the previous subsection when assessing the costs of formal employment. To input the valuation of vacations by workers, we use exactly the same value that is calculated as the cost of vacations to employers. In this sense, vacations can be regarded as a transfer from firm to worker. Thus, if we calculate the difference between aggregate total payroll taxes and aggregate benefits, vacations and other transfers, such as the thirteenth salary, are canceled out, and we can use the result as government surplus in the model. We find that the net valuation of variable benefits is around 31% of the base salary for unskilled workers and around 27% for skilled workers.

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

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

B Appendix: Informality Trends by Economic Activity

In this appendix, we show that the decline in the informality rate in Brazil was widespread in the economy and that it was not caused by a reallocation of workers across sectors. In the PME survey, workers report the economic activity to which their main job belongs, choosing one of 60 categories. In Table A.3, we list the top-15 economic activities in terms of employment. Together, they account for 76% of the workforce in 2003 and 78% in 2012. For each activity, we compute the formality rates in 2003 and 2012 and also the share of the workforce employed therein. Note that since the PME targets workers in large metropolitan areas, few of them are employed in agricultural or extractive activities.

Table A.3 – Informality Trends per Economic Activity

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

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

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

$$\text{Total contribution}_i = F_{i,2012}P_{i,2012} - F_{i,2003}P_{i,2003}$$

$$\text{Within contribution}_i = P_{i,2003} \cdot (F_{i,2012} - F_{i,2003})$$

$$\text{Between contribution}_i = F_{i,2012} \cdot (P_{i,2012} - P_{i,2003})$$

where $P_{i,t}$ and $F_{i,t}$ respectively denote the share of the workforce engaged in, and the formality rate of, activity i in year t . The sum of the within contributions describes what would happen if the share of workers in each activity

remained constant from 2003 to 2012, but the formality rates within each activity changed. The sum of between contributions accounts for the part of the decline in informality that can be attributed to changes in the size of each activity, given the formality rates in 2012. As can be seen in the bottom row of Table A.3, the decline in informality can be accounted for almost exclusively with changes within each activity.

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

C Appendix: Solution to the Problem of the Firm

Consider problem 1 and denote $\frac{\partial \Pi^{z,j}(n_s, n_u)}{\partial n_i} = \pi_i^{z,j}(n_s, n_u)$. The optimality of controls v_s, v_u yields

$$-\xi_i^j(v_i) + q(\theta_i)\pi_i^{z,j}(n_s^+, n_u^+) = 0$$

Also, differentiating the value function in n_i yields

$$(1 + rdt)\pi_i^{z,j}(n_s, n_u) = \frac{\partial \psi^{z,j}(\cdot)}{\partial n_i} dt + (1 - \lambda_i^j dt)\pi_i^{z,j}(n_s^+, n_u^+)$$

If we differentiate $\psi^{z,j}(\cdot)$ in n_i and restrict attention to steady-state equilibria, where $n_i^+ = n_i$, the two equations above result in 3 and 2 respectively.

The remainder of this appendix shows how to rewrite the first-order conditions in a way that is more computationally efficient, requiring fewer numerical integrations. Those first-order conditions are analogous to Equation 12 in Cahuc, Marque and Wasmer (2008). The same integration by parts used to get from (11) to (12) in their paper works in our case. For informal firms, we can use their results. For formal firms, we show the derivation below.

C.1 Informal Firms

Assuming the production function is homogeneous of degree α and that the informality penalty $\rho(\cdot)$ is homogeneous of degree PE , the FOC is

$$\frac{F_i^z(n_s^{inf}(z), n_u^{inf}(z))}{1 - \sigma(1 - \alpha)} - \frac{\rho_i(n_s^{inf}(z), n_u^{inf}(z))}{1 + \sigma(PE - 1)} = rU_i + \frac{r + \lambda_i^{inf}}{(1 - \sigma)q(\theta_i)} \xi_i^{inf} \left(\frac{\lambda_i^{inf} n_i^{inf}(z)}{q(\theta_i)} \right) \quad i = s, u$$

where $\rho_i(\cdot)$ is the marginal informality penalty and $\xi_i^{inf}(\cdot)$ is the marginal vacancy-posting cost for informal vacancies of skill type i . The denominators in the left-hand side are the inverses of “overemployment factors” from Cahuc, Marque and Wasmer (2008).

The wage function is

$$w_i^{inf}(n_s, n_u) = (1 - \sigma)rU_i + \sigma \left[\frac{F_i^z(n_s, n_u)}{1 - \sigma(1 - \alpha)} - \frac{\rho_i(n_s, n_u)}{1 + \sigma(PE - 1)} \right] \quad i = s, u$$

which we use to get firm-specific wages after solving for optimal employment values.

C.2 Formal Firms, Unconstrained Regime

The first-order condition is

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

and the bargained wage is

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

The derivatives of bargained wages l with respect to employment i are

$$\frac{\partial w_l^{z,for}(\cdot)}{\partial n_i} = \frac{1}{1+\tau_l} \int_0^1 \epsilon^{\left(\frac{1-\sigma}{\sigma} + \frac{1+\tau_i}{a_i}\right) \frac{a_l}{1+\tau_l}} \frac{\partial F^z \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_l}{1+\tau_l}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_l}{1+\tau_l}} n_u \right)}{\partial n_l \partial n_i} d\epsilon$$

Now note that the integral in the expression for $w_i^{z,for}(n_s, n_u)$ above can be rewritten as follows, after an integration by parts:

$$\begin{aligned} V &= \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_i}{1+\tau_i}} \frac{\partial F^z \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i} d\epsilon \\ &= \tilde{\sigma}_i \left[\frac{\partial F^z(n_s, n_u)}{\partial n_i} - \sum_{i=s,u} n_l \frac{1+\tau_l}{a_l} \frac{a_i}{1+\tau_i} \int_0^1 \epsilon^{\left(\frac{1-\sigma}{\sigma} + \frac{1+\tau_l}{a_l}\right) \frac{a_i}{1+\tau_i}} \frac{\partial F^z \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i \partial n_l} d\epsilon \right] \\ &= \tilde{\sigma}_i \left[\frac{\partial F^z(n_s, n_u)}{\partial n_i} - \sum_{i=s,u} n_l \frac{1+\tau_l}{a_l} a_i \frac{\partial w_i^{z,for}(\cdot)}{\partial n_l} \right] \end{aligned}$$

where in the second line, we assumed that the production function is such that $\lim_{\epsilon \rightarrow 0} \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_i}{1+\tau_i} + 1} \frac{\partial F^z \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i} = 0$, and in the third line, we plugged in the previous result with the i and l indexes inverted. Appendix D shows that $\frac{\partial w_s^{z,for}(\cdot)}{\partial n_u} = \frac{a_u}{a_s} \frac{\partial w_u^{z,for}(\cdot)}{\partial n_s}$. That allows us to write

$$\sum_{i=s,u} n_l (1+\tau_l) \frac{\partial w_l^{z,for}(\cdot)}{\partial n_i} = \frac{\partial F^z(n_s, n_u)}{\partial n_i} - \frac{V}{\tilde{\sigma}_i}$$

Plugging this result back in the first-order condition, we obtain

$$\Upsilon_i^{z,unc}(n_s^{for}(z), n_u^{for}(z)) F_i^z(n_s^{for}(z), n_u^{for}(z)) = \frac{1+\tau_i}{a_i} \left[\frac{\tilde{\sigma}_i}{\sigma_i} (rU_i - b_i) + \frac{r + \lambda_i^{for}}{(1-\sigma_i)q(\theta_i)} \xi_i^{for} \left(\frac{\lambda_i^{for} n_i^{for}(z)}{q(\theta_i)} \right) \right] \quad i = s, u$$

where

$$\begin{aligned} \Upsilon_s^{z,unc}(n_s, n_u) &= \frac{1}{\sigma F_s^z(n_s, n_u)} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_s}{1+\tau_s}} F_s^z \left(\epsilon n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_s}{1+\tau_s}} n_u \right) d\epsilon \\ \Upsilon_u^{z,unc}(n_s, n_u) &= \frac{1}{\sigma F_u^z(n_s, n_u)} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_u}{1+\tau_u}} F_u^z \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_u}{1+\tau_u}} n_s, \epsilon n_u \right) d\epsilon \end{aligned}$$

are the overemployment factors (corrected for payroll taxes and benefits) and $\tilde{\sigma}_i = \frac{\sigma(1+\tau_i)}{\sigma(1+\tau_i) + (1-\sigma)a_i}$. Here we use numerical integration.

The solution in this regime is valid if the wage for unskilled workers is above the minimum wage.

C.3 Formal Firms, Strictly Binding Regime

Assuming the minimum wage binds for unskilled workers, the first-order condition for skilled workers can be written in the same form as above:

$$\Upsilon_s^{z, str}(n_s^{for}(z), n_u^{for}(z)) F_s^z(n_s^{for}(z), n_u^{for}(z)) = \frac{1 + \tau_s}{a_s} \left[\frac{\tilde{\sigma}_s}{\sigma_s} (rU_s - b_s) + \frac{r + \lambda_s^{for}}{(1 - \sigma_s)q(\theta_s)} \xi_i^{for} \left(\frac{\lambda_s^{for} n_s^{for}(z)}{q(\theta_s)} \right) \right] \quad i = s, u$$

but with a different formula for the overemployment factor:

$$\Upsilon_s^{z, str}(n_s, n_u) = \frac{1}{\sigma F_s^z(n_s, n_u)} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_s}{1+\tau_s}} F_s^z(\epsilon n_s, n_u) d\epsilon.$$

The first-order condition for unskilled labor is

$$F_u^z(n_s^{for}(z), n_u^{for}(z)) = (1 + \tau_u)\bar{w} + \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_s}{1+\tau_s}} F_{su}^z(\epsilon n_s, n_u) d\epsilon + \frac{r + \lambda_u^{for}}{q(\theta_u)} \xi_u^{for} \left(\frac{\lambda_u^{for} n_u^{for}(z)}{q(\theta_u)} \right).$$

The strict regime is valid if the freely bargained (unconstrained) wage for unskilled workers, when evaluated at the strict solution, is at or below the minimum wage.

C.4 Formal Firms, Strategic Solution

When neither the unconstrained nor the strict regime is valid, the solution of the firm problem is called "strategic." This solution is given by employment values satisfying the unconstrained solution for skilled labor, such that the freely bargained unskilled wage is exactly the minimum wage.

D Appendix: Solution to the Wage Bargaining Equation

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

The Nash bargaining equation is

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

Replacing equations 2 and 4 in the expression above, we find the following system of nonlinear differential equations:

$$c_i w_i(n_s, n_u) = (1 - \sigma)(rU_i - b_i) + \sigma \left[F_i(n_s, n_u) - (1 + \tau_s)n_s \frac{\partial w_s(\cdot)}{\partial n_i} - (1 + \tau_u)n_u \frac{\partial w_u(\cdot)}{\partial n_i} \right] \quad (11)$$

where $c_i = [(1 - \sigma)a_i + \sigma(1 + \tau_i)]$, and we allow for taxes to vary by skill level, since we consider progressive payroll taxes in one of our quantitative exercises.

To solve this system, the first step is to write it in a more convenient form. Taking the partial derivative of 11 with respect to n_u when $i = s$ yields

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

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

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

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

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

Using the definition of c_i , we obtain

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

Which we can use to write the system of equations defined in 11 as

$$c_i w_i(n_s, n_u) = (1 - \sigma)(rU_i - b_i) + \sigma \left[F_i(n_s, n_u) - (1 + \tau_i) \left(\chi_{i,s} n_s \frac{\partial w_i(\cdot)}{\partial n_s} + \chi_{i,u} n_u \frac{\partial w_i(\cdot)}{\partial n_u} \right) \right] \quad (12)$$

where

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

Following Cahuc, Marque and Wasmer (2008) (henceforth CMW), we first solve the equation for the case in which $\chi_{i,j} = 1$. Later, we generalize the solution. CMW's insight is to perform a change of coordinates that allows us to express the term multiplying $(1 + \tau_i)$ in equation 12 in a simpler manner, effectively obtaining a univariate

differential equation as the result. The transformation we need is

$$n_s = \zeta \cos \delta$$

$$n_u = \zeta \sin \delta$$

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

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

Which is the term multiplying $(1 + \tau_i)$ in equations 12 if $\chi_{i,j} = 1$. Following the same notation, let $\hat{F}_{n_i}(\zeta, \delta) = \frac{\partial F(\zeta \cos \delta, \zeta \sin \delta)}{\partial n_i}$ denote the marginal product function in the new coordinate system. We can then rewrite the differential equations as

$$\frac{\partial \hat{w}_i(\zeta, \delta)}{\partial \zeta} + \frac{c_i}{\sigma(1 + \tau_i)\zeta} \hat{w}_i(\zeta, \delta) = \frac{1 - \sigma}{\sigma(1 + \tau_i)\zeta} (rU_i - b_i) + \frac{1}{(1 + \tau_i)\zeta} \hat{F}_{n_i}(\zeta, \delta) \quad (13)$$

We guess the following form for the solution:

$$\begin{aligned} \hat{w}_i(\zeta, \delta) &= C(\zeta, \delta) \zeta^{-\frac{c_i}{\sigma(1 + \tau_i)}} + D(\delta) \\ \frac{\partial \hat{w}_i(\zeta, \delta)}{\partial \zeta} &= C'(\zeta, \delta) \zeta^{-\frac{c_i}{\sigma(1 + \tau_i)}} - C(\zeta, \delta) \frac{c_i}{\sigma(1 + \tau_i)} \zeta^{-\frac{c_i}{\sigma(1 + \tau_i)} - 1} \end{aligned} \quad (14)$$

With $C'(\cdot) = \frac{\partial C(\cdot)}{\partial \zeta}$. Plugging these expressions back into the differential equation, we get

$$\begin{aligned} D(\delta) &= \frac{1 - \sigma}{c_i} (rU_i - b_i) = D \\ C'(\zeta, \delta) &= \zeta^{\frac{c_i}{\sigma(1 + \tau_i)} - 1} \frac{1}{1 + \tau_i} \hat{F}_{n_i}(\zeta, \delta) = \zeta^{\frac{1 - \sigma}{\sigma} \frac{a_i}{1 + \tau_i}} \frac{1}{1 + \tau_i} \hat{F}_{n_i}(\zeta, \delta) \end{aligned}$$

We can integrate the latter equation to obtain

$$C(\zeta, \delta) = \frac{1}{1 + \tau_i} \int_0^\zeta x^{\frac{1 - \sigma}{\sigma} \frac{a_i}{1 + \tau_i}} \hat{F}_{n_i}(x, \delta) dx + \kappa(\delta)$$

Replacing in 14, we get

$$\hat{w}_i(\zeta, \delta) = \frac{1 - \sigma}{c_i} (rU_i - b_i) + \frac{\zeta^{-\frac{1 - \sigma}{\sigma} \frac{a_i}{1 + \tau_i} - 1}}{1 + \tau_i} \left[\int_0^\zeta x^{\frac{1 - \sigma}{\sigma} \frac{a_i}{1 + \tau_i}} \hat{F}_{n_i}(x, \delta) dx + \kappa(\delta) \right]$$

In order to pin down the integration constant $\kappa(\delta)$, we assume that $\lim_{\zeta \rightarrow 0} \zeta \hat{w}_i(\zeta, \delta) = 0$, in a manner similar to CMW. This assumption means that payroll goes to zero as firm size decreases while keeping the ratio of skilled to unskilled workers constant, and it is valid as long as marginal productivities do not increase too fast as the number

of workers goes to zero (technically, faster than $1/\zeta$ as $\zeta \rightarrow 0$). This is the case for the CES-like production function we use in our quantitative exercises. Then, the equation above implies $\kappa(\delta) = 0$.

