DISCUSSION PAPER SERIES

IZA DP No. 9951

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

Daniel Haanwinckel
Rodrigo R. Soares

May 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Daniel Haanwinckel
University of California, Berkeley

Rodrigo R. Soares
Sao Paulo School of Economics-FGV and IZA

Discussion Paper No. 9951
May 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9951
May 2016

ABSTRACT

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

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

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

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

Corresponding author:

Rodrigo R. Soares
Sao Paulo School of Economics - FGV
Rua Itapeva 474, 12o Andar
01332-000 São Paulo, SP
Brazil
E-mail: rodrigo.reis.soares@fgv.br

* 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 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 LACEA Annual Meeting. 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.

1 Introduction

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

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

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

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

In order to accommodate decreasing returns to scale and imperfect substitutability between different types of labor within a search model, we draw from the intra-firm bargaining theory proposed by Cahuc, Marque and

Wasmer (2008), who build on Stole and Zwiebel (1996a), and extend it in three directions. First, we characterize an equilibrium where labor can move between the formal and informal sectors. Second, we consider firms with different productivity levels, as opposed to a single representative firm. And third, we incorporate a more realistic set of labor regulations, including minimum wages, which adds a non-trivial degree of complexity to the characterization of the solution.

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

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

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

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

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

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

In addition to the theoretical points and the quantitative exercises mentioned before, the paper makes two conceptual contributions to the informality literature. First, it shows that both the cross-sectional and time-series variations in informality are consistent with a model in which informality is entirely due to the existence of labor market regulations. The model reproduces several stylized facts related to informality and its recent evolution resorting only to regulatory distortions and to search and matching frictions commonly associated with the functioning of the labor market. Second, it rationalizes three interrelated and widely documented patterns that are incompatible with previous informality models: the presence of skilled and unskilled workers in both the formal and informal sectors, the rising share of skilled workers by firm size (and formality status), and the declining formal wage premium by skill level (becoming null or negative at the top). Many authors suggest that the heterogeneity in the formality wage premium indicates that the informal sector is composed of two distinct tiers. For the more productive workers at the top tier, informality is a matter of opportunity, which is reflected on their wages being equal to or higher than they would be in the formal sector. For the bottom tier, informality is strictly worse than formal employment, since informal workers earn lower wages and lack valuable mandated benefits. In our model, the two tiers are clearly identified by the two skill levels, and the pattern of decreasing wage gap results from the binding minimum wage for unskilled workers.[1]

[1] Bargain and Kwenda (2011) find this pattern in fixed-effects models using data from Brazil, Mexico and South Africa. Botelho

Our model builds upon many search models from the informality literature, but differ from them in key aspects. Boeri and Garibaldi (2007) and Boeri, Garibaldi and Ribeiro (2011) propose simple models with worker heterogeneity, but without the possibility of substitutability between different types of labor and with poor institutional characterizations. In both papers, the equilibrium displays complete segregation of workers by skill level across the formal and informal sectors. Albrecht, Navarro and Vroman (2009) introduces uncertainty about workers' productivity in the formal sector and a richer institutional setting, but maintains the one-to-one matching between workers and firms, in addition to assuming strong structural differences between sectors and no compliance decision on the side of the firms. Ulyssea (2010), Bosch and Esteban-Pretel (2012), and Meghir, Narita and Robin (2015) have more sophisticated compliance decisions and are better equipped in institutional details, but forgo worker heterogeneity. Ulyssea (2010) still assumes substantial structural differences between sectors, while Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015) assume that formal and informal firms differ only in their choice to abide by labor regulations.[2] On the institutional side, Ulyssea (2010) incorporates unemployment insurance and severance payments, and Meghir, Narita and Robin (2015) accounts for both these dimensions and minimum wages.[3]

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

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

and Ponczek (2011) reach similar conclusion with Brazilian data under different specifications (also using panel data), and observe that the formal wage premium decreases as workers become older and more educated. Lehman and Pignatti (2007) find similar results for the Ukrainian labor market. The idea of a two-tiered informal sector goes back at least to Fields (1990). Günther and Launov (2012) develop an econometric model of selection to test the hypothesis of heterogeneity inside the informal sector. They find that there are two distinct groups in the informal sector in Côte d'Ivoire. Some of these authors, as well as others, have used the term "segmentation" to describe the bottom tier of the informal sector. By that, they mean that wages are not fully determined by individual productivity and compensating differentials. This interpretation, present in Fields (1975) and Rauch (1991), is different from the original concept of segmented labor markets, as described in Dickens and Lang (1985) or Cain (1976). In the case we discuss, increases in education (or, more generally, productivity) can lead every worker to better jobs, a view that contrasts with labor market duality. In addition, the significant flow of workers in and out of the informal sector, particularly among those with lower skills, undermines the hypothesis of strong non-economic barriers of entry to the so-called primary sector. To our knowledge, Araujo, Ponczek and Souza (2016) present the only alternative model that explains the decreasing wage gap among salaried workers, but in a very specific setting (one-to-one random matching model with asymmetric information, where workers can take employers to court). Bargain et al. (2012) account for heterogeneity in income gaps between formal and informal self-employed workers.

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

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

labor market outcomes in Brazil between 2003 and 2012 and conducts some policy experiments. Section 6 concludes the paper.

2 Empirical Context

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

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

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

The average informal worker in Brazil earns a lower wage, is less educated, and works in a smaller firm than her formal counterpart. The first claim is evident from the top row in Table 1. While the average formal hourly wage was 4.83 Brazilian Reais in 2003 (around 1.60 US dollars), the average informal wage was 32% lower (2.67 Brazilian Reais). Table 2 also presents the distribution of workers across sectors, firm sizes, and educational categories. By comparing the totals along rows for each sector, the differences in average schooling become clear: 40% of informal employees had less than 8 years of schooling, while the analogous number was less than 28% in the formal sector. The differences in firm size can be seen in the column totals. While only a minority (roughly 1/16) of formal employees worked in firms with 5 workers or less, this fraction was over one third for informal employees.