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

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

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

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

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

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

Also, we further impose that

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

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

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

We only need one solution, the simplest being

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

since $1/\chi_{i,j} = \chi_{j,i}$. Then, using $\partial F/\partial n_j = \chi_{j,i} n_j^{\chi_{j,i}-1} \partial \tilde{F}/\partial M_{i,j}$ and $\partial F/\partial n_i = \partial \tilde{F}/\partial M_{i,i}$ as $\chi_{i,i} = 1$, the system 12 can be rewritten as

$$c_i \tilde{w}_i(M_{is}, M_{iu}) = (1-\sigma)(rU_i - b_i) + \sigma \left[\frac{\partial \tilde{F}(M_i)}{\partial M_{ii}} - (1+\tau_i) \left(M_{is} \frac{\partial \tilde{w}_i(M_i)}{\partial M_{is}} - M_{iu} \frac{\partial \tilde{w}_i(M_i)}{\partial M_{iu}} \right) \right] \quad (15)$$

System 15 is equivalent to system 12 in the case where $\chi_{i,j} = 1$. Thus, the solution for $\tilde{w}_i(M_{is}, M_{iu})$ is known:

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

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

coordinate system, we obtain

$$w_i(n_s, n_u) = \frac{1 - \sigma}{c_i}(rU_i - b_i) + \frac{1}{1 + \tau_i} \int_0^1 \epsilon^{\frac{1-\sigma}{\sigma} \frac{a_i}{1+\tau_i}} \frac{\partial F \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i} d\epsilon \quad (16)$$

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

$$\frac{\partial w_i(n_s, n_u)}{\partial n_j} = \frac{1}{1 + \tau_i} \int_0^1 \epsilon^{\frac{a_i}{1+\tau_i} \left(\frac{1-\sigma}{\sigma} + \frac{1+\tau_j}{a_j} \right)} \frac{\partial^2 F \left(\epsilon^{\frac{1+\tau_s}{a_s} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tau_u}{a_u} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i \partial n_j} d\epsilon \quad (17)$$

To compare the solution we found to that in CMW, write $\tilde{\sigma}_i = \frac{\sigma(1+\tau_i)}{\sigma(1+\tau_i)+(1-\sigma)a_i} = \frac{\sigma(1+\tau_i)}{c_i}$. Then, equation 16 can be stated as

$$a_i w_i(n_s, n_u) = (1 - \tilde{\sigma}_i)(rU_i - b_i) + \frac{a_i}{1 + \tau_i} \int_0^1 \epsilon^{\frac{1-\tilde{\sigma}_i}{\tilde{\sigma}_i}} \frac{\partial F \left(\epsilon^{\frac{1+\tilde{\sigma}_i}{\tilde{\sigma}_i} \frac{a_i}{1+\tau_i}} n_s, \epsilon^{\frac{1+\tilde{\sigma}_i}{\tilde{\sigma}_i} \frac{a_i}{1+\tau_i}} n_u \right)}{\partial n_i} d\epsilon \quad (18)$$

This expression is very similar to the solution in CMW, except for the terms a_i and $a_i/(1 + \tau_i)$. Consider the case where $\alpha_i = 1 + \tau_i$: the valuation of formal benefits by workers is exactly equal to the total costs incurred by firms. In this case, $\tilde{\sigma}_i = \sigma$, and the only difference between our solution and that in CMW is a term a_i multiplying w_i on the left-hand side. This factor accounts for the fact that the "true" wage in this economy is $(1 + \tau_i)w_i = a_i w_i$, which is both the value that firms pay and how workers value total compensation.

If $\tau_i \neq a_i - 1$, then there is a wedge between firm disbursements and the valuation of total pay by workers, and $\tilde{\sigma}_i \neq \sigma$. Note that this does not mean that the share of rents appropriated by workers is different; instead, this is an adjustment inside the integral term to compensate for the term $a_i/(1 + \tau_i)$ outside the integral, keeping the Nash bargaining equation valid. However, even in the case where σ is the same for all workers, we can have $\tilde{\sigma}_i \neq \tilde{\sigma}_j$. This would lead to nontrivial interactions between different types of labor, similar to how heterogeneity in bargaining power affects wages in CMW.

Finally, note that although we have assumed the same bargaining power for all workers, it is easy to extend the model to the more general case with type-specific bargaining power. This would lead to an expression similar to 18, but with $\tilde{\sigma}_i = \frac{\sigma_i(1+\tau_i)}{\sigma_i(1+\tau_i)+(1-\sigma_i)a_i}$. Similarly, extending the solution to more than two types of workers would be trivial, essentially requiring a simple change in notation. See CMW—in particular, how they define the matrix $\mathbf{NA}_i(z)$.

D.1 Minimum Wages and Wage Bargaining

The solution we found above for the wage bargaining differential equation, $w_i(n_s, n_u)$, does not take into account the possibility of a minimum wage. If we set a rule that constrains wages to no less than a constant value, then the previous solution is only correct in the interior of the subset of the (n_s, n_u) space in which the minimum wage is less than the freely bargained wage. For other values of (n_s, n_u) , the minimum wage binds for the skilled, unskilled, or both.

Figure A.1 shows an example of how wages can be affected by the minimum wage according to firm size. For small values of n_s and n_u , marginal productivities are high, and bargained wages are above the minimum wage. As

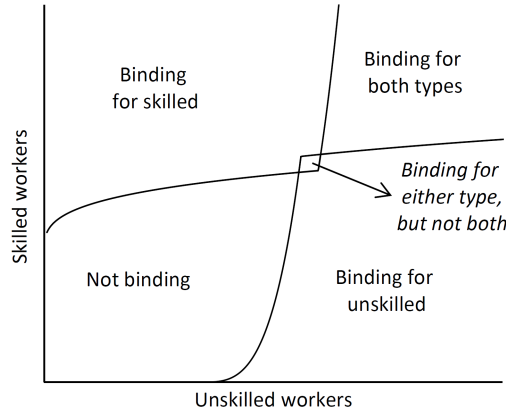


Figure A.1 – Minimum Wage Status According to Firm Size

the quantity of either type of worker increases, it is possible that marginal productivities decrease so much that the minimum wage binds. For high values of both inputs, it is possible that all wages equal the minimum wage. In this example, the curves are upward-sloping because there is complementarity between labor types ($\frac{\partial^2 F^z(n_s, n_u)}{\partial n_s \partial n_u} > 0$). They would be straight or downward-sloping if the cross derivative were null or negative, respectively.

It is also possible that for certain values of (n_s, n_u) , there is a multiplicity of wages satisfying the bargaining conditions: either type of worker might receive the minimum wage, but not both. This pathology is caused by discontinuities in the marginal value of workers, which we discuss below. In our applications, there is no possibility that the minimum wage binds for the skilled, no matter how many workers of this type are hired. The reason is that the first term in the wage equation 16, related to the reservation wage, is strictly greater than the minimum wage in all simulations. Hence, we are not concerned about this multiplicity problem.

If the minimum wage binds for only one type of worker, the unconstrained solution for the other type is no longer adequate. This is because contrary to the implications of the wage bargaining differential equation, marginal changes in the amount of the unconstrained type do not affect wages of the constrained type. From now on, for ease of exposition and focusing on our empirical application, we restrict attention to the case in which the minimum wage binds for unskilled workers but not for skilled workers.

To find the correct skilled-wage function in this case, we observe that the differential equation 11 simplifies to

$$c_i w_s(n_s, n_u) = (1 - \sigma)(rU_i - b_i) + \sigma \left[F_s(n_s, n_u) - (1 + \tau_s) n_s \frac{\partial w_s(n_s, n_u)}{\partial n_s} \right] \quad (19)$$

as the term $\frac{\partial w_u(n_s, n_u)}{\partial n_s}$ is set to zero. This is a univariate differential equation in n_s , similar to 13. The solution is analogous:

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

Note that skilled wages are still a function of the number of both skilled and unskilled workers, but not the same function as before. When the cross derivative of the production function $\frac{\partial^2 F^z(n_s, n_u)}{\partial n_s \partial n_u}$ is positive, as in our quantitative exercises, then we should expect this new wage function to be strictly greater than the unconstrained

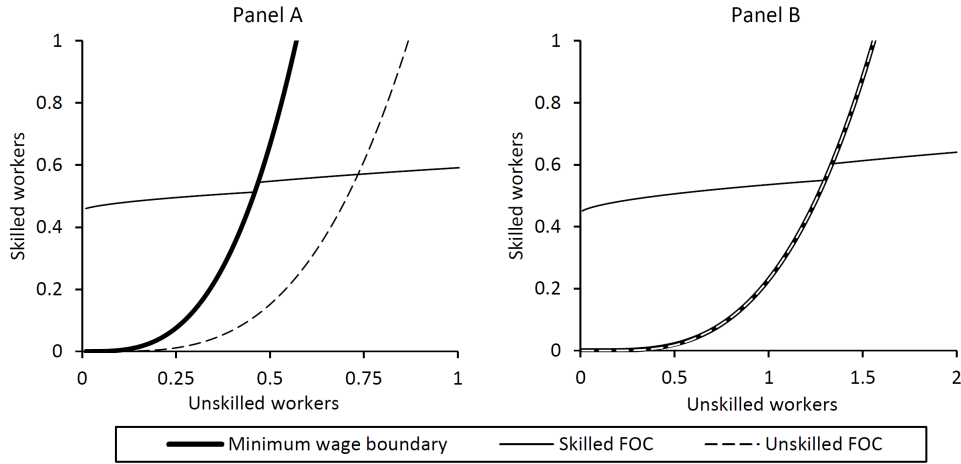


Figure A.2 – The Problem of the Firm and Minimum Wages

one for the same values of n_s and n_u . The reason is that in the unconstrained case, hiring an additional skilled worker leads to an increase in unskilled wages due to the effect on unskilled marginal productivities, which reduces the surplus being bargained over (from the point of view of the firm and the single skilled worker with whom it is bargaining). This "negative" effect does not exist (at the margin) when the minimum wage binds: the surplus is bigger, and so are bargained wages. Note that this implies a discontinuity in the wage function at the points that separate the regions where the minimum wage is or is not binding.

D.2 Minimum Wages and the Solution to the Problem of the Firm

Finally, we discuss how the existence of the minimum wage might change the problem of choosing the optimal firm size. The discontinuity in the wage function, discussed above, is caused by discrete changes in the net marginal value of workers $\pi_i^{for}(\cdot)$ (see equation 2) at the boundary of region of the (n_s, n_u) space where the minimum wage is binding. This discontinuity might lead to cases in which there is no exact solution to the firm's first-order condition, equation 3. We continue to restrict attention to the case in which the minimum wage binds only for unskilled workers.

In figure A.2, we show how the minimum wage can affect the problem of the firm. In Panel A, we illustrate the problem of a formal firm with average productivity ($z = 1$) in our baseline calibration. The heavy solid line marks the transition between a non-binding and a binding minimum wage for unskilled workers—that is, it is the vertical line in figure A.1. The other lines are the optimality conditions for the number of skilled and unskilled workers (equation 3). The solid line marks the combinations of (n_s, n_u) in which the marginal value of a skilled worker, $\pi_s^{for}(n_s, n_u)$, is equal to the expected search cost $\frac{\xi_s^{for}}{q(\theta_s)}$. Above this line, there are too many skilled workers, which drives down their marginal productivity and makes the marginal value less than the search cost. The same reasoning is valid for the dashed line: to the right of it, the marginal value of unskilled workers is less than the expected search cost, and the converse is true to the left of the line. As before, the upward slope of all curves comes from complementarity between labor inputs.

The unique solution to the problem of the firm in Panel A is the point where the two first-order conditions are

satisfied. Since this point is to the right of the heavy solid line, the minimum wage is binding at the optimal firm size. Note that there is a discontinuity in the skilled worker's first-order condition as it crosses the minimum wage boundary. Since the marginal value of skilled workers increases when the minimum wage binds for the unskilled, it becomes optimal to hire more skilled workers immediately to the right of the boundary. There is a similar discontinuity in the value of the unskilled worker, but in the opposite direction: to the right of the boundary, hiring an additional unskilled worker no longer benefits the firm by bringing down unskilled wages. However, in this case, the discrete decrease is not enough to reduce the marginal value of the unskilled to below the search cost. This is why the dashed line lies to the right of the minimum wage boundary.

Panel B describes a case in which there is no solution to the problem of the firm FOCs because of the discontinuities associated with the minimum wage. It follows from a change in the baseline model that increases overall productivity (parameter A in the quantitative experiments section), making the minimum wage bind by a smaller margin. The difference between Panel B and Panel A is that the discrete fall in the marginal value of the unskilled workers causes it to drop from a number strictly greater than the expected search costs to one strictly less. As a consequence, there is no point in the graph at which the unskilled first-order condition is satisfied. Neither is the skilled first-order condition is satisfied at the intersection of the three lines.

In such a situation, the firm would strategically choose a point to the left of that intersection (where the minimum wage does not bind) since, there, bargained wages for skilled workers would be discontinuously lower than immediately to the right of the intersection. There is no similar discontinuity in the unskilled wage because it cannot drop below the minimum wage, and thus, unskilled wages are approximately equal on both sides of the boundary. In our numerical applications, the optimal firm size in those situations is chosen by finding the point (n_s^*, n_u^*) that satisfies the first-order condition for skilled workers and lies immediately to the left of the discontinuity.

Note that in the absence of the minimum wage, we would expect the firm to hire more unskilled workers since the dashed line would lie to the right of the heavy solid line. Whether the firm would hire more or fewer skilled workers depends on the degree of complementarity between the two types of labor in the production function.

E Appendix: Numerical Solution of the Model

This appendix describes how to find the four endogenous equilibrium variables—labor market tightnesses θ_s, θ_u and unemployment values U_s, U_u —given a complete parameterization of the model. Along the way, it also shows how to simulate moments from the model.

Throughout the appendix, we refer to MATLAB code that is posted online at Daniel Haanwinckel’s website (currently located at <https://sites.google.com/view/haanwinckel>). All of the procedures described here are implemented in the `Matlab/Model.m` file. The model is implemented as a MATLAB class, which is a structure that contains both data (e.g., parameters) and procedures/functions (e.g., `solveForEquilibrium()` or `plotFirmProblem()`). The first parameter in these functions—almost always named `this` in the code—are references to the model object from which parameters should be taken. This is useful for three reasons. First, it eliminates the need to use model parameters as function arguments. Second, it also makes it easy to store, modify, and document a series of auxiliary parameters, such as tolerance levels for optimization. Third, the class can also easily store auxiliary variables used for speeding up certain parts of the code, as described below.

When calling these functions, this first parameter is not passed along with the other ones inside the parentheses. Instead, you should use dot notation, as the following example illustrates:

```
m = Model(); %Create a model object with default parameters
eq = m.solveForEquilibrium(); %Solve the model (m is the “this” parameter)
m.plotFirmProblem(eq); %Plots optimal employment and wage values in the equilibrium
m_highMW = m.clone(); %Creates another model object identical to m...
m_highMW.mw = m.mw * 1.1; %... but with a 10% higher minimum wage
eq_highMW = m_highMW.solveForEquilibrium(eq.rU,eq.q); %Solve using the previous equilibrium as
a starting point
disp(['Inf. rate change: ' num2str(eq_highMW.informality-eq.informality)]);
```

E.1 Problem of the Firm in the Informal Sector

The function `Model.firm_problem_inf(·)` solves the problem of the informal firm for one particular productivity level z . It takes as parameters a grid id number—which determines a z level, according to the `Model.z_grid` property—as well as the reservation wages rU_i , the vacancy filling probabilities $q(\theta_i)$, and functions `ini_log_n(·)` that determine initial points for each z . The z grid has 10 atoms, distributed uniformly in log space between 1 and 10,000 (that is: $10,000^{(x-1)/9}$ for $x \in \{1, \dots, 10\}$).

The function loops over all values of the discrete distribution of the informality penalty. For each, it attempts to solve the first-order conditions, choosing the log of skilled and unskilled employments using MATLAB’s `fsolve`, with tolerances `TolX` and `TolFun` set to 10^{-15} . In case of failure, a new starting point is chosen randomly (uniform shifts in the $[-20,20]$ range for each skill level). A total of 50 attempts are allowed, though the problem is almost always solved in the first attempt for reasonable parameterizations. The procedure returns the optimal employment values and corresponding wages for each informality penalty value, as long as the maximum error in the FOCs is 10^{-10} or less. Otherwise, it returns NaN. That prevents the program from using invalid solutions as if they were

correct.

E.2 Problem of the Firm in the Formal Sector

The problem of the formal firm differs from that of the informal one because of the minimum wage. There are three possible cases discussed below. The `Model.firm_problem_for(.)` function takes the same parameters as `Model.firm_problem_inf(.)`, solves all three cases, and returns employment and wage values in 2×3 matrices.

The main difference, relative to the informal problem, is that because payroll taxes and benefits vary by skill, the integrand in the Stole–Zwiebel bargaining formula is not homothetic. Thus, each evaluation of the first-order conditions requires numerical integration. We use trapezoidal integration via MATLAB’s `trapz` function. In that integration, a series of intermediate calculations are common across productivity levels, regardless of firm productivity or chosen employment values. To speed up calculations when solving for equilibrium, the computations are executed only once for every evaluation of the equilibrium objective function (defined below). These calculations are in the function `Model.calculate_auxiliary_params`, and the results are stored in the `Model.fpAux` structure. The speed gains are substantial.

As with the informal problem, numerical tolerances for MATLAB’s `fsolve` are set to 10^{-15} , and the output is considered a success as long as the maximum error in the FOC is 10^{-10} or less.

E.2.1 Unconstrained Case

The integrands of the numerical procedure are defined in `Model.integrand_for_s_unc(.)` and `Model.integrand_for_u_unc(.)`, and the integration procedure is executed in `Model.upsilon_for_Fi_unc(.)`. `Model.foc_for_unc(.)` returns the FOC error as functions of z , employment choices, and general equilibrium variables.

E.2.2 Strictly Binding Minimum Wages

The algorithm assumes that the minimum wage only applies to low-skill workers. In this case, `Model.integrand_for_strict(.)` provides a 2×1 integrand. The first element is integrated to obtain the left-hand side of the skilled FOC, following the structure of the unconstrained case, where results follow the general structure of Cahuc, Marque and Wasmer (2008), equation 12. The second element provides the derivative of skilled workers’ wages with respect to low-skill employment. The FOC is returned in the `Model.foc_for_strict(.)` function.

E.2.3 Strategic Solution

As discussed in the main text, when the minimum wage is marginally binding, we assume that the selected employment values are such that the skilled FOC holds, and the bargained unskilled wage is exactly the minimum wage. This is implemented in the `Model.foc_for_strategic(.)` function. It is important to note that when the unskilled reservation wage is too high, it becomes impossible to satisfy the second condition. The `Model.firm_problem_for(.)` function verifies whether that condition is true and, if it is, does not attempt to solve the strategic case.

E.2.4 Diagnostics

The `Model` class provides a `plotFormalProblem(·)` function that can be useful for verifying the issue that leads to strategic solutions. It makes plots similar to those in Appendix D, with two types of first-order conditions (unconstrained and constrained) and the curve that determines the region where the minimum wage binds.

E.3 Solving the Problem for all Firms and Determining Solution Types

The `Model.allFirmProblems(·)` function takes equilibrium variables and starting points as parameters and returns a structure that defines the solution of the firm problem for all firms in the economy. Productivity values z are floating-point numbers ranging from 1 to 10,000 (meaning that holding employment constant, the most productive firm in the economy would produce 10,000 times as much as the least productive one). The problem of the firm is solved in a grid of size 10, with values equally distributed in log space.

The first step of `Model.allFirmProblems(·)` is to solve the problem of the firm for all z levels in the grid. Next, these solutions are interpolated using MATLAB's `interp1` function, using the `pchip` option (shape-preserving piecewise cubic interpolation). The following step is to determine which formal solution regime applies for each z . When the reservation wage rU_u is too high, constrained regimes will never be optimal, so the unconstrained solution is selected. When that condition does not hold, the program proceeds to find thresholds in z space that separate regimes. The lowest z where the solution is unconstrained is such that the freely bargained unskilled wage, when evaluated at the solution of the unconstrained firm problem, is exactly the minimum wage. To find that number, we use `fsolve`, and if that fails, a binary search procedure is used (defined in the `binarySearch.m` file). Tolerances in this procedure are set to 10^{-16} . The highest z where the solution is strictly binding is such that the bargained unskilled wage, when evaluated using the constrained employment values, is equal to the minimum wage. As in the case above, that value is found using `fsolve` or a binary search. In both cases, the algorithm also verifies whether each of these regimes encompass all or none of the firms by evaluating those conditions at the extreme values of the z distribution. Finally, functions for formal employment values and wages are provided. These functions verify which regime holds for the requested z and return the corresponding values.

The final step is to create one function that returns instantaneous profits for formal or informal firms and another that returns the formality status by comparing instantaneous profits in each sector.

E.4 Aggregating Over Firms

The function `Model.equilibriumConditions(·)` takes equilibrium variables and starting points as parameters and returns the conditions that define general equilibrium in the model. It also returns a structure that contains the solution of the problem of the firms and some aggregate measures of the economy that were calculated as intermediate steps to find the equilibrium conditions. Another function, `Model.addEquilibriumStatistics(·)`, takes the equilibrium object as an argument and calculates a series of other useful aggregates.

Calculating those aggregates is done by integrating over the distribution of firms, where integrands are employment and wage values for the corresponding firm types for all combinations of z and the informality penalty. In that integration, the process loops over the discrete values of the informality penalty. For each one, it defines a series of

thresholds in z space that define firms that are marginal between formality and informality, and, within the formal sector, points that separate the different solution regimes (unconstrained, strategic, and strictly binding).

To find all z values corresponding to firms that are indifferent between sectors, we first define a grid of 10,000 points evenly distributed in $\log(z)$ space and check the formality status at each point. Whenever two consecutive points in the grid correspond to different formality decisions, we run a binary search procedure to find the marginal value. Note that we do not assume only one transition (from small informal firms to large formal firms) because such an assumption is not a general prediction of the model for all parameter values (though it holds for the estimated parameters). The points that separate the formal regimes were calculated in step 1.3, above.

Within each z interval corresponding to a single formality decision and formal regime, the integrations are performed using MATLAB’s integral function, with tolerances set to 10^{-15} .

When calculating additional measures that are important for estimation but not needed in the equilibrium-finding procedure, the steps are similar. However, the segments in z space are now also separated according to firm-size categories for firms of each sector—5.5, 10.5, 100.5, and 500.5. This is so that the model can calculate moments such as share of formal workers in firms with 11 or more workers (here interpreted as 10.5 or more). Those calculations are performed in the `Model.addEquilibriumStatistics` function.

E.5 Equilibrium-Finding Procedure

The goal of this procedure is to find the values of unemployment U_i and labor market tightnesses θ_i that are consistent with general equilibrium conditions of the model. The choice variables are $\log(rU_i)$ and $\log(q(\theta_i))$, respectively. The objective function for this optimization is defined in the `Model.equilibriumObjFun(·)` function. It returns the equilibrium errors obtained using `Model.equilibriumConditions(·)`, as well as the corresponding Jacobian (calculated with forward finite differences). The numerical Jacobian is explicitly provided for efficiency. When re-solving firm problems following marginal changes in equilibrium parameters, the optimal solutions for the baseline equilibrium variables are used as starting points.

The function `Model.solveForEquilibrium(·)` attempts to find an equilibrium and, if it does, returns an equilibrium structure with predicted aggregate moments and solutions of firm problems. Equilibria are searched for using MATLAB’s `fsolve`, with tolerance options set to 10^{-10} . A solution is considered valid if the equilibrium conditions are met to 10^{-8} precision. If they are not, a maximum of 9 reattempts are made, shifting the minimization algorithm (`trust-region-dogleg` in odd-numbered attempts, `levenberg-marquardt` in even-numbered attempts) or the starting point (using shocks to the base starting points drawn from a uniform distribution over [-2,2]).

While the `Model` class provides default starting values, `Model.solveForEquilibrium(·)` accepts additional parameters defining initial reservation wages and vacancy-filling probabilities. It also accepts a “reference equilibrium” parameter which, if provided, sets the initial points for firm problems based on that equilibrium’s solutions. This feature is useful when calculating how equilibrium properties change with marginal shocks to parameters.

F Appendix: Estimation and Counterfactuals

This section describes the estimation procedures in detail. As with Appendix E, it also serves as an overview of the code that accompanies the paper.

Step 0: Read PME data using Datazoom

The main data source used in the paper is the Brazilian’s Monthly Employment Survey (Portuguese acronym: PME), conducted by the Brazilian’s Statistics Bureau (IBGE). The microdata is publicly available on the IBGE website. To clean the data, we use Datazoom, a Stata program developed by PUC-Rio’s department of economics.³⁴ To run the code in Step 1, put the clean data files generated by Datazoom in the `Stata/Panel` folder.

Step 1: Estimate Dynamic Moments and Means from the PME

The file `Stata/getMoments.do` contains code to obtain raw data moments used in the estimation procedure, starting from Datazoom’s clean data files. The outputs of that code are `.csv` files with region–education-group level statistics to be used in the following steps of the estimation procedure. The code takes two parameters:

- `year`, either 2003 or 2012;
- `bootstrapId`, which is either 0, for obtaining moments from the actual sample, or a positive number, if one wishes to resample with replacement before calculating the moments (see details in the Bootstrap step below).

The `Stata/runAll.do` code contains a loop to call `getMoments` for all combinations of year, `bootstrapId`, and `singleRegion` needed for the paper. Results are stored in the `CleanData` folder.

Here is a list of procedures in `getMoments.do`, in the order they are executed:

1. Load the Datazoom files for the corresponding year and keep only data for the months of April through December. This provides the longest comparable periods of both years under a constant minimum wage. If `bootstrapId` is not 0, set the random number generator seed to a number that depends on this argument, and then resample individuals in the data (including their whole industry), within regions.
2. Keep only individuals who are either salaried workers (formal or informal) or unemployed. Mark observations with total work hours of less than 5 or more than 70 as missing for that variable. Mark hourly wage values below the 0.5 and above the 99.5 quantiles as missing data. In both cases, the goal is to reduce the impact of potential measurement error. These workers will not be used in wage regressions but might still provide useful information regarding shares of the workforce in each education group or transition probabilities.
3. Redefine wages as multiples of the 2003 minimum wage and create auxiliary variables to be used in the next steps, including firm-size categories by formality status, educational group, age group, and employment status. Education and wage groups are defined according to their values as of the first time an individual is observed, so individuals belong to only one group in all periods they are observed in.