These stylized facts are consistent with many papers that discuss the empirical regularities of informality in the developing world, such as La Porta and Shleifer (2008) and Maloney (2004). They have been traditionally interpreted as evidence that informality is circumscribed to low-earning, unskilled workers, but a closer look at the data reveals that this assertion is not accurate. Table 1 shows that the informality rate among workers with a college degree is 17.3%, not dramatically lower than the overall rate of 28.4%. Moreover, informal workers with college earn almost three times as much as the average formal employee. Note that these individuals are not self-employed professionals defaulting on taxes or social security contributions, since we have restricted our sample to wage earners. The table also suggests that there is no labor market segmentation in the traditional sense: as workers become more educated, they are more likely to be employed formally and also more likely to receive higher wages if they stay in the informal sector. Finally, the fact that some informal firms are willing to pay high wages for skilled workers shows that the technology used by these firms displays significant returns to human capital, contradicting many depictions of labor market duality in which informal firms are presented as being structurally different from

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

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

Source: IBGE/PME, author's calculations.

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

Table 2 – Educational Distribution of Workers by Sector and Firm Size, Brazil, 2003

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

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

formal ones.

But it is also useful to highlight that formal schooling does not seem to encompass all dimensions of skill relevant to the labor market. To illustrate this point, Table 3 shows the distribution of wages in the formal sector by educational level. There is a wide dispersion in wages across all levels of schooling, with the exception of college or more. For example, among those with complete high school and college drop outs, there is a fraction of 8.4% earning roughly one minimum wage, while 15.5% earn more than 5 times the minimum wage. Wage dispersion seems almost as large within as across educational categories, despite the fact that average wages – and, therefore, skills – do increase with years of schooling.

We can look at data on firm size in Table 2 to infer the hiring behavior of firms in both sectors. Comparisons between different columns in the same sector show that, as firm size increases, the proportion of educated workers also increases. In other words, larger firms are more likely to have a higher fraction of educated workers. An important takeaway is that this pattern is observed for workers in both sectors, suggesting again that the technologies used

Table 3 – Formal Wage Distribution by Schooling Levels and Workforce Composition, Brazil, 2003 or 2012 (when indicated)

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

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

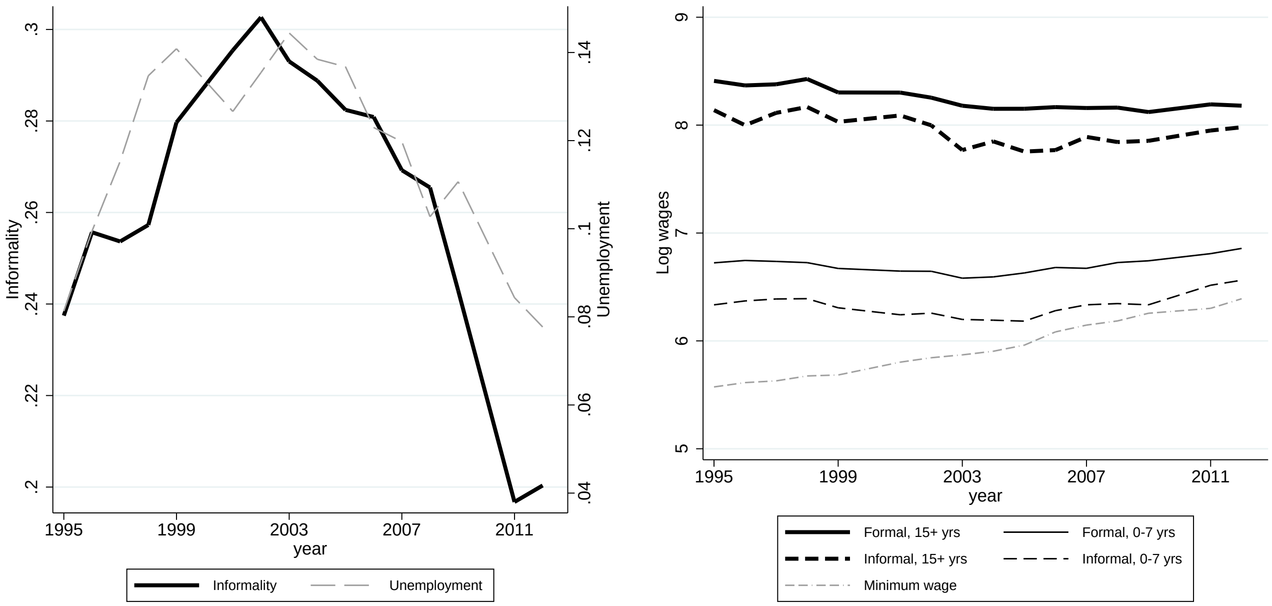

Figure 1 – Evolution of Informality, Unemployment and Real Wages for Salaried Workers, Brazil, 1995-2012
Source: IBGE/PNAD, author's calculations. The sample is restricted to the six metropolitan regions surveyed in the IBGE/PME.

by formal and informal firms, at the margin, are not substantially different.

Now we turn to the evolution of informality in Brazil since the 1990s. Figure 1 shows that the rate of informality was rising up to 2002, but then started declining sharply.[4] In Appendix B, we show that the decline was widespread in the economy and not driven by workforce reallocation (i.e., a movement of employment to sectors of economic activity that are intrinsically more formal). What makes this pattern intriguing is the observation that, while the upward trend has been credited to increasing costs of formal employment during the 1990s, these costs continued to rise even after the reversal.[5] In particular, the minimum wage increased dramatically throughout the period, accumulating real gains of 60% from 1995 to the end of 2003, and another 61% from 2003 to 2012.

There is some evidence that the enforcement of labor regulation in Brazil has become more efficient, a factor that could also bring down both unemployment and informality rates.[6] But enforcement cannot account for other important shifts in labor market outcomes: Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015), for example, predict that the formal wage premium should increase as a consequence of more enforcement, which is the opposite of what happens in the data.

The changing composition of the workforce, evident in the last columns in Table 3, may have contributed to the patterns described here, despite rarely appearing in the literature as an important determinant of informality.

[4]In Figure 1, we use data from the National Household Survey (PNAD) instead of the PME, because of methodological changes in PME in 2002.

[5]Barros and Corseuil (2001) explain how the 1988 Constitution significantly raised employment costs (payroll and firing costs and mandated benefits). Bosch, Goñi-Pacchioni and Maloney (2012) claim that these changes were the most important factor behind the increase in informality during the 1990s. We present a brief discussion of changes in labor legislation and tax rates after 2003 in Appendix A.

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

Two intuitive arguments hint at this potentially important role. First, since informality is much lower among the highly educated, increases in the share of skilled workers should mechanically lead to a decline in informality due to a compositional effect (abstracting from equilibrium considerations).[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. When coupled with the increases in TFP observed in Brazil during this period (documented, for example, by Ferreira and Veloso, 2013), changes in the relative supply of skills seem promising as a main driving force behind the evolution of labor market outcomes.

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

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. 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.