³⁴The PUC-Rio Econ website is currently located at <http://www.econ.puc-rio.br>. Datazoom’s website is currently located at <http://www.econ.puc-rio.br/datazoom/english/>.

4. For each region:
 - (a) Restrict the data to observations in that region.
 - (b) Split workers into age–education groups according to the value of these variables as of the first time the individual is observed in the panel. Age groups are (i) 16 through 19 complete years; (ii) 20 through 24 complete years; (iii) 25 through 29 complete years; (iv) 30 through 59 complete years. Education groups are (i) less than 8 years of schooling; (ii) at least 8, but less than 11 complete years of schooling (complete primary and incomplete secondary); and (iii) at least 11 years of schooling (complete secondary or more).
 - (c) For each age–education group:
 - i. Run a regression with individuals of that group to obtain the formal wage premium, controlling for age (defined in months), time fixed effects (calendar months), individual fixed effects, and weighting by the PME sampling weights. The regression sample only includes workers with a single job in order to avoid measurement error in hourly wages, since only total hours worked is available.
 - ii. Run a similar regression where the main independent variables are firm-size indicators instead of formality status. One of these indicate being in a firm with 6 to 10 employees, and another indicates being in a firm with 11 or more employees. This regression includes the same controls and weights as above but does not include the formality dummy.
 - iii. Get the average probability of transitioning from unemployment to employment for that education group–region combination.
 - iv. Use collapse to get other region–education-level statistics, such as mean log wages, formality rates, shares of the salaried workforce in each firm-size category, and total counts, all weighted by PME sampling weights.
 - (d) Save the region–education statistics and regression results in temporary files.
5. Create auxiliary variables to be used in the estimation of job destruction rates. These include tenure, indicators for censoring in tenure (individuals still employed at the last time they are observed), and worker-level mean log wages and mean age relative to the mean for workers in the same region and educational group.
6. Estimate a Cox proportional hazards model for the probability of transitioning from salaried positions to unemployment. Use as controls region fixed effects, age–education group fixed effects, *relative* age, and a cubic polynomial in *relative* mean log wage. Here, *relative* means the individual value minus the mean for workers in the same region–age–educational group. The idea is to obtain coefficients for a “typical” worker in each region–age–education group, using the relative measures to net out individual variation in within-group productivity. That variation, which does not exist in the model, could introduce a bias in the effect of informality on job destruction rates if (i) more productive workers are more likely to be formally employed, and (ii) more productive workers have lower job destruction rates.
7. Reconstruct region–age–education–sector-level job destruction rates (where sector is formal or informal) based on the proportional hazard coefficients estimated above. Normalize the level of these estimates such that the

Table A.4 – Estimated Shares of Skilled Workers by Age–Education Cell

Age Groups	Education Groups (by Years of Schooling)		
	0-7	8-10	11+
16-19	0.00	0.10	0.30
20-24	0.15	0.24	0.50
25-29	0.26	0.39	0.74
30-59	0.35	0.50	1.00

mean job destruction rate in the salaried population matches the mean probability of job destruction in the sample. Add those job destruction rates to the previously saved region–age–education-level statistics and save the final results in the `CleanData` folder.

Step 2: Data Reduction: from Age-Education to Skill

This appendix explains in detail how we identify skilled and unskilled workers in the data. This step yields moments at the region–skill level, which will be used as targets in Step 4 below (the minimum distance procedure). It also yields two sets of parameters: shares of skilled workers in each region and job destruction rates.

This procedure is independent of the structure of the model. It relies only on a small set of assumptions:

1. There are two latent worker types, skilled and unskilled; each one of the twelve age–education groups is a convex combination of skilled and unskilled workers, with skill shares in each group being common across regions.
2. The first group (16–19 years old, less than 8 years of schooling) is composed of unskilled workers only, and the last group (30–59 years old, 11 or more years of schooling) is composed of skilled workers only.
3. There are no job-to-job transitions, transition rates are continuously compounded, and job destruction hazards have the form $\lambda_{i,r}^j = \lambda_r \lambda_i \lambda^j$. If a worker is observed in a formal job in a month and in an informal job in the next, for example, the procedure assumes there must have been an interval between these two points in time when the worker was unemployed. This assumption ensures that estimated job destruction and job finding rates are comparable to their corresponding values in the structural model.

We solve an optimization problem where choice variables are skill shares in each age–education group (except the first and the last), job destruction parameters, job-finding rates by region–skill, and a series of region–skill-level variables corresponding to various labor market moments: mean log wages, wage premiums, shares of individuals by firm size, etc. The goal of this optimization problem is to minimize deviations between the region–age–education moments and the corresponding moments implied by the region–skill-level summary (obtained by averaging the values for each skill level using the skill shares in the relevant subpopulation). The sum of squared deviations are weighted by region size, defined as the share of all workers in the data located in each region. The composition of skills by age–education group generated by this procedure is shown in Table A.4.

This step provides estimates of the job destruction parameters and region-specific shares of skilled workers. These parameters are shown in Table 2, Panel A. As expected, job destruction rates are higher for unskilled workers

and for those in the informal sector. Additionally, the model finds that a little more than half of the workforce is skilled, with little variation across regions.

Let us now formalize these procedures and fill in details. There are three groups of choice variables in the minimization problem. The first contains the shares of skilled workers in the workforce for each age–education group except the first and the last, $\tilde{\eta}_g$ (where $g \in \{2, \dots, 11\}$). The second group contains parameters that determine instantaneous job transition rates: region shifters of the job destruction hazard, λ_r ; the relative hazard associated with being skilled, λ_s ; the increased hazard associated with the informal sector, λ^{inf} ; the region–skill-specific job-finding rate, $(\theta q)_{r,i}$; and region–skill-specific shares of vacancies posted by formal firms, $\phi_{r,i}$. The third group contains some region–skill-specific moments, such as mean log wages and the formal wage premium. The full list is provided below.

The first step in the objective function of the optimization procedure is to convert the instantaneous transitions implied by job destruction rates, job-finding rates, and shares of formal vacancies into monthly transition rates. This is done by exponentiating the instantaneous transition matrix appropriately. This step provides monthly transition rates from unemployment to employment, from informal employment to unemployment, and from formal employment to unemployment, all by region and skill group.

The second step is to use the transition rates and share of skilled vacancies to calculate the implied informality rate for each skill group in each region, using the following expression:

$$\frac{(1 - \phi_{r,s}) / \lambda_{r,i}^{inf}}{\phi_{r,s} / \lambda_{r,i}^{for} + (1 - \phi_{r,s}) / \lambda_{r,i}^{inf}}$$

The third step is to calculate shares of skilled workers in particular subsamples of each age–education group: employed workers, formal workers, informal workers, and unemployed workers. This is important because when calculating a given statistic for an age–education group by taking a weighted average between skilled and unskilled workers, the appropriate weights are different according to the statistic. For example, employed workers are positively selected in terms of skill, while informal workers are negatively selected, relative to formal workers.

Finally, we can calculate the loss function. That is a sum of squared errors, where each error is a deviation between the region–age–education group and the value implied by the two-types model for a particular moment (indexed by l). Let the moment implied by the model be denoted by $M_{i,l,r}^{Model}$. The indices $i \in \{s, u\}$ denote skill and r denotes regions. The moments in the data are $M_{g,l,r}^{Data}$, where g denotes age–education groups. The deviations are weighted by the share of workers in each region–age–education group, $S_{g,r}$. Formally, the loss function is

$$L = \sum_{r=1}^6 \sum_{l=1}^{11} \sum_{g=1}^{12} S_{g,r} [M_{g,l,r}^{Data} - ((1 - w_{g,l,r})M_{u,l,r}^{Model} + w_{g,l,r}M_{s,l,r}^{Model})]^2$$

where $w_{g,l,r}$ are the appropriate mixing weights between skilled and unskilled workers for group g in region r when calculating the moment l . Below, we list the weight used for each moment l .

Note that while the estimated moments vary by regions, the shares of skilled workers in each age–education group are assumed to be the same in all regions. These estimates are also assumed constant over time in the counterfactual exercises. This is what gives empirical content to the concept of skill in the model.

The moments l are as follows:

- Mean log wage. For this l , the model moment $M_{i,l,r}^{Model}$ is a choice variable. The mixing weights $w_{g,l,r}$ are share of skilled workers among employed workers of that region–age–education group.
- Formal wage premium. As above, $M_{i,l,r}^{Model}$ is a choice variable and the weights are the share of skilled workers among employed workers.
- Firm-size wage premiums for the 6–10-worker and 11-or-more-worker categories, relative to the 1–5-worker category. The same applies regarding $M_{i,l,r}^{Model}$ and the weights.
- Informality rate. Here, $M_{i,l,r}^{Model}$ is implied by the transition rates and share of formal vacancies, so it is not an additional degree of freedom. The weights are for employed workers.
- Month-to-month transitions: unemployment to employment, formal employment to unemployment, informal employment to unemployment. As above, $M_{i,l,r}^{Model}$ are implied by transition rates and $\phi_{i,r}$. The weights are for unemployed workers, formal workers, and informal workers, respectively.
- Shares of formal and informal workers in each of three firm-size categories available in the PME: 1 to 5 employees, 6 to 10 employees, and 11 employees or more. $M_{i,l,r}^{Model}$ are choice variables, and the weights are for formal workers and informal workers.

To solve the minimization problem, we use MATLAB’s `lsqnonlin`, making log or logit transformations of the variables so that constraints are not needed. We use 5 sets of starting points and pick the best result from among them.

The first starting point has $\tilde{\eta}_g = 0.5$, $\lambda^u = 2$, $\lambda^{inf} = 4$, $(\theta q)_{r,s}$ set to the monthly transition from unemployment to employment for workers in the highest age–education group ($g = 12$), $(\theta q)_{r,u}$ set in the same way but using $g = 1$, $\phi_{r,i}$ set to the informality rate for groups 12 and 1. For the remaining choice variables, where $M_{i,l,r}^{Model}$ is a choice variable, we set $M_{s,l,r}^{Model} = M_{12,l,r}^{Data}$ and $M_{u,l,r}^{Model} = M_{1,l,r}^{Data}$. The other starting points shift the one above with shocks drawn from a standard normal distribution.

The last step in this first part of the estimation is to renormalize the employment shares by firm size so that they add up to one.

This process directly yields two sets of model parameters: the job destruction rates and the shares of skilled workers in each region, $\eta_r = \sum_{g=1}^{12} S_{g,r} \tilde{\eta}_g / \sum_{g=1}^{12} S_{g,r}$. The other moments estimated at the region–skill level are used as targets in Step 4, below.

All of the procedures above are coded in `Matlab/getDataSummary.m`. The code file `Matlab/createMomentsFile.m` runs that procedure. It also creates corresponding bootstrap estimates of the variance–covariance matrix for the moments used in the second part of the estimation, as described in detail in Step 6.

Step 3: Obtain Other Estimation Targets: Almeida–Carneiro Elasticity, Labor Share, and Share of Formal Workers in Very Large Firms

To obtain the effect of enforcement of labor regulations on the informality rate, we use Almeida and Carneiro’s methodology and data, which are publicly available at the article’s webpage.³⁵ For consistency with our definition of informality, the dependent variable is the informality rate among salaried workers. Other than that, we use regression specification identical to the main one in their paper, with state fixed effects and a series of municipality controls. The desired outcome is the instrumental variables estimate of the effect of enforcement on informality, where enforcement is measured as the share of workers in the municipality that were the target of labor inspections. The instrument is an interaction between the log number of inspectors in a state and the distance between each municipality and the nearest enforcement office.

The labor share of income is obtained from the Brazilian National Accounts System. Standard errors for this estimate are not provided. We circumvent this problem by using time-series variation in this number. More precisely, we assume that the labor share changes smoothly over time, and deviations from that smooth pattern should reflect sampling and aggregation errors within the calculation procedure. Then, we fit the labor share numbers from 1995 through 2008 in a polynomial in time. Specifically, we define the time variable as $t = year - 2003$, regress the labor shares in a polynomial in t , and use the standard error of the intercept as the standard error of the labor share estimate. The time series presents a clear convex pattern, and both the quadratic and cubic polynomials provide a tight fit. The standard errors are similar under both specifications, so we choose the largest of them (associated with the quadratic specification). Finally, we also consider non-diagonal terms in the eight rows of the covariance matrix to be zero.

The third national-level target is the ratio of formal workers in firms with at least 500 employees to formal workers in firms with at least 100 employees. That number is obtained from the CEMPRE tables. We have no measure of the precision of this statistic. But given that it comes from the universe of formal firms in the government registry, we believe this to be small enough to be negligible, compared to the other targets in the estimation. Nevertheless, in order to be conservative, we set its standard error to 0.02.

The file `Stata/getNationalMoments.do` performs the Almeida–Carneiro estimation and saves the estimate, along with the labor share, to a `.csv` file in the `CleanData` folder. It also has an option for providing bootstrap estimates, which is further described in Step 6 below.

Step 4: Minimum Distance Estimation of the Remaining Parameters

As described in the main text, the estimation of the remaining 39 parameters— $d_{s,r}$, $d_{u,r}$, A_r , T_r , P_r , P^D , α , γ , B_1 , B_2 , ξ_s , ξ^{for} , ξ^{inf} , ξ^S —is done by minimizing the distance between 123 moments in the data and their model equivalents. The moments are ten statistics at the region–skill level plus three national-level targets.

Let us denote the estimated parameters as a vector $x \in X \subset \mathbb{R}^{39}$, the moments calculated from the model as a function $h : X \rightarrow \mathbb{R}^{123}$, and the value of these moments in the population of interest as β . Under the assumption that there is a unique vector of parameters x_0 that minimizes $[\beta - h(x)]' W [\beta - h(x)]$, we can obtain an estimate

³⁵<https://www.aeaweb.org/articles?id=10.1257/app.4.3.64>

of x_0 by solving the following minimization problem:

$$\hat{x} = \operatorname{argmax}_{x \in X} \left[\hat{\beta} - h(x) \right]' W \left[\hat{\beta} - h(x) \right] \quad (20)$$

where $\hat{\beta}$ is an estimate for β , and W is a symmetric weighting matrix.

Under some assumptions, which include consistency of $\hat{\beta}$, differentiability of $h(\cdot)$, uniform convergence of the minimand, and that x_0 is the unique solution to $E \{H(x)'W[\beta - h(x)]\} = 0$ (where $H(x) = \nabla h(x)$ is the Jacobian matrix of h at x), \hat{x} converges to x_0 as the sample from which $\hat{\pi}$ is calculated increases in size.³⁶

The weights W are based on the shares of the workforce to which the statistics refers. For job-finding rates, informality rates, mean log wages, and the formal wage premium, we use the share of the national workforce in the corresponding region. For the firm-size wage premiums, we use the share of workers in that region in the lowest firm-size group (1 through 5 workers) plus the share in the firm-size group corresponding to the premiums (6–10 workers or 11+ workers). For the firm-size categories, the weights are the share of the national workforce in that formality status and firm-size category. For the labor share and the enforcement elasticity of informality, the weights are one. Finally, for the tail of the distribution of firm sizes among formal workers, the weight is the national share of formal workers.

Except for mean log wages and the wage premium variables, the moments are measured in relative terms—that is, the value divided by the moment in the data. This helps balance the relative importance of each moment such that estimation becomes scale-independent.

In that optimization problem, every function evaluation requires solving for the equilibrium six times, once for each region. Calculating the Jacobian of the objective function requires 84 additional computations of equilibria.³⁷ Solving each equilibrium is costly since it combines two layers of optimization (general condition and problem of individual firms), and the inner layer requires numerical integration in every function evaluation (for formal firms). Even using techniques to speed up the calculation of an equilibrium, described in Appendix E, the average CPU time consumed in each iteration is measured in hours, not seconds.

Two further issues compound the computational demands in estimation. First, the parametric space is large and the model complex, meaning that the objective function might have several local minima. Addressing that issue requires testing a large number of starting points in estimation. Second, the multistep nature of the estimation process suggests the use of bootstrapping for statistical inference.

A first step in tackling those challenges is to set convenient starting points in evaluations of the Jacobian. When evaluating the estimation loss function, and also when solving individual firm problems, we always use the same starting points for equilibrium variables. But when calculating numerical derivatives with marginal changes of

³⁶The asymptotic variance of \hat{x} is given by:

$$AVAR \left[\sqrt{N} (\hat{x} - x_0) \right] = [H(x_0)'WH(x_0)]^{-1} H(x_0)'WVWH(x_0) [H(x_0)'WH(x_0)]^{-1}$$

where V is the covariance matrix of the estimates $\hat{\pi}$. This matrix can be estimated by replacing $H(\hat{x})$ for $H(x_0)$ and \hat{V} for V , where \hat{V} is a consistent estimate for V . However, that formula does not account for sampling variation affecting the precision of parameters estimated in previous steps of the estimation procedure ($\eta_r, \lambda_{i,r}^j$). For that reason, we opt for a bootstrap procedure that includes all steps in the estimation.

³⁷Each of the national parameters requires recalculating all of the regions, and each of the region-specific parameters requires recalculating a single region. Note that this requires providing a user-defined Jacobian to the estimation objective function in MATLAB to avoid recalculation of every region's equilibria when a region-specific parameter is changed.

the estimated parameters, we take the starting points from the calculated equilibria. Since the individual changes in parameters are marginal (set to 10^{-5}), those starting points are close to the optimal points, accelerating the computations.

The second measure to make estimation manageable is to use a simplified version of the objective function, with one or two regions, when executing a thorough exploration of the parametric space. In that initial exploration, we check a large number of starting points to increase the likelihood that the minimum found is global. Then, having noticed that the best points are all located in a narrow region of the parametric space, we select starting points around that region for the main estimation procedure per se.

There, we use 100 processors performing independent explorations of the parametric space. Each processor samples 5 starting points, evaluates them, and uses the best one as the starting point in a gradient descent procedure. We use MATLAB's `fminunc` with the quasi-Newton algorithm, setting the maximum number of function evaluations and iterations to 10,000, `TolX` to 10^{-5} , and `TolFun` to 10^{-6} . In case of failure (e.g., an equilibrium could not be found for a certain region when evaluating the loss function or the Jacobian), the algorithm will proceed to the next best sampled point. Of the 100 processors, 50 are used for finding the solution for the baseline value of $\gamma = 0.4366$. The other 50 are split equally among the other two values of γ .

Step 5: Counterfactuals

The procedures are defined in the `CounterfactualMaker` class. We start by gathering data and model information corresponding to 2003 from the estimation outcomes. Next, we read the PME data for 2012, which has been preprocessed using the procedures described in Step 1. Subsequently, we perform the data reduction procedure in Step 2 for 2012, with the only difference being that the population shares in each age–education group are taken from the 2003 values instead of estimated. The other choice variables in the optimization procedure are the same.

To find the model value corresponding to 2012, we start by imputing observed changes to the share of skilled workers in each region (implied by changes in the shares of the workforce in each age–education group), payroll taxes, formal benefits, enforcement of regulation, and minimum wages, as described in the text. Next, we calculate target levels for skilled and unskilled mean log wages in 2012, so that the change in the model is the same as the one in the data. Finally, we use MATLAB's `fsolve` to find values of A and B_1 that make the model match these targets, setting the tolerances to 10^{-3} .

To calculate the robustness tests, we undo the change in education and use `fsolve` again to choose the corresponding choice variables to match the changes in log wages and informality rates, as described in the text.

After having the fully parameterized model in both periods, calculating any particular counterfactual is a matter of changing parameters and solving the model for equilibrium.

Step 6: Bootstrap

In the bootstrap procedure, we redo all estimation and counterfactual procedures (Steps 1–5), using the same code but resampled versions of the data. We perform 100 replications, each running on a separate computer core for speed gains.

For the PME data, we sample individuals within each region with replacement, keeping their whole history in the panel identical. The number of workers in each region is kept constant.

The Almeida–Carneiro data from which we obtain the effect of enforcement on inequality is at the municipality level. There, we resample municipalities within states, keeping the number of municipalities in each state constant.

For the labor share of income, we perform a parametric bootstrap. The draws come from a normal distribution with mean equal to the point estimate and standard deviation equal to the standard error computed in Step 3. The remaining national moment, pertaining to the tail distribution of formal firm sizes, is assumed to have little error compared to the other moments since it is calculated from administrative data on the universe of formal firms in Brazil. To be conservative, we assume that it has standard error 0.02 and perform a parametric bootstrap analogous to that of the labor share of income.

In the fourth step of each replication, we attempt to start the gradient descent method using a starting point very close to the point estimate. In case of failure, the algorithm will sample 10 points using random uniform $[-0.05, 0.05]$ shifts in all transformed choice variables and use them as starting points instead, starting with those with the lowest loss function.

Once the bootstrap replications are completed, the standard error of any function of estimated model parameters (including the parameters themselves and counterfactual measures) is obtained by taking the standard deviation of the corresponding function among the replications.

G Appendix: Comparative Statics Counterfactuals

In this appendix, we follow up on the quantitative exercise from subsection , which explored the determinants of the reductions in informality observed in Brazil between 2003 and 2012. We perform a complementary exercise here by considering one at a time the main exogenous changes observed during the period, as in a comparative statics exercise. In order to validate the performance of the model along these dimensions as well, we confront the comparative statics results with the empirical evidence currently available from reduced-form estimates.

Table A.5 presents the results, with each row describing a particular labor market outcome. The first column shows the changes observed in the data between 2003 and 2012. Each following column considers how changes in one parameter affect labor market outcomes in the model by comparing the baseline calibration with a new steady-state equilibrium where only the parameter in question is set to its 2012 level. As in the main text, since the effects of changes in payroll taxes and benefits were negligible, we omit them from the table.

Table A.5 – Comparative Statics, Changes in the Brazilian Labor Market between 2003 and 2012

Changes in:	(1)	(2)	(3)	(4)	(5)	(6)
	All changes	Changing one parameter at a time:				
		Fraction skilled $\Delta = 12.1\text{p.p.}$	Minimum wage $\Delta = 61.2\%$	Enforcement $\Delta = 33.9\%$	Productivity/Demand A B_1 $\Delta = 24\%$ $\Delta = 93\%$	
Outcomes						
Informality	-0.0603	-0.0641	0.1101	-0.0158	-0.0335	-0.0342
Unemployment	-0.0434	-0.0617	0.0407	0.0039	-0.0452	-0.0296
Wages (ln)	0.3437	0.0622	-0.0182	0.0024	0.1843	-0.0018
Skilled	0.0774	-0.1513	-0.0055	0.0003	0.2604	-0.0410
Unskilled	0.5152	0.1565	-0.1296	-0.0030	0.1759	0.1106
Formal Wage Premium	-0.0353	-0.0818	0.4261	-0.0007	-0.0956	-0.0687
Firm Size Premium (6–10)	-0.0065	-0.0147	-0.0274	0.0023	-0.0027	-0.0042
Firm Size Premium (≥ 11)	-0.0017	0.0035	0.0063	0.0019	0.0007	-0.0006

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 remained constant at their 2003 levels, except for one parameter (which is updated to its 2012 level).

Workforce Composition

In column 2 of Table A.5, we plug in the change of 12.1 percentage points corresponding to the increase in the share of skilled workers in the model and analyze the impact of labor market outcomes.

We find that the predicted changes are in line with our discussion from section 2. Both unemployment and informality decrease sharply as a consequence of a more skilled workforce, falling by 6.3 and 6.5 percentage points, respectively. Wages for unskilled workers increase by 15.8%, while they decrease by 14.8% for skilled workers. This is a direct consequence of the relative increase in the supply of skilled workers. The labor market for skilled workers becomes less tight (and the reverse happens for unskilled workers). Because firms hire more skilled labor in the new equilibrium, the productivity of unskilled work increases due to complementarities in the production function. The combination of a tighter labor market for unskilled labor and higher productivity is behind the steep increase in the informal wage. The same mechanisms discussed in the main text are also at work here.

We know of no reduced-form empirical study that analyzes the aggregate labor market effect of changes in the educational composition of the labor force. Various papers, such as Menezes Filho, Mendes and de Almeida (2004),

describe the strongly positive individual-level correlation between schooling and formality. Other papers, such as Barbosa Filho and Moura (2015), assume a stable individual-level relationship between schooling and informality and perform Oaxaca–Blinder type exercises that analyze the role of demographic changes as determinants of changes in informality. But no paper allows for the possibility that changes in the educational composition of the labor force directly affect labor market equilibrium outcomes, conditional on individual schooling. This highlights the relevance of the type of analysis conducted in this paper, where we can systematically address the endogenous labor market response to this type of compositional change.

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

Minimum Wage

The minimum wage increased by 61.2% from 2003 to 2012. The effects of a change of this magnitude in the calibrated model are shown in column 3 of Table A.5. The change in the minimum wage alone would have led to increases in informality of 11.3 percentage points and in unemployment of 4.3 percentage points. Wages of skilled workers would remain largely unchanged, while average wages of unskilled workers would fall by 13.4 percentage points due to the large increase in informality. The reason for this decline is the reduced demand for unskilled labor by formal firms, which increases unemployment and lowers the outside option of workers being hired by informal firms.

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

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

approximately reproduce the qualitative patterns documented by the empirical literature on minimum wages in Brazil.

Enforcement of Regulation

The third column in Table A.5 shows how a change in enforcement of 33.9% impacts our baseline calibration. First, informality decreases by 1.6 percentage points, as expected. We argued in section 2 that the effects of increased enforcement on unemployment are ambiguous in many models, and this is also true in ours. There is an extensive margin effect because firms that change their compliance decisions may hire more workers, and there is an intensive margin effect because the remaining informal firms hire fewer workers. In our calibration, unemployment increases by 0.4 percentage points with the increase in enforcement. The qualitative responses of informality and unemployment generated by the model are consistent with reduced-form evidence from exogenous variations in labor inspections provided by Almeida and Carneiro (2012). The only noticeable change in wages is a small decline in earnings for unskilled workers. In this respect, our model replicates the results found in Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015).

Productivity Changes

The exercise discussed in the main text calibrated changes in technology corresponding to, on average, 24% for A and 93% for B_1 . As mentioned there, these parameters are intended to capture aggregate changes affecting the Brazilian economy during this period, such as improvements in TFP (estimated to be around 22% by Ferreira and Veloso, 2013), terms of trade, and aggregate domestic demand (estimated to be, respectively, 77% and 81% by Messina and Silva, 2018). The comparative results related to these changes are presented in columns 5 and 6 in the table, respectively.

Unemployment declines by 5 percentage points, and wages rise by 18%, with the increase in A , which can be understood as a TFP, or aggregate demand, shock. There is also a reduction of 3.6 percentage points in informality, consistent with many other models where informal employment is countercyclical. This particular pattern generated by the model—with unemployment and informality being countercyclical, but the former responding more than the latter to changes in aggregated conditions—is also consistent with the empirical evidence for Brazil presented in Bosch and Esteban-Pretel (2012). Wages rise for all workers, but particularly for the skilled.

The change in B_1 also generates sizable reductions in informality and unemployment, both of the order of 3 percentage points, but almost no change in average wages. The constant average wages are a composite of declining wages for skilled workers (by 3.9%) and rising wages for unskilled workers (by 10.9%). This parameter likely captures the terms-of-trade shock experienced by Brazil during the 2000s, with an increase in the relative price of non-tradables (see the discussion in Messina and Silva, 2018).

H Appendix: Some Preliminary Evidence on Educational Composition and Labor Market Outcomes

This appendix provides some tentative empirical evidence on the relationship between the educational composition of the population and labor market equilibrium outcomes—particularly, informality and schooling. Since we could not find any empirical study focusing on this relationship and providing this type of evidence, we thought it would be useful to generate some preliminary results in this direction.