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

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

[7] In fact, Mello and Santos (2009) and Barbosa Filho and Moura (2012) 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.

3.1 Labor Markets

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

We make no distinction between formal and informal firms in the search process. The aggregate $V_i = V_i^{for} + V_i^{inf}$ is the sum of all vacancies posted by formal and informal firms, and unemployed workers search simultaneously in both sectors. After a worker is matched to a vacancy, the probability that this vacancy is offered by a formal firm is given by $\phi_i = \frac{V_i^{for}}{V_i}$, which is simply the fraction of vacancies posted by formal firms in market i. With this assumption, as with many others, we try to minimize the structural differences between formal and informal sectors and focus instead on the regulatory asymmetries. Our modeling of the search process is most similar to that in Bosch and Esteban-Pretel (2012). Other models with undirected search, such as Ulyssea (2010) and Meghir, Narita and Robin (2015), assume exogenous differences in the matching technology across sectors.

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 continuous and twice differentiable, where n_s and n_u denote units of skilled and unskilled labor. The term z is an exogenous productivity parameter distributed across firms according to a distribution function $G(z)$. We assume that $F^z(\cdot)$ is strictly concave in (n_s, n_u) for any z in the support of $G(z)$, and increasing in z. Moreover, we assume that $\sigma_{z,n_s} < \sigma_{z,n_u}$, where $\sigma_{i,j}$ denotes the partial elasticity of substitution between inputs i and j. Given fixed hiring costs, firms with higher z employ relatively more skilled workers. The parameter z is most easily interpreted as entrepreneurial talent, as in Lucas (1978), with the idea that entrepreneurs cannot efficiently manage a large number of skilled workers if they are not highly talented themselves.

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

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

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

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

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

Job relations are destroyed at exogenous separation rates s^{for} and s^{inf}, which depend on the compliance decision. This allows the model to capture the empirical pattern of higher labor turnover among informal firms.[8] The dynamics of labor quantities inside each firm are given by

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

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

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

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

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

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

Denote by $J_i^{z,j}(n_s, n_u)$ the marginal value of an additional worker of type i in a firm of type z, with compliance status j: $J_i^{z,j}(n_s, n_u) = \frac{\partial \Pi^{z,j}(n_s,n_u)}{\partial n_i}$. We derive the first order conditions for the firm's problem in Appendix C. From now on, we restrict attention to steady-state solutions where the numbers of workers of different types are constant in each firm. By imposing $\dot{n}_i = 0$ in the F.O.C.'s, the expressions simplify to:

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

$$J_i^{z,j}(n_s, n_u) = \frac{\xi}{q(\theta_i)}, \tag{3}$$

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

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 s^j. The instantaneous rent is given not only by the difference between marginal product and wage, but also by the effect of this additional employee on the wages of all other workers currently employed by the firm, due to changes in marginal productivities. At the time of the hiring decision or bargaining, previous vacancy costs are sunk and thus do not appear in this expression.

Equation 3 is the optimality condition in a steady state. Its interpretation is straightforward: the value of the marginal worker must be equal to the expected cost of hiring another worker, which is the cost ξ per vacancy multiplied by the expected number of vacancies needed to hire a worker. By combining both expressions, we find an equation similar to the standard first order condition of the firm in which marginal product equals marginal cost:

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

The case for informal firms is analogous, just omitting the payroll tax τ and adding the marginal effect of n_i on the informality penalty $\rho(n_s + n_u)$.

3.3 Wage Determination

Wage is determined through Nash bargaining, with workers and firms sharing the rents created by the match. The share of the surplus appropriated by a worker is given by the exogenous parameter σ, which corresponds to the bargaining power of workers. Differently from the standard model in Pissarides (2000), we do not assume homogeneous labor nor constant returns to scale in the production function, and allow workers and firms to engage in renegotiation after the initial match. As discussed in Stole and Zwiebel (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. We follow the solution developed by Cahuc, Marque and Wasmer (2008), who analyze this type of problem in a context with search frictions.

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

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

We can write the flow equations that define the value of employment as

$$rE_i^{for}(w) = a_i w + b_i + s^{for}\left[U_i - E_i^{for}(w)\right], \text{ and} \tag{4}$$

$$rE_i^{inf}(w) = w + s^{inf}\left[U_i - E_i^{inf}(w)\right], \tag{5}$$

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

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

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

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

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

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

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

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

Now we explain how the introduction of a minimum wage changes these results. If the bargained wage in a formal firm for one type of worker – typically, the unskilled – is lower than the minimum wage, then the minimum wage restriction is binding. The Nash bargaining equation is not satisfied anymore for unskilled workers; indeed, in this situation, these workers receive a share of rents larger than σ. This also implies that the previous wage function for skilled workers is not valid anymore, since the term $\frac{\partial w_u^{z,for}}{\partial n_s}$ in equation 2 is equal to zero (marginal changes in the number of skilled workers do not affect wages of unskilled workers, which are binding at the minimum wage). In Appendix D, we show that the wage equation for skilled workers in the formal sector when the minimum wage binds for unskilled workers is

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

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

In Appendix D, we show that, because of this discontinuity, there might not be a solution to the first order conditions when the unconstrained (freely bargained) unskilled wage is slightly lower than the minimum wage. In these cases, firms engage in the strategic manipulation of firm size described above.[9] In our quantitative exercises,

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

we deal explicitly with this issue by assuming that firms in this situation choose employment figures that (i) satisfy the first order condition for skilled workers and (ii) lie immediately to the "left" (in terms of n_u) of the region of the (n_s, n_u) space where the minimum wage binds for unskilled workers. Details are laid out in Appendix D.

Now we turn to the analysis of wage determination in equilibrium. If we replace equation 3 in 6, and take into account that the bargaining equation is not satisfied if the minimum wage is binding, we have

$$\begin{aligned}(1-\sigma)\left[E_i^{for}\left(w_i^{z,for}\right)-U_i\right] &\geq \sigma\frac{\xi}{q(\theta_i)}, \quad i\in\{s,u\}, \quad \text{with} > \text{only if } w_i^{z,for}=\bar{w}, \text{ and}\\ (1-\sigma)\left[E_i^{inf}\left(w_i^{z,inf}\right)-U_i\right] &= \sigma\frac{\xi}{q(\theta_i)}, \quad i\in\{s,u\}.\end{aligned}$$

Recalling expressions 4 and 5, notice that E_i^j does not depend directly on firm size or productivity. So neither n_i nor z appear in the expressions above. In a steady-state equilibrium, wages paid for a worker of a given type, working in a firm in a given sector, are the same for all firms in that sector irrespective of firm size. In other words, in equilibrium, there are only four wages in this economy: w_s^{for}, w_u^{for}, w_s^{inf} and w_u^{inf}.