We use data from the 1991, 2000, and 2010 Brazilian censuses and consider microregions as the relevant definition of local labor markets. Microregions are sets of contiguous municipalities defined by the Brazilian Census Bureau (IBGE) that share similar geographic and socioeconomic conditions . This geographic unit has been repeatedly used in previous literature as the relevant definition of local labor markets in Brazil (see, for example, Kovak, 2013). In order to minimize heterogeneity, we focus on a sample of men between the ages of 20 and 50 who live in urban areas and are not in school.

Our goal is to analyze the relationship between educational composition and labor market equilibrium outcomes at the level of local labor markets. Therefore, for simplicity, the independent variable of interest is the share of individuals in the microregion with at least 8 years of schooling. The dependent variables are microregion formality or employment rates netted out of compositional effects. Specifically, the dependent variables are microregion fixed effects in individual-level regressions, run separately for each year, where the dependent variable is either formality status (among salaried workers) or an indicator of employment (among the entire sample). The individual-level regressions control for a quartic polynomial on age, dummies for race, and, in some specifications (as indicated in the table), dummies for educational levels.

The microregion regressions include as demographic controls the shares of the sample in two age categories (30–39 and 40–50) and the log of population (all calculated based on the sample used in the individual-level regressions explained above). To allow for differential trends across local labor markets with different initial conditions, we also control for an interaction of the initial (1991) formality rate (from the individual-level regressions that do not control for schooling) with year dummies. In some specifications, we also control for the shares of employment in eight broadly defined sectors (agribusiness and extractive industries, excluding mining; mining; manufacture; construction; utilities; retail; services; and government), as well as for interactions of year fixed effects with a set of initial (1991) socioeconomic characteristics (schooling, which is the independent variable of interest; average earnings; and employment, which is one of the dependent variables considered). All regressions include microregion and year fixed effects and are weighted by the inverse of the standard error of the dependent variable (obtained from the individual-level regressions). Standard errors are clustered at the microregion level.

The results from these regressions are presented in Table A.6. In the table, each coefficient corresponds to a different regression, with the rows indicating different specifications and dependent variables (controlling and not controlling for education in the first stage and using informality and employment as dependent variables). The columns correspond to different sets of controls, as indicated at the bottom of the table. All coefficients in the table refer to the same independent variable: the fraction of the population in the microregion with at least 8 years of schooling. We present results both controlling and not controlling for education in our “first stage” to distinguish

Table A.6 – Average Local Schooling and Local Formality and Employment, Conditional on Individual Schooling; Brazilian Microregions, 1991–2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Dep.Var.: Formality</i>							
No control for indiv. school.	1.151*** (0.0603)	0.306*** (0.0938)	0.407*** (0.0890)	0.283*** (0.0928)	0.367*** (0.0754)	0.389*** (0.0882)	0.356*** (0.0903)
Control for indiv. school.	0.955*** (0.0592)	0.203** (0.0947)	0.312*** (0.0895)	0.184** (0.0916)	0.285*** (0.0751)	0.294*** (0.0889)	0.222** (0.0913)
<i>Dep.Var.: Employment</i>							
No control for indiv. school.	0.280*** (0.0269)	0.0983* (0.0530)	0.0381 (0.0455)	-0.0254 (0.0573)	0.141*** (0.0485)	0.0254 (0.0455)	0.0585 (0.0455)
Control. for indiv. school.	0.149*** (0.0247)	0.0268 (0.0496)	-0.0247 (0.0430)	-0.0791 (0.0540)	0.0759* (0.0453)	-0.0367 (0.0430)	-0.0198 (0.0430)
<i>Fixed effects:</i>							
Microregion	No	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	No	Yes	Yes	Yes
State-Year	No	No	No	Yes	No	No	No
<i>Controls:</i>							
Demographic	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1991 Form. \times Year	No	No	Yes	Yes	Yes	Yes	Yes
Sectoral Shares	No	No	No	No	Yes	No	No
Avg. Earnings	No	No	No	No	No	Yes	No
1991 Other \times Year	No	No	No	No	No	No	Yes

Obs.: Standard errors (in parentheses) clustered at the microregion level; *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Data from the Brazilian censuses (1991, 2000, and 2010). Sample composed of males between ages 20 and 50, not in school, living in urban areas. Each number is the coefficient of the the share of individuals with at least 8 years of schooling from a different microregion-level regression (509 microregions, 1,527 observations). Dependent variables are microregion formality and employment rates, netted out of compositional effects (microregion fixed effects from individual-level regressions, run separately for each year, where the dependent variable is either an indicator of formality or employment, and independent variables are a quartic polynomial on age, dummies for race, and dummies for educational category, as indicated in the table). Demographic controls are the shares of the population in two age categories (30–39 and 40–50) and the log of population (both calculated with the sample used in the individual-level). 1991 Formality \times Year is the 1991 formality dependent variable (taken from the first-stage regression without individual schooling) interacted with year dummies. Sectoral shares are shares of the employed population in each of 8 broadly defined sectors (agribusiness and extractive industries, excluding mining; mining; manufacture; construction; utilities; retail; services; and government). 1991 Other \times Year include interactions of year fixed effects with 1991 levels of three other variables: the independent variable (schooling), average earnings, and the employment dependent variable. Regressions are weighted by the inverse of the standard error of the dependent variable (obtained from the individual-level regressions).

the individual-level association between schooling and formality from the equilibrium effect of the composition of the population on the incidence of formality, conditional on individual schooling.

The first two rows show that there is a robust correlation between the share of the population with at least 8 years of schooling and the formality rate. As expected, the coefficients are reduced in magnitude as we include microregion fixed effects and move from column 1 to 2, but they remain roughly stable across the various specifications between columns 2 and 7. This means that the correlation between the fraction of skilled individuals and formality is not related to differential trends across states or microregions with different initial characteristics, nor to overall patterns of development and growth (as reflected in demographic patterns, average earnings, or sectoral composition of employment).

The estimates in the first row do not control for individual schooling when calculating the conditional informality rate in the “first stage.” They therefore capture both the individual relationship between schooling and formality and the potential aggregate effect of the composition of the population on individual-level formality probabilities (through equilibrium labor market outcomes). The second row, in turn, controls for schooling in the “first stage,” so its results reflect the equilibrium response to changes in the educational composition of the population, conditional on individual-level schooling. The fact that the results from the second row are consistently significant indicates that the aggregate effects of the composition of the labor force on labor market equilibrium outcomes are indeed relevant. The relative magnitude of the coefficients across the first two rows would suggest that more than 60% of the aggregate correlation between educational composition of the population and informality may be due to these equilibrium effects, while less than 40% would be due to the direct relationship between schooling and informality at the individual level.

The magnitude of the point estimates in the first row (excluding column 1) are between 53% and 77% of the quantitative effects from the comparative statics exercise in column 2 of Table A.5. This difference may be due to the lack of a truly exogenous source of identification in our empirical results from this section, due to differences in the definitions of skilled workers, or due to different samples (most importantly, the exercise here uses all microregions in Brazil, while the calibration was conducted using data from the PME, which includes only the country’s main metropolitan regions).

The results related to employment, shown in the 3rd and 4th rows, are much less robust. Some specifications point to a positive and statistically significant relationship between educational composition and employment, but most results are small in magnitude and not statistically significant. Overall, we do not find a systematic relationship between educational composition of the population and employment rates. The problems alluded to in the previous paragraph could also be interfering with employment results, in addition to the fact that employment tends to be more sensitive to short-term changes in economic activity.

I Appendix: Additional Tables and Figures

This appendix contains the additional tables and figures mentioned in the main text.

I.1 Standard Errors for Tables from Section 4 (Quantitative Exercises)

Table A.7 – Standard Errors of Counterfactuals from Table 5 (Individual Contribution of Each Factor, Changes in the Brazilian Labor Market from 2003 to 2012)

Outcomes:	(1)	(2)	(3)	(4)	(5)	(6)
	All changes	All changes, except:				
		Fraction skilled	Minimum wage	Enforcement	Productivity/Demand A	B_1
Informality	0.0091	0.0367	0.0148	0.0079	0.0080	0.0077
Unemployment	0.0073	0.0161	0.0072	0.0065	0.0050	0.0054
Wages (ln)	0.0106	0.0999	0.0080	0.0096	0.0138	0.0103
Skilled	0.0093	0.1483	0.0095	0.0094	0.0100	0.0104
Unskilled	0.0152	0.0409	0.0152	0.0151	0.0259	0.0189
Formal Wage Premium	0.0107	0.0332	0.0232	0.0107	0.0156	0.0113
Firm Size Premium (6 – 10)	0.0020	0.0061	0.0029	0.0031	0.0035	0.0022
Firm Size Premium (≥ 11)	0.0041	0.0187	0.0080	0.0040	0.0083	0.0064

Table A.8 – Standard Errors of Counterfactual from Table 6 (Hypothetical Policy Experiments)

Outcomes	(1)	(2)	(3)	(4)
	1 p.p. reduction in payroll tax $\Delta\tau = -0.01$	Progressive payroll tax $\tau_s = 0.7143$		Doubling transfer to unskilled $\Delta b_u^F = 0.05$
		$\Delta\tau_u = -0.01$	$\Delta\tau_u = -0.10$	
Informality	0.0004	0.0003	0.0014	0.0009
Skilled	0.0005	0.0004	0.0005	0.0019
Unskilled	0.0030	0.0030	0.0047	0.0048
Unemployment	0.0019	0.0019	0.0020	0.0021
Skilled	0.0004	0.0004	0.0004	0.0032
Unskilled	0.0067	0.0067	0.0071	0.0100
Wages (log)	0.0011	0.0011	0.0014	0.0021
Skilled	0.0005	0.0002	0.0004	0.0015
Unskilled	0.0003	0.0002	0.0017	0.0008
Labor Share	0.0005	0.0005	0.0006	0.0008
Govt Surplus	0.0006	0.0005	0.0029	0.0025
Net Output	0.0004	0.0001	0.0005	0.0017

I.2 Main Results with Alternative Values of γ

Table A.9 – Model Fit and Validation with $\gamma = 0.2$, Country Averages, 2003 and 2012

Outcomes	Fit in 2003		Validation: $\Delta 2003-2012$	
	Data ₂₀₀₃	Model ₂₀₀₃	Δ Data ₂₀₀₃₋₂₀₁₂	Δ Model ₂₀₀₃₋₂₀₁₂
Informality Rate	0.2927	0.2515	-0.1053	-0.0514
Skilled	0.1492	0.1445	-0.0374	-0.0183
Unskilled	0.5089	0.4188	-0.1526	-0.0425
Unemployment (implied)	0.1344	0.1623	-0.0750	-0.0527
Skilled	0.0311	0.0332	-0.0160	0.0023
Unskilled	0.2514	0.3077	-0.1049	-0.0622
Wages (\ln)	0.7860	0.9055	0.3382	0.3435
Skilled	1.3454	1.4699	0.0773	0.0773
Unskilled	-0.0268	0.0286	0.5151	0.5149
Formal wage premium	0.0617	0.0311	-0.0214	-0.0495
Skilled	0.0429	-0.0537	-0.0215	0.0000
Unskilled	0.0892	0.1564	-0.0047	-0.0960
Firm-size premium 6-10 workers	0.0494	-0.0033	-0.0121	-0.0039
Skilled	0.0592	0.0016	-0.0173	-0.0091
Unskilled	0.0330	-0.0108	-0.0074	0.0039
Firm-size premium 11 or more workers	0.0732	0.0965	-0.0108	-0.0061
Skilled	0.0837	0.0958	-0.0128	-0.0024
Unskilled	0.0611	0.0965	-0.0190	-0.0141
% formal workers firms 6-10 employees	0.0775	0.0333	-0.0271	-0.0046
Skilled	0.0579	0.0232	-0.0154	-0.0023
Unskilled	0.1280	0.0558	-0.0490	-0.0027
% formal workers workers firms ≥ 11 employees	0.8552	0.8734	0.0492	0.0182
Skilled	0.8968	0.9212	0.0240	0.0072
Unskilled	0.7431	0.7654	0.1057	0.0090
% informal workers firms 6-10 employees	0.1234	0.1501	-0.0116	-0.0183
Skilled	0.1145	0.1189	-0.0051	-0.0033
Unskilled	0.1267	0.1657	-0.0128	-0.0220
% informal workers workers firms ≥ 11 employees	0.5133	0.4863	0.0345	-0.0180
Skilled	0.6797	0.6049	-0.0203	-0.0401
Unskilled	0.4383	0.4226	0.0299	-0.0311

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

Table A.10 – Individual Contribution of Each Factor with $\gamma = 0.2$, Changes in the Brazilian Labor Market from 2003 to 2012

	(1)	(2)	(3)	(4)	(5)	(6)
	All changes	All changes, except:				
Outcomes:		Fraction skilled	Minimum wage	Enforcement	Productivity/Demand A	B_1
Informality	-0.0514	0.0323	-0.0725	-0.0391	-0.0241	-0.0415
Unemployment	-0.0527	0.0342	-0.0874	-0.0548	-0.0203	-0.0419
Wages (ln)	0.3435	0.2709	0.3000	0.3432	0.1325	0.3498
Skilled	0.0773	0.2250	0.0732	0.0770	-0.1774	0.0947
Unskilled	0.5149	0.2553	0.4673	0.5197	0.3352	0.4651
Formal Wage Premium	-0.0495	0.1601	-0.1172	-0.0493	0.0521	-0.0242
Firm Size Premium (6 – 10)	-0.0039	-0.0085	-0.0142	-0.0013	-0.0024	-0.0010
Firm Size Premium (≥ 11)	-0.0061	0.0262	0.0028	-0.0070	0.0270	-0.0014

Note: Column 1 shows 2003-2012 changes predicted by our validation exercise with low γ (i.e., the same values from Table A.9, Column 4). 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 its 2003 level).