This result comes immediately from the fact that the matching technology and the cost of posting a vacancy are the same across firms of different sizes. The intuition behind it is that, regardless of productivity, all firms adjust the number of employees so as to equate the marginal value of workers to the expected search cost, which does not depend on productivity or firm size. Thus, the value added by the marginal worker in equilibrium is the same across the productivity distribution. Finally, since we assume that the worker's bargaining power is not related to firm size or productivity, the solution to the Nash bargaining cannot vary with z.[10]

3.4 Equilibrium

So far, we have described the behavior of firms taking θ_i and U_i as given. In equilibrium, these values have to be consistent with the aggregate behavior of firms and workers. The labor market tightness, as explained in subsection 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

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

where n_i^z denotes the optimal employment of type i workers for a firm with productivity z. Since, in equilibrium, $\dot{n}_i=0$ for all firms, $v_i^z = s^j n_i^z/q(\theta_i) \Longrightarrow V_i^j = s^j N_i^j/q(\theta_i)$. We can therefore find the expressions that pin down θ_i,

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

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

[10]This result greatly simplifies the solution and interpretation of the model. The reason why it differs from wage dispersion as featured in wage posting models is that the choice of vacancies, along with decreasing marginal returns, provide additional degrees of freedom to the firm, so that firms with different productivities can drive marginal productivities down to some common value associated with the outside option of workers. In contrast, wage posting models usually assume that increased wages are the only possible dimension of firm effort in the search process. Our modeling choice comes at the cost of eliminating the possibility of accommodating the widely documented firm size wage premium within the model. A simple way to account for this pattern would be to assume that the bargaining power of workers increases with z, as a result of greater worker unionization, for example. Pratap and Quintin (2006) and Badaoui, Strobl and Walsh (2010) provide a discussion of the relationship between the formality wage premium and the firm size wage premium.

$$
\begin{aligned}
rU_i &= \theta_i q(\theta_i)\left[\phi_i E_i^{for}(w_i^{for}) + (1-\phi_i)E_i^{inf}(w_i^{inf}) - U_i\right] \\
&= \begin{cases} \frac{\sigma}{1-\sigma}\xi\theta_i & \text{, if } w_i^{for} > \bar{w} \text{ and} \\ \frac{\theta_i}{1+\frac{\phi_i\theta_i q(\theta_i)}{r+s^{for}}}\left[\phi_i q(\theta_i)\frac{a_i\bar{w}+b_i}{r+s^{for}} + (1-\phi_i)\frac{\sigma}{1-\sigma}\xi\right] & \text{, otherwise.} \end{cases}
\end{aligned} \tag{8}
$$

For simplicity, since we incorporate unemployment benefits in the parameters a_i and b_i, we assume that individuals derive no utility flow from unemployment. The instantaneous return of being unemployed is the expected value of finding a job and leaving unemployment. In case a worker finds a job, which happens with probability $\theta_i q(\theta_i)$, there is a probability $\phi_i = \frac{V_i^{for}}{V_i^{for}+V_i^{inf}} = \frac{s^{for}N_i^{for}}{s^{for}N_i^{for}+s^{inf}N_i^{inf}}$ that the match is with a formal firm. The second expression is the result of inserting the first order condition of the firm, equation 3, in 8.

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

1. the wage functions solve the system of differential equations given by expressions 2 and 6;
2. the labor schedules n_s^z and n_u^z 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.

3.5 Discussion: Compensating Differentials

From the final Nash bargaining equations, we can show that:

$$
E_i^{For}\left(w_i^{for}\right) \geq E_i^{inf}\left(w_i^{inf}\right), \quad i \in \{s, u\}.
$$

This expression holds as an equality if the minimum wage is not binding for skill level i. In this case, we can use the definition of $E_i^j(w_i^j)$ to show that

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

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

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

4 Fitting the Model

We fit the model to the Brazilian labor market in 2003, calibrating some of the parameters and estimating others using a minimum distance procedure. We choose 2003 as the baseline year because it is close to the reversal of the informality trend (Figure 1) and it is when the second wave of the Informal Urban Economy survey (*Economia Informal Urbana*, ECINF) was conducted by the Brazilian Census Bureau (IBGE). The ECINF targeted small urban firms, most of which were unregistered, thus providing an estimate of the number of informal firms in the economy. We use the survey's micro data in the next section, but, since the ECINF is relatively small and was not repeated after 2003, it is not our main source.

Most of the data we use come from the Monthly Employment Survey (*Pesquisa Mensal do Emprego*, PME), also conducted by IBGE. The PME is a household survey that provides information on employment, wages, occupational choice, formality status, and other characteristics of the workforce, including educational attainment. Because there was an increase in the minimum wage on April 1st, 2003, we restrict the sample to the months of April through December of that year.[11] We use two other data sources from IBGE: the Central Registry of Firms (*Cadastro Central de Empresas*, CEMPRE), a registry of formal firms, and the annual projections of the size of the workforce.

4.1 Functional Forms

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

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

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

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

raise the productivity of skilled labor. If $\gamma \in (0,1)$, unskilled workers are more productive in a firm with a higher z and with more skilled workers.[12]

The parameter z is assumed to follow a Generalized Pareto distribution, to account for the fact that the majority of firms are small but a large part of the workforce is employed by large firms (see IBGE, 2005). We set the location parameter to zero, so that the smallest firms have z arbitrarily close to zero. Also, we normalize the scale parameter to $1-T$, where T is the shape (tail) parameter, so that average productivity is normalized to one.[13] Increases in T are thus mean-preserving spreads that add probability mass to extreme values of productivity. The cumulative distribution of productivity is given by[14]

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

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

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

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

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

4.2 Calibrated Parameters

Table 4 presents a first subset of the parameter values we use.

A non-trivial problem in our calibration exercise is how to map observed traits at the individual level to skills in the model. In the model, skills map directly into wages. In the relevant case from the perspective of the quantitative analysis, formal sector minimum wages bind only for unskilled workers. This gives an empirical counterpart of skills for formal workers that does not match perfectly with schooling. Unskilled workers in the model represent workers in the data who receive close to the minimum wage when employed in the formal sector. If they receive significantly more than the minimum wage in a formal job, then they must correspond to skilled workers in the model. As mentioned in section 2, there is a wide dispersion of wages for each level of schooling in the data, indicating that the definition of skill in the model does not map easily into schooling (despite being highly correlated with it).