Table A.11 – Model Fit and Validation with $\gamma = 0.565$, Country Averages, 2003 and 2012

Outcomes	Fit in 2003		Validation: $\Delta 2003-2012$	
	Data ₂₀₀₃	Model ₂₀₀₃	Δ Data ₂₀₀₃₋₂₀₁₂	Δ Model ₂₀₀₃₋₂₀₁₂
Informality Rate	0.2927	0.2901	-0.1053	-0.0560
Skilled	0.1492	0.1477	-0.0374	-0.0208
Unskilled	0.5089	0.5051	-0.1526	-0.0301
Unemployment (implied)	0.1344	0.1360	-0.0750	-0.0506
Skilled	0.0311	0.0311	-0.0160	0.0025
Unskilled	0.2514	0.2546	-0.1049	-0.0725
Wages (ln)	0.7860	0.8055	0.3382	0.3387
Skilled	1.3454	1.3424	0.0773	0.0774
Unskilled	-0.0268	0.0318	0.5151	0.5151
Formal wage premium	0.0617	-0.0248	-0.0214	-0.0250
Skilled	0.0429	-0.0547	-0.0215	-0.0014
Unskilled	0.0892	0.0296	-0.0047	-0.0600
Firm-size premium 6-10 workers	0.0494	0.0090	-0.0121	-0.0071
Skilled	0.0592	0.0134	-0.0173	-0.0107
Unskilled	0.0330	-0.0006	-0.0074	-0.0004
Firm-size premium 11 or more workers	0.0732	0.0838	-0.0108	0.0053
Skilled	0.0837	0.1149	-0.0128	-0.0061
Unskilled	0.0611	0.0396	-0.0190	0.0033
% formal workers firms 6-10 employees	0.0775	0.0409	-0.0271	-0.0061
Skilled	0.0579	0.0250	-0.0154	-0.0031
Unskilled	0.1280	0.0776	-0.0490	-0.0016
% formal workers workers firms ≥ 11 employees	0.8552	0.8489	0.0492	0.0238
Skilled	0.8968	0.9202	0.0240	0.0096
Unskilled	0.7431	0.6750	0.1057	0.0062
% informal workers firms 6-10 employees	0.1234	0.1726	-0.0116	-0.0232
Skilled	0.1145	0.1196	-0.0051	-0.0016
Unskilled	0.1267	0.1948	-0.0128	-0.0278
% informal workers workers firms ≥ 11 employees	0.5133	0.4995	0.0345	-0.0225
Skilled	0.6797	0.6658	-0.0203	-0.0426
Unskilled	0.4383	0.4279	0.0299	-0.0375

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

Table A.12 – Individual Contribution of Each Factor with $\gamma = 0.565$, Changes in the Brazilian Labor Market from 2003 to 2012

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All changes, except:				
Outcomes:	changes	Fraction skilled	Minimum wage	Enforcement	Productivity/Demand A	B_1
Informality	-0.0560	0.0215	-0.0803	-0.0425	-0.0133	-0.0067
Unemployment	-0.0506	0.0019	-0.0803	-0.0530	-0.0150	-0.0234
Wages (ln)	0.3387	0.2764	0.3086	0.3388	0.1014	0.3370
Skilled	0.0774	0.2135	0.0775	0.0771	-0.1952	0.1301
Unskilled	0.5151	0.3652	0.4822	0.5208	0.2828	0.3212
Formal Wage Premium	-0.0250	0.0768	-0.0743	-0.0247	0.0867	0.0636
Firm Size Premium (6 – 10)	-0.0071	-0.0072	-0.0172	-0.0039	-0.0041	-0.0010
Firm Size Premium (≥ 11)	0.0053	0.0130	0.0132	0.0053	0.0379	0.0202

Note: Column 1 shows 2003-2012 changes predicted by our validation exercise with high γ (i.e., the same values from Table A.11, Column 4). 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 group (which is held constant at its 2003 level).

I.3 Alternative Counterfactuals Holding Skill Composition Constant

Table A.13 – Alternative Counterfactuals Holding Skill Composition Constant, Changes in the Brazilian Labor Market between 2003 and 2012

Changes in: Outcomes	(1)	(2)	(3) (4)	
	Data	Model	Alternative Counterfactuals	
			P_r	T_r
Informality	-0.1053	-0.0603	-0.0604	-0.0603
Unemployment	-0.0750	-0.0434	-0.0136	-0.0036
Wages (ln)	0.3382	0.3437	0.2485	0.2514
Skilled	0.0773	0.0774	0.0770	0.0773
Unskilled	0.5151	0.5152	0.5156	0.5151
Formal Wage Premium	-0.0214	-0.0353	-0.0230	0.0109
Firm Size Premium (6–10)	-0.0121	-0.0065	-0.0019	-0.0072
Firm Size Premium (≥ 11)	-0.0108	-0.0017	0.0057	0.0251

Note: Columns 1 and 2 repeat values from our main validation exercise in Table 4. Column 3 shows the alternative counterfactual where workforce composition is held constant at its 2003 level, but the enforcement parameter P_r changes from 2003 to 2012 so that the predicted fall in informality by region is the same as in the baseline model. Column 4 shows a similar counterfactual where the parameter being changed is T_r instead of P_r .

I.4 Cross-sectional Features of the Equilibrium, Metropolitan Regions

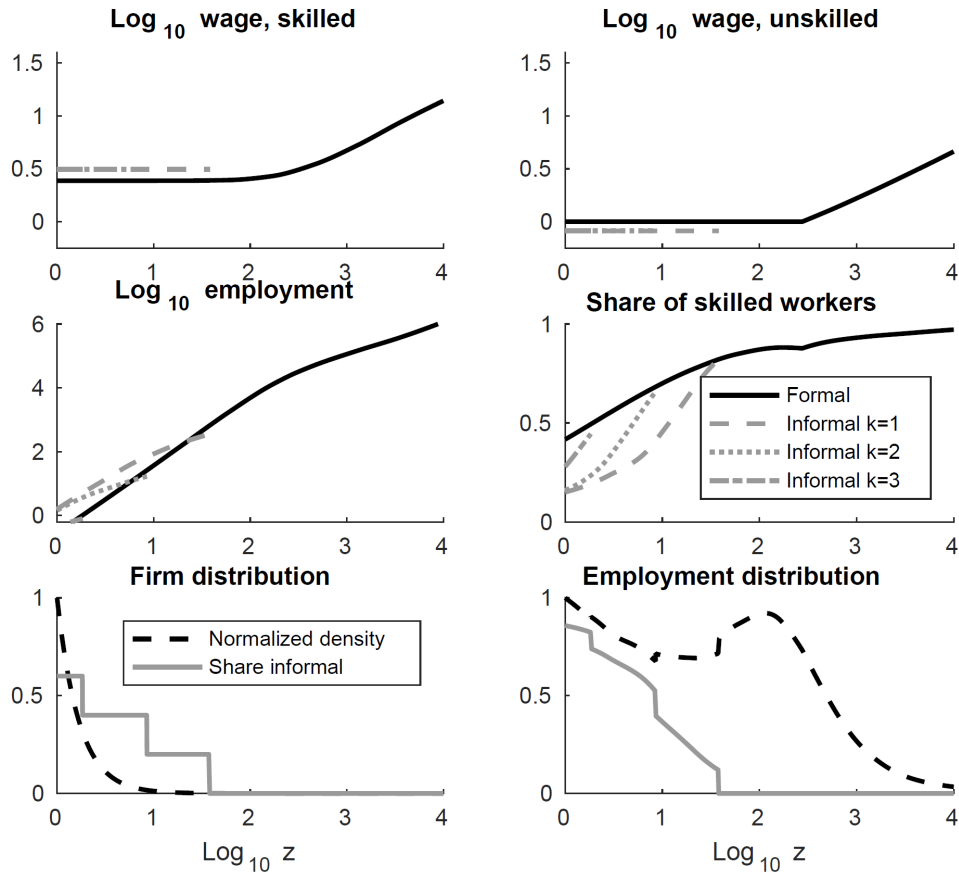


Figure A.3 – Cross-sectional Features of the Equilibrium, Salvador Metropolitan Region, 2003

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 or workers that are informal conditional on productivity. The latter curve has discontinuities because the distribution of informality cost shifters k is discrete.



Figure A.4 – Cross-sectional Features of the Equilibrium, Belo Horizonte 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 $k \geq 3$ are formal for all values of z , there are only two such curves. The bottom two panels show distributions of firms and total employment by firm productivity z . They also show shares of firms or workers that are informal conditional on productivity. The latter curve has discontinuities because the distribution of informality cost shifters k is discrete.

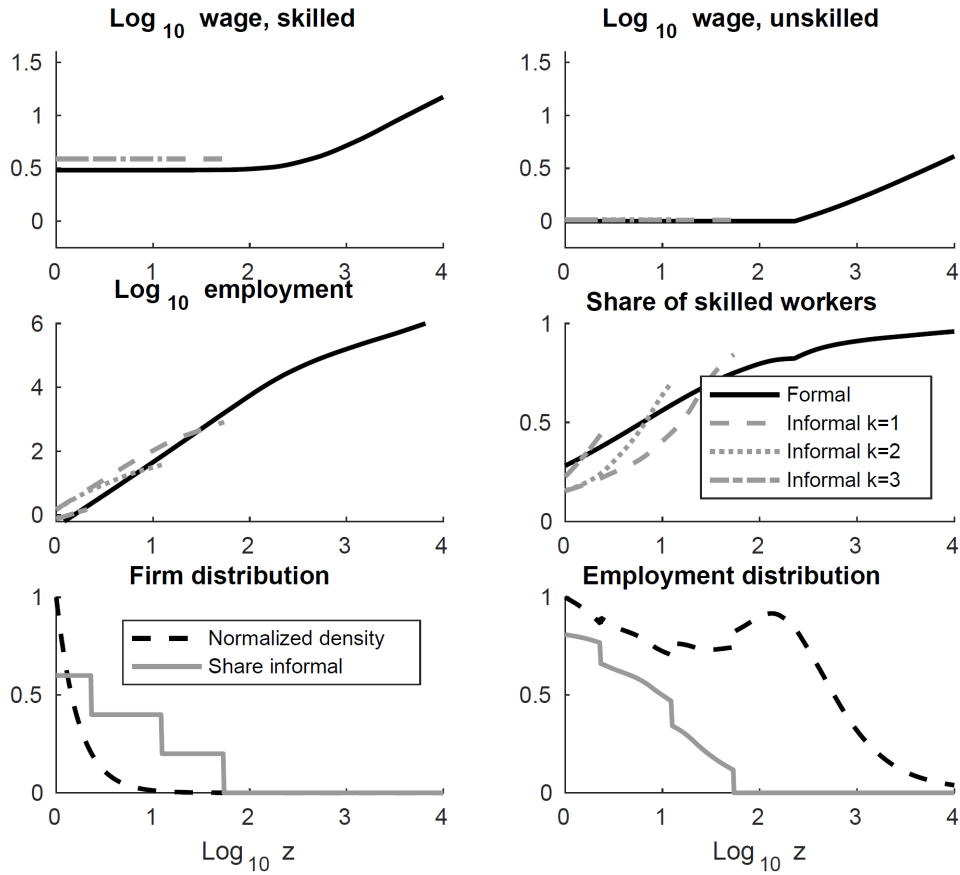


Figure A.5 – Cross-sectional Features of the Equilibrium, Rio de Janeiro Metropolitan Region, 2003

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 or workers that are informal conditional on productivity. The latter curve has discontinuities because the distribution of informality cost shifters k is discrete.

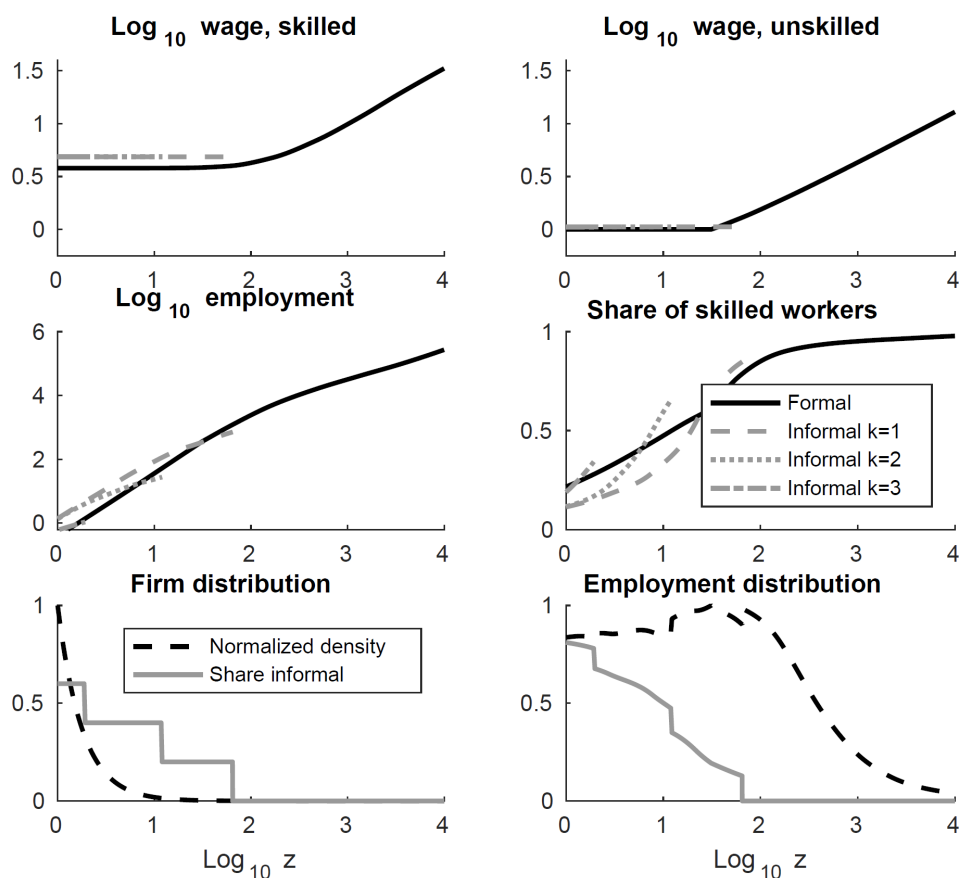


Figure A.6 – Cross-sectional Features of the Equilibrium, São Paulo Metropolitan Region, 2003

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 or workers that are informal conditional on productivity. The latter curve has discontinuities because the distribution of informality cost shifters k is discrete.