Our approach is to combine an aggregate definition of the share of skilled workers with the individual level

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

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

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

Table 4 – Parameters Imputed from the Data or from the Literature

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

implications of the model in terms of the relationship between wages and skills. We assume that the measure η of skilled workers corresponds to the fraction of the workforce with 8 or more years of schooling, but let the quantitative model determine the allocation of workers of different skill levels to the formal and informal sectors based on the distribution of wages observed in the data. Though inevitably somewhat arbitrary, our choice of 8 or more years of schooling to represent skilled workers is based on the distributions of schooling and wages in the Brazilian labor market, discussed in section 2, and on the definition of skills that arise from the model (earning more than the minimum wage in the formal sector).[15]

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

The job destruction rates s^j are taken from estimates of the duration of employment spells in Gonzaga (2003). 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 same values for the parameters of the matching function as Ulyssea (2010). Finally, we assume symmetric bargaining, meaning that the bargaining power of workers is set to 0.5.

4.3 Minimum Distance Estimation

We use a minimum distance procedure to estimate the remaining seven parameters displayed in Table 5. The algorithm minimizes differences between a set of eight moments taken from the data, listed in Table 6, and the

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

equivalent values implied by the model. Formally, the minimum distance estimator is defined as:

$$\hat{x} = \underset{x \in X}{\operatorname{argmax}} \, [\hat{\pi} - h(x)]' \, W \, [\hat{\pi} - h(x)] \tag{9}$$

where $\hat{\pi}$ is the vector of the logarithms of the targets in Table 6, x is a vector of the seven parameters being estimated, $h(x)$ is the mapping from the parameter space X to the model outcomes corresponding to the moments $\hat{\pi}$ (measured in logs), and W is a weighting matrix. We use logs rather than levels to define the distance measure in relative terms, thus reducing concerns regarding the scaling of moments.

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

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

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

Table 6 shows that the estimated model matches all target variables with considerable accuracy. Moreover, the standard errors of the estimated parameters are very small, due mostly to the very large sample sizes from the PME survey.

Before we proceed to the next subsection, it is interesting to use our baseline specification to characterize some properties of the equilibrium, particularly as it relates to the cross-sectional distribution of firms. Each row in Table

Table 5 – Estimated Parameters - Minimum Distance Procedure

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

Table 6 – Moments Used in Estimation

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

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

Table 7 – Firms in the Model

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

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

7 describes firms in a specific position in the distribution of productivity. The top row refers to the smallest firms in the model and the bottom row refers to the largest ones. The columns show the productivity parameter, the number of workers, the fraction of skilled workers, and the compliance status. The model generates an equilibrium where the fraction of skilled workers increases monotonically with firm size (in both the informal and formal sectors) and formal firms are larger than informal ones. This profile reproduces patterns observed in the data (as in Table 2) but incompatible with previous search models of informality: the presence of skilled and unskilled workers in both sectors and a higher share of skilled workers in formal firms. Also interestingly, the smallest firms in the model have approximately one employee, even though this is not imposed as a restriction.

5 Quantitative Results

5.1 The Recent Reduction in Informality in Brazil

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

Throughout the analysis, we often refer to Table 8, where each row contains a particular labor market outcome. The first column describes how the Brazilian labor market changed from 2003 to 2012 using the same data and definitions used in the calibration. Each following column considers how changes in one or more parameters affect labor market outcomes in the model, by comparing the baseline calibration with a new steady-state equilibrium where only the parameters in question are set to their 2012 levels.

In the period we study, the unemployment rate fell by 7.2 percentage points and the informality rate dropped by 10.7 points. Average wages increased by 28%, but, as mentioned in section 2, the gains were larger for low-skill formal workers and for informal workers. Informal wages, for example, increased by 42%, as compared to a wage growth of 22% for formal skilled workers.

5.1.1 Minimum Wage

The minimum wage increased by 61.2% from 2003 to 2012. The effects of a change of this magnitude in the calibrated model are shown in column 2 of Table 8. Wages for skilled workers in both sectors are only marginally affected. However, for informal workers, wages fall by 3.2%. The reason for this decline is the reduced demand for unskilled labor by formal firms, which increases unemployment and lowers the outside option of workers being hired by informal firms.

From the changes in the minimum wage alone, informality increases by 5.7 percentage points and unemployment by 1.6 percentage point. The increase in unemployment seems small when compared to the magnitude of the increase in the minimum wage. The reason is that part of the effect on unemployment is attenuated by marginal

Table 8 – Quantitative Experiments, Changes in the Brazilian Labor Market between 2003 and 2012

Changes in:	(1) Data	(2) Model	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Minimum wage	Payroll tax	Benefits	Enforcement	Fraction skilled	All but productivity	Productivity	All simultaneously
Outcomes		$\Delta\bar{w} = 61.2\%$	$\Delta\tau = -0.63$p.p.	Δa_i and $\Delta b_i \simeq 0$	$\Delta C = 33.9\%$	$\Delta\eta = 12.9$p.p.		$\Delta A = 25.3\%$	
Unemployment (p.p.)	-7.2	1.6	0.0	0.0	0.4	-6.0	-3.1	-3.2	-5.0
Informality (p.p.)	-10.7	5.7	-0.3	0.2	-3.4	-12.6	-8.9	-1.2	-9.1
Wages (%):									
Average	28.4	0.6	0.2	0.0	-0.8	1.2	2.2	24.3	28.4
Formal, skilled	22.4	-1.5	0.3	0.0	0.0	-8.8	-9.6	26.5	14.2
Formal, unskilled	61.2	61.2	0.0	0.0	0.0	9.6	61.2	0.0	61.2
Informal	42.2	-3.2	-0.4	0.3	-8.4	48.5	16.2	44.4	61.7
Product^a (%)	27.0	-1.8	0.1	0.0	0.3	8.0	6.9	26.6	35.5
Govt. net revenues (%)^b	-	-23.8	-1.1	1.4	3.9	7.1	-15.5	40.1	29.7

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

firms entering informality (and thus not being subject to the minimum wage anymore), and also by the fact that informal unskilled wages decrease, leading to increased labor demand by informal firms.

This logic resembles the traditional view of the informal sector, where for some workers informality is an alternative to unemployment (Fields, 1975, Rauch, 1991, 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 more timid increase in unemployment, is in line with evidence from the Brazilian labor market. Though there are no well identified studies of the labor market response to increases in the minimum wage currently available, the existing evidence, such as Foguel, Ramos and Carneiro (2001) and Lemos (2009), seems to indicate that informality tends to rise and employment responds only mildly – if at all – to minimum wage increases. Our comparative statics reproduces the qualitative patterns documented by the empirical literature on minimum wages in Brazil.

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

5.1.2 Payroll Taxes

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

5.1.3 Mandated Benefits

There were minor changes in mandated benefits, specifically in the formulas for calculating the income tax and social security contributions, which are both deducted from the wage of formal employees and thus are included in our parameter a_i. However, on average, they did not result in sizable changes in deductions. When we recalculate the parameters a_i and b_i using 2012 data (Appendix A), we find that the differences are negligible. Hence, they do not have any relevant impact on labor market outcomes, as column 4 from Table 8 shows.

5.1.4 Enforcement of Regulation

We use data from the Ministry of Labor to estimate changes in the enforcement of labor regulations from 2003 to 2012. Reports of labor inspections, available in MTE (2013), show that the number of workers targeted by inspections rose during the last decade both in absolute terms and as a fraction of the workforce.[16] We use the relative increase as a proxy for increases in enforcement in the model. We find that the fraction of the workforce targeted by inspections rose by about 33.9% from 2003 to 2012. We therefore raise the parameter C in the model by this same proportion.

The fifth column of Table 8 shows how this change impacts our baseline calibration. First, informality decreases by 3.4 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 decision may hire more workers, and also an intensive margin effect because the remaining informal firms hire fewer workers. In our calibration, unemployment increases by 0.4 percentage point with the increase in enforcement. The qualitative responses of informality and unemployment generated by the model are consistent with reduced-form evidence from exogenous variations in labor inspections provided by Almeida and Carneiro (2012). The only noticeable change in wages is a decline in earnings for informal workers. In this respect, our model replicates the results found in Bosch and Esteban-Pretel (2012) and Meghir, Narita and Robin (2015). Government revenues increase, but one should be cautious about drawing additional implications from this result since we do not take into account operational costs associated with increased enforcement.

5.1.5 Workforce Composition

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

We find that the predicted changes are in line with our discussion from section 2. Both unemployment and informality decrease sharply as a consequence of a more skilled workforce, falling, respectively, by 6 and 12.6 percentage points. Wages for informal workers increase by 48.5%, while they decrease for skilled formal workers by 8.8%. This is a direct consequence of the relative increase in the supply of skilled workers. The labor market for skilled workers becomes less tight (and the reverse happens for unskilled workers). Because firms hire more skilled labor in the new equilibrium, the productivity of unskilled work increases due to complementarities in the production function. The combination of a tighter labor market for unskilled labor and higher productivity is behind the steep increase in the informal wage. Wages for unskilled formal workers also rise, by 9.6%, meaning that the minimum wage is not binding anymore in the new equilibrium. The model therefore predicts that, absent the observed increases in minimum wages between 2003 and 2012, the minimum wage would have become non-binding under the 2012 composition of the Brazilian labor force.

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

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

We know of no reduced-form empirical study that analyzes the aggregate labor market effect of changes in the educational composition of the labor force. Various papers, such as Menezes Filho, Mendes and de Almeida (2004), describe the strongly positive individual-level correlation between schooling and formality. Other papers, such as Barbosa Filho and Moura (2012), assume a stable individual-level relationship between schooling and informality and perform Oaxaca-Blinder type exercises analyzing the role of demographic changes as determinants of changes in informality. But no paper allows for the possibility that changes in the educational composition of the labor force directly affect labor market equilibrium outcomes, conditional on individual schooling. This highlights the relevance of the type of analysis conducted in this paper, where we can systematically address the endogenous labor market response to this type of compositional changes.

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

5.1.6 Estimating Changes in Productivity

Now we consider the performance of the model when the five dimensions discussed above are brought together. The results are shown in column 7. These changes explain 83% of the observed decline in informality, but only 43% of the decline in unemployment. Also, average wages and product increase by far less than observed in the data, explaining in both cases just a fraction of the actual change. The increase in GDP per capita in the data – listed as product in the table – reflects in part an overall increase in TFP in the Brazilian economy during this period. Ferreira and Veloso (2013), for example, estimate an yearly growth of TFP in Brazil between 1.5% and 2.5%

per year from 2003 to 2009. These observations suggest that there was an increase in overall productivity in the economy that, not surprisingly, cannot be captured by the model, which does not display capital accumulation or technological change. To calibrate TFP gains in the model, we raise the parameter A until the increase in average wages – taking into account all changes during the period – match that observed in the data. We find that TFP in the model must increase by 25.3% between 2003 and 2012 in order for the model to reproduce the increase in average wages in the data. This number falls close to the upper bound of the cumulative growth in TFP that would be obtained from the Ferreira and Veloso (2013) estimates.

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

5.1.7 Explaining the Evolution of Labor Market Outcomes

In column 9, we consider changes in minimum wages, taxes, benefits, enforcement, skills, and productivity simultaneously. The qualitative implications of the model, in terms of direction of predicted changes, matches precisely the pattern of movements observed in the Brazilian labor market between 2003 and 2012: reductions in unemployment, reductions in informality, and increases in average wages, with proportionally larger gains for informal and unskilled workers. Quantitatively, the model does a good job in explaining the reduction in informality, generating a decline of 9.1 percentage points while the observed decline was 10.7 points. It also predicts a decline in unemployment of 5 percentage points, which corresponds to 69% of the observed decline of 7.2 points. Predictions regarding wages are close to the empirical patterns, though the model underestimates by more than a third the gains for formal skilled workers. Overall, the model is able to explain quantitatively the main outcomes of the Brazilian labor market with a reasonable degree of precision.

Going back to the discussion in section 2, we can use the model to determine which factor was the main driver behind the reductions in informality and unemployment. Table 8 already addressed this issue, by looking at the effect of each factor one at a time. In Table 9, we conduct the opposite comparative statics exercise: we analyze what happens in the model when all but one of the factors discussed before is taken into account. We find that the declines in both unemployment and informality would have been considerably larger – respectively, 6.5 and 11 percentage points – if the minimum wage had not increased. We also reinforce the idea that the change in labor force composition was the main driver behind the observed reductions in informality: without a larger fraction of skilled workers, informality would have increased by 4 percentage points, instead of declining by 9.1, and unemployment would have remained roughly stable. In short, the model is unable to reproduce the reductions in informality and

Table 9 – Individual Contribution of Each Factor, Changes in the Brazilian Labor Market from 2003 to 2012

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

Notes: [a]Change from 2003 to 2007 (IBGE/SCN is only data available up to 2007). [b]Product is total production in the model net of search costs and the informality penalty.

unemployment when changes in labor force composition are ignored. The relevance of enforcement (informality penalty) is of second order: without changes in this parameter, the decline in informality would have been three percentage points smaller. As before, the effects of changes in payroll taxes and benefits are negligible.