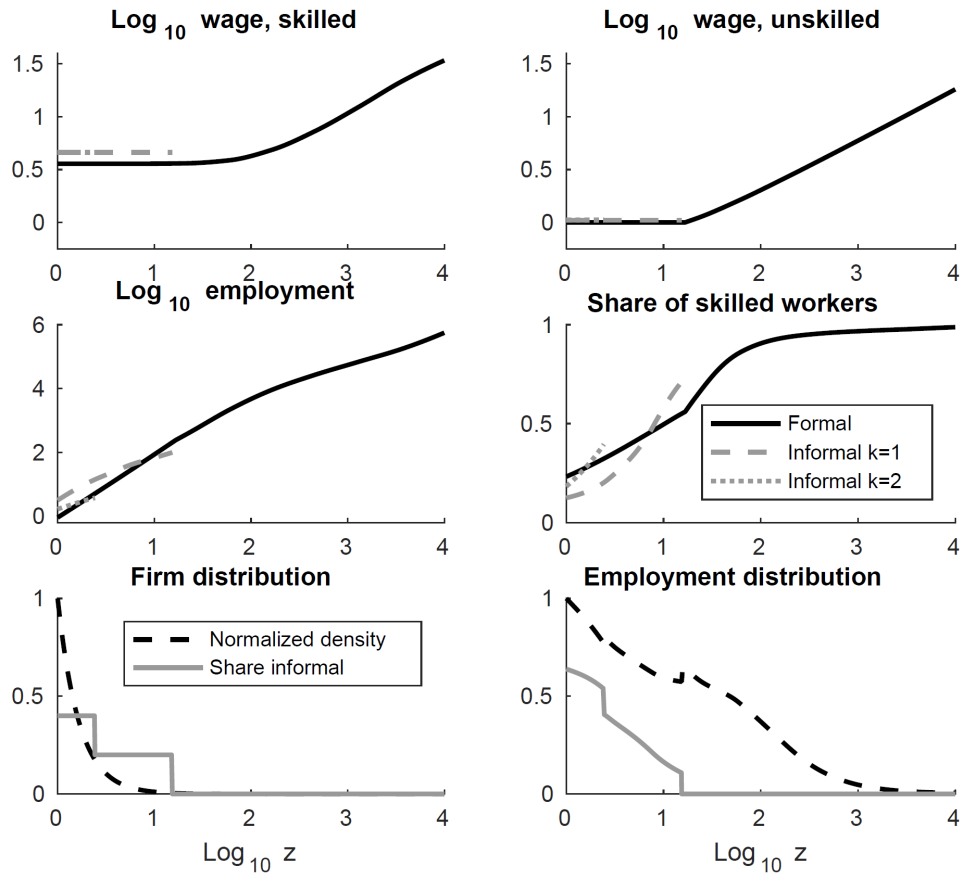


Figure A.7 – Cross-sectional Features of the Equilibrium, Porto Alegre Metropolitan Region, 2003

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 $k \geq 3$ are formal for all values of z , there are only two such curves. The bottom two panels show distributions of firms and total employment by firm productivity z . They also show shares of firms or workers that are informal conditional on productivity. The latter curve has discontinuities because the distribution of informality cost shifters k is discrete.

I.5 Extra Figures

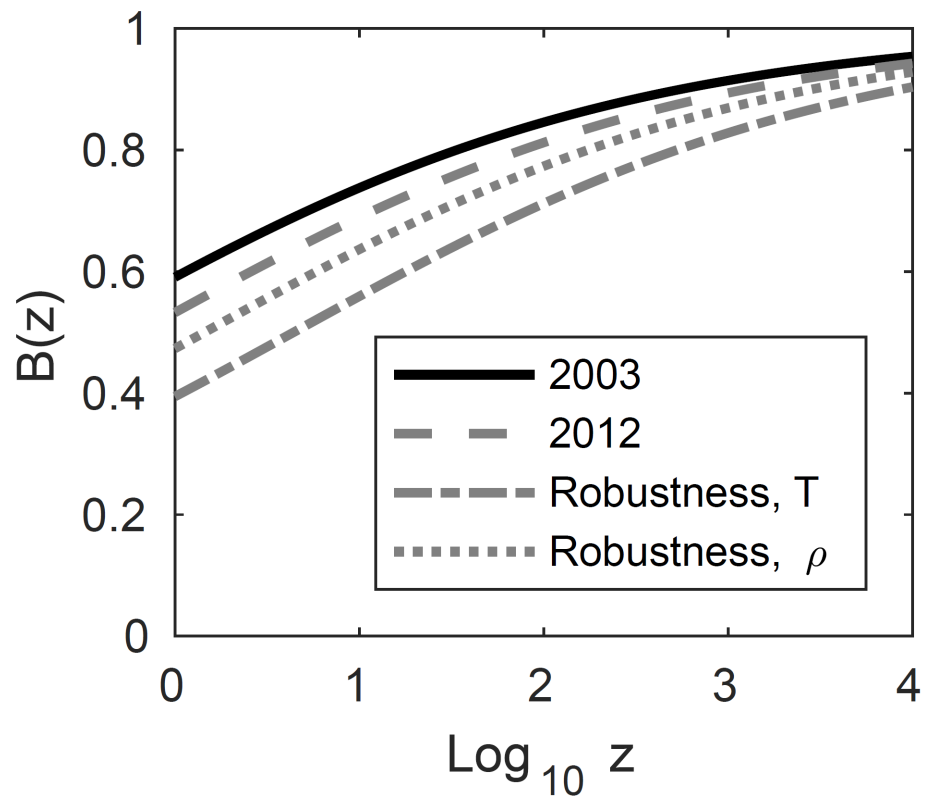


Figure A.8 – Shape of $B(z)$ Function, Recife Metropolitan Region

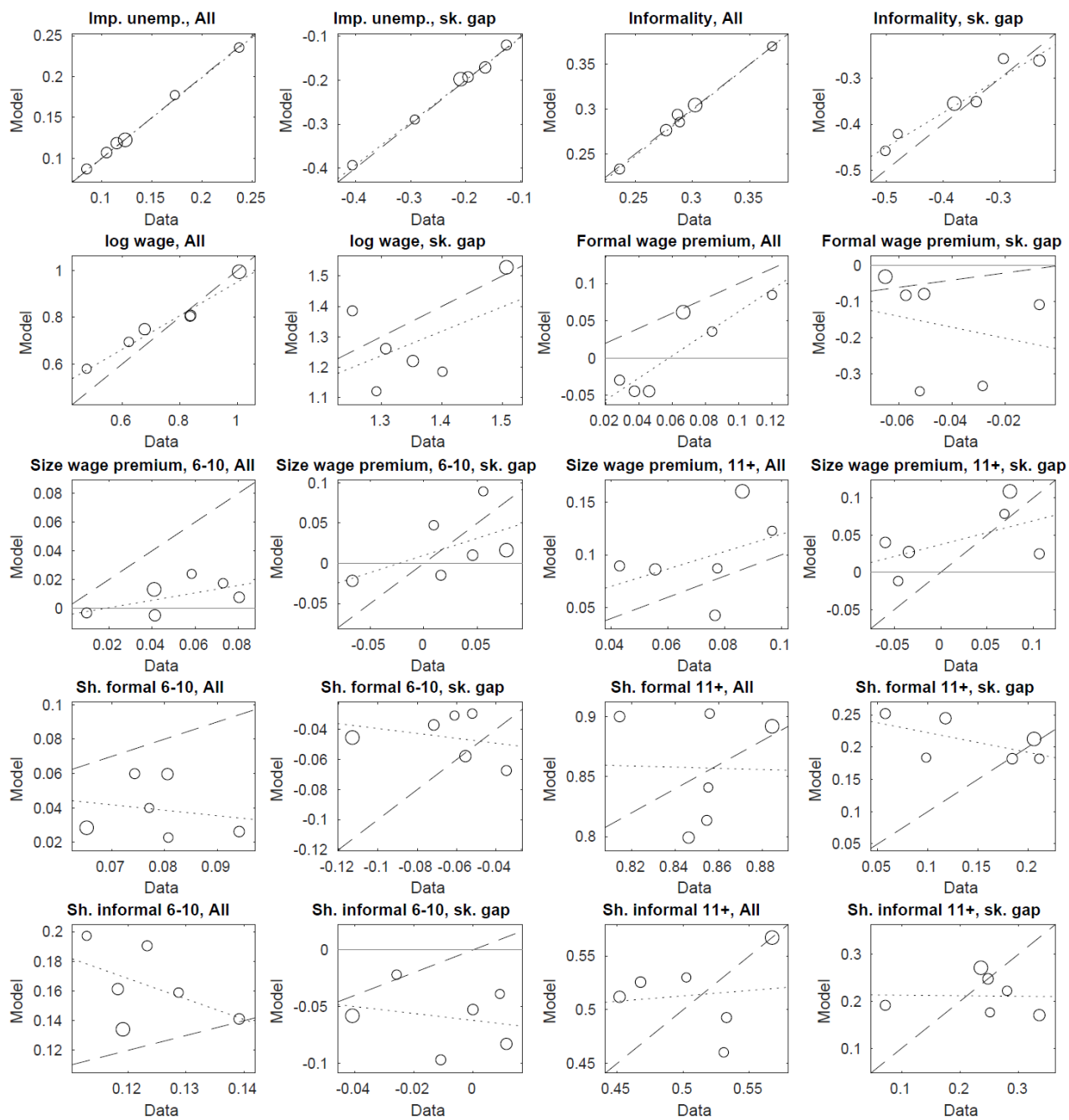


Figure A.9 – Model Fit across Regions, 2003

Note: In each panel, the scatterplots show the relationship between levels in the data and those predicted by the model in each region, with circle sizes corresponding to workforce size. The dashed line is the 45 degree curve (representing perfect fit), the dotted line is a regression line fitted to the scatterplot, and the solid line is the horizontal axis.

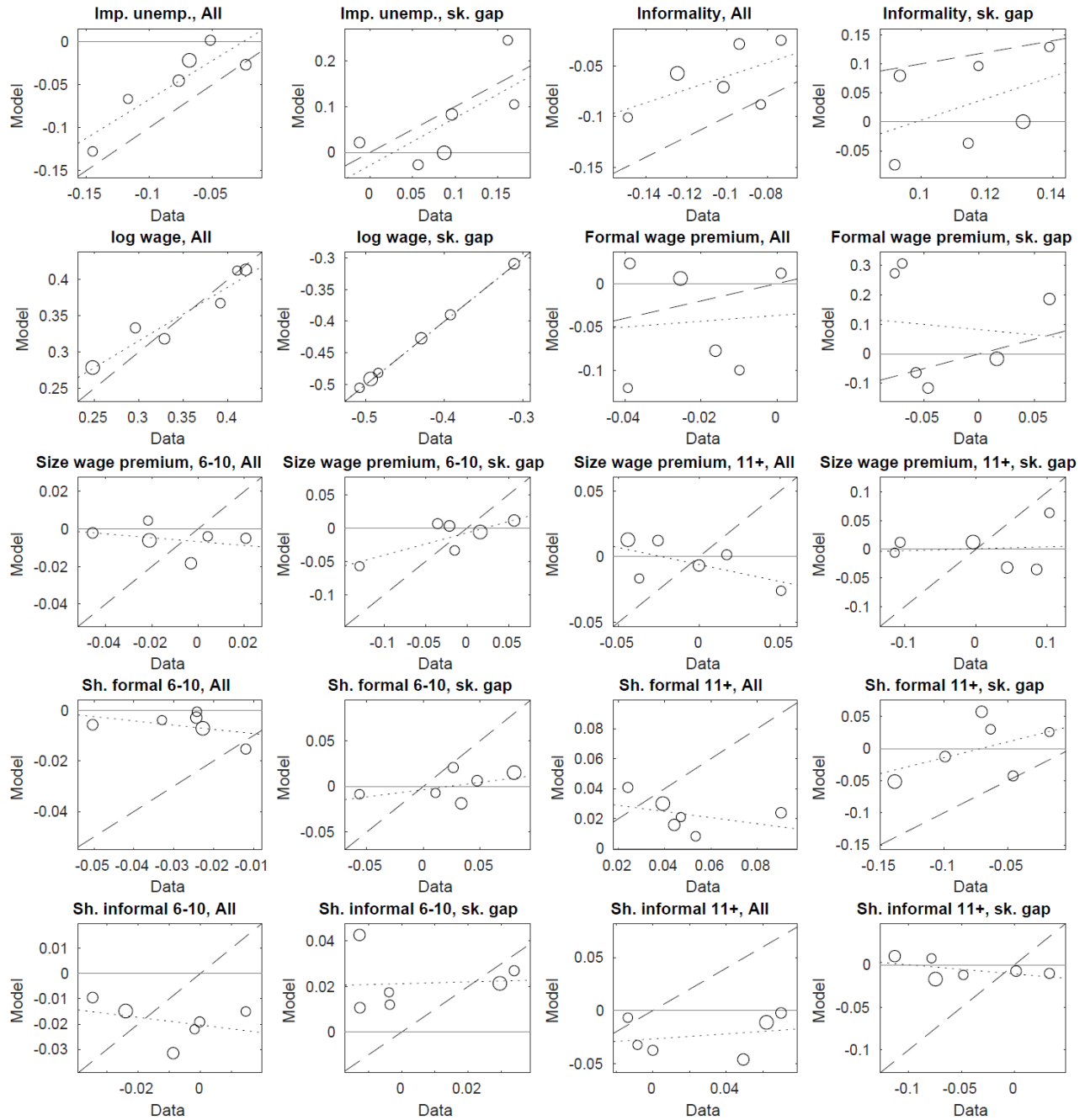


Figure A.10 – Model Validation across Regions, 2003-2012

Note: In each panel, the scatterplots show the relationship between changes in the data and those predicted by the model in each region, with circle sizes corresponding to workforce size. The dashed line is the 45 degree curve (representing perfect fit), the dotted line is a regression line fitted to to scatterplot, and the solid line is the horizontal axis.