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

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

Table 10 – Hypothetical Policy Experiments

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

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

effect of increased productivity on informality is ambiguous. Our results suggest that this theoretical ambiguity may indeed be quantitatively relevant. It also shows that increases in productivity alone are not enough to rationalize the changes seen in the Brazilian context.

5.2 Policy Experiments

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

The first policy we consider is a reduction in payroll taxes for low wage workers. In column 3 from Table 8, we showed that a lower payroll tax rate can lead to a decline in informality with no adverse effect on unemployment. On the other hand, it also leads to a reduction in government revenues that is substantial when compared to the decline in informality. However, informal firms are relatively more intensive in unskilled labor. In addition, only a fraction of government revenues come from payroll taxes on low skill workers, since their wages are lower and they account for a small fraction of formal employment. Thus, an intermediate alternative might be for governments to subsidize the employment of low wage formal workers through a progressive payroll tax, with the tax rate increasing with the wage. Proposals like this have been considered as ways to subsidize low wage workers in developed countries (see Lee and Saez, 2012), but rarely feature in the informality discussion in the developing world.

In Table 10, we examine the progressive payroll tax policy using as starting point the model as of 2012 (column 9 in Table 8). In the first column, we show as a reference point the result of simply reducing the overall tax rate by 1 percentage point (to 0.7043). As argued above, although this reduction leads to reductions in informality, there are significant costs in terms of government revenue. In columns 2 and 3, we consider similar policies where the reduction in payroll taxes is restricted to low wage workers (in the model, equivalent to low skill). The policy achieves similar or better results for employment and formalization and, in addition, for some values of τ_u government revenues actually increase. The formalization induced by lower taxes among low skill workers is sufficient to induce marginal firms to comply, and thus enlarges the tax base. The taxes raised from skilled workers in firms that formalize more than offset the foregone revenue from low skill workers in infra-marginal firms. On top of that, wages increase substantially for unskilled workers in the informal sector because of a tighter labor market. This policy therefore is also likely to have positive effects on poverty alleviation.

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

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

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

We believe that these concerns are not enough to compromise the qualitative implications of the analysis, though the quantitative results from Table 10 should not be taken at face value. To assess the relevance of the first two issues, we examine data from the ECINF, which surveyed small firms in the formal and informal sectors in Brazil. For each of the small firms covered by ECINF, we have information on number of employees, formal status, wages, and schooling levels. Regarding the first point, we examine the degree of wage dispersion within firms in the informal sector. In 64% of the informal firms with five employees – the largest firms surveyed by ECINF and those more likely to be marginal firms in the informal sector – the highest wage was at least 50% above the lowest wage. In 20% of them, the highest wage was more than three times the lowest wage. The data also show that, in most of these firms, workers belong to very different educational categories. This evidence suggests that there is a substantial degree of skill heterogeneity within marginal informal firms, as implied by the model.

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

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

6 Concluding Remarks

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

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

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

References

Albrecht, James, Lucas Navarro, and Susan Vroman. 2009. "The Effects of Labour Market Policies in an Economy with an Informal Sector." *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, Vladimir P. Ponczek, and André P. Souza. 2016. "Informal wages in an economy with active labor courts." *Applied Economics*, forthcoming.

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

Barbosa Filho, Fernando Holanda, and Rodrigo Leandro Moura. 2012. "Evolução Recente da Informalidade no Brasil: Uma Análise segundo Características da Oferta e Demanda de Trabalho." *Texto para discussão IBRE-FGV*, 17.

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

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

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

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

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

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

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

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

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

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.

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

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

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.

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

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

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

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

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

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

Galiani, Sebastian, and Frederico Weinschelbaum. 2012. "Modelling Informality Formally: Households and Firms." *Economic Inquiry*, 50(3): 821–838.

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

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

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

IBGE. 2005. "Estatísticas 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(5): 1960–1976.

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(9-10): 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(2): 224–237.

Lucas, Robert E. Jr. 1978. "On the Size Distribution of Business Firms." *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: 1509–1546.

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

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

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

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

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

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

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

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

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

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

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

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

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

Ulyssea, Gabriel. 2014. "Firms, Informality and Welfare." Unpublished manuscript.

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

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

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

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

Costs of Formal Labor

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

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

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

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

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
FGTS balance on dismissal (B.2)	*B.1 x average duration*	*304.36*
Severance payment	50% of B.2 x prob. dismissal	4.58
FGTS temporary extra	0.5% of B	0.57
Employer INSS contribution	20% of B	22.89
SAT, INCRA, S system	5.3% of B	6.07
Total with contributions (C)		**157.72**
Vacation adjustment	1/11 of C	14.34
Total cost		172.06
Payroll tax rate (τ)		**0.7206**

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

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

Valuation of Mandated Benefits

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

Table A.2 shows our calculations of the value attributed to benefits and contributions that calculated as fractions of the base salary. When taken together, these regulations compose the variable benefits parameters in the $benefits$ expression, a_s and a_u. The first five rows are similar to those in Table 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 raw total wage: the social security (INSS) deduction and the income tax (IRPF). For the low wage workers, we use the lowest brackets: zero income tax in both years and social security deductions of 7.65% (in 2003) or 8.00% (in 2012). For the high wage workers, we calculate the deductions

Table A.2 – Valuation of Variable Benefits

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

for each individual worker in the PME data set that receives more than the minimum wage, using the corresponding tax rates and brackets in each year. Then, we calculate the average deduction per worker.

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

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

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

Finally, we calculate the unemployment insurance parameters (b_s^D, b_u^D). Unemployed workers who were previ-

Table B.1 – Informality Trends per Economic Activity

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

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

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

Appendix B: Informality Trends by Economic Activity

In this Appendix, we show that the decline in the informality rate in Brazil was widespread in the economy, and also that it was not caused by reallocation of workers across sectors. In the PME survey, workers report the economic activity to which their main job belongs, choosing one of 60 categories. In Table B.1, we list 15 economic activities with the largest number of workers. Together, they account for 76% of the workforce in 2003, and 78% in 2012. For each activity, we compute the formality rates in 2003 and 2012, and also the share of the workforce employed therein. Note that, since the PME targets workers in large metropolitan areas, few of them are employed in agricultural or extractive activities.

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

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

where $P_{i,t}$ and $F_{i,t}$ denote the share of the workforce in and the formality rate of activity i in year t, respectively. The sum of the within contributions describe what would happen if the share of workers in each activity remained constant from 2003 to 2012, but the formality rates within each activity changed. The sum of between contributions accounts for the part of the decline in informality that can be attributed to changes in the size of each activity, given the formality rates in 2012. As can be seen in the bottom row of Table B.1, the decline in informality can be accounted for almost exclusively with changes within each activity.

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

Appendix C: Solution to the Problem of the Firm

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

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

Also, differentiating the value function in n_i yields:

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

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

Appendix D: Solution to the Wage Bargaining Equation

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

The Nash bargaining equation is:

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

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

$$c_i w_i(n_s, n_u) = (1-\sigma)(rU_i - b_i) + \sigma \left[F_i(n_s, n_u) - (1+\tau_s) n_s \frac{\partial w_s(\cdot)}{\partial n_i} - (1+\tau_u) n_u \frac{\partial w_u(\cdot)}{\partial n_i} \right] \tag{10}$$

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

The first step to solve this system is to write it in a more convenient way. Taking the partial derivative of 10 with respect to n_u when $i = s$ yields:

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

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

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

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

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

Using the definition of c_i, we obtain:

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

Which we can use to write the system of equations defined in 10 as:

$$c_i w_i(n_s, n_u) = (1-\sigma)(rU_i - b_i) + \sigma \left[F_i(n_s, n_u) - (1+\tau_i) \left(\chi_{i,s} n_s \frac{\partial w_i(\cdot)}{\partial n_s} + \chi_{i,u} n_u \frac{\partial w_i(\cdot)}{\partial n_u} \right) \right] \tag{11}$$

where

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

Following Cahuc, Marque and Wasmer (2008) (henceforth CMW), we first solve the equation for the case in which $\chi_{i,j} = 1$. Later, we generalize the solution. The insight in CMW is to perform a change of coordinates that allows us to express the term multiplying $(1+\tau_i)$ in equation 11 in a simpler manner, effectively obtaining a univariate differential equation as the result. The transformation we need is:

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

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

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

Which is the term multiplying $(1+\tau_i)$ in equations 11 if $\chi_{i,j}=1$. Following the same notation, let $\hat{F_{n_i}}(\rho,\phi) = \frac{\partial F(\rho\cos\phi,\rho\sin\phi)}{\partial n_i}$ denote the marginal product function in the new coordinate system. We can then rewrite the differential equations as:

$$\frac{\partial\hat{w}_i(\rho,\phi)}{\partial\rho}+\frac{c_i}{\sigma(1+\tau_i)\rho}\hat{w}_i(\rho,\phi) = \frac{1-\sigma}{\sigma(1+\tau_i)\rho}(rU_i-b_i)+\frac{1}{(1+\tau_i)\rho}\hat{F_{n_i}}(\rho,\phi) \tag{12}$$

We guess the following form for the solution:

$$\hat{w}_i(\rho,\phi) = C(\rho,\phi)\rho^{-\frac{c_i}{\sigma(1+\tau)}}+D(\phi) \tag{13}$$
$$\frac{\partial\hat{w}_i(\rho,\phi)}{\partial\rho} = C'(\rho,\phi)\rho^{-\frac{c_i}{\sigma(1+\tau)}}-C(\rho,\phi)\frac{c_i}{\sigma(1+\tau)}\rho^{-\frac{c_i}{\sigma(1+\tau)}-1}$$

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

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

We can integrate the latter equation to obtain:

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

Replacing in 13, we get:

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

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

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

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

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

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

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

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

Also, we further impose that

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

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

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

We only need one solution, the simplest being

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

since $1/\chi_{i,j} = \chi_{j,i}$. Then, using $\partial F/\partial n_j = \chi_{j,i} n_j^{\chi_{j,i}-1} \partial \tilde{F}/\partial M_{i,j}$ and $\partial F/\partial n_i = \partial \tilde{F}/\partial M_{i,i}$ as $\chi_{i,i} = 1$, the system 11 can be rewritten as

$$c_i \tilde{w}_i(M_{is}, M_{iu}) = (1-\sigma)(rU_i - b_i) + \sigma\left[\frac{\partial \tilde{F}(M_i)}{\partial M_{ii}} - (1+\tau_i)\left(M_{is}\frac{\partial \tilde{w}_i(M_i)}{\partial M_{is}} - M_{iu}\frac{\partial \tilde{w}_i(M_i)}{\partial M_{iu}}\right)\right] \tag{14}$$

System 14 is equivalent to system 11 in the case where $\chi_{i,j} = 1$. Thus, the solution for $\tilde{w}_i(M_{is}, M_{iu})$ is known:

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

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

$$w_i(n_s, n_u) = \frac{1-\sigma}{c_i}(rU_i - b_i) + \frac{1}{1+\tau_i}\int_0^1 z^{\frac{1-\sigma}{\sigma}\frac{a_i}{1+\tau_i}}\frac{\partial F\left(z^{\frac{1+\tau_s}{a_s}\frac{a_i}{1+\tau_i}}n_s, z^{\frac{1+\tau_u}{a_u}\frac{a_i}{1+\tau_i}}n_u\right)}{\partial n_i}dz \tag{15}$$

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

$$\frac{\partial w_i(n_s, n_u)}{\partial n_j} = \frac{1}{1+\tau_i} \int_0^1 z^{\frac{a_i}{1+\tau_i}\left(\frac{1-\sigma}{\sigma}+\frac{1+\tau_j}{a_j}\right)} \frac{\partial^2 F\left(z^{\frac{1+\tau_s}{a_s}\frac{a_i}{1+\tau_i}} n_s, z^{\frac{1+\tau_u}{a_u}\frac{a_i}{1+\tau_i}} n_u\right)}{\partial n_i \partial n_j} dz \tag{16}$$

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 15 can be stated as:

$$a_i w_i(n_s, n_u) = (1-\tilde{\sigma}_i)(rU_i - b_i) + \frac{a_i}{1+\tau_i} \int_0^1 z^{\frac{1-\tilde{\sigma}_i}{\tilde{\sigma}_i}} \frac{\partial F\left(z^{\frac{1+\tilde{\sigma}_i}{\tilde{\sigma}_i}\frac{\tilde{\sigma}_s}{1-\tilde{\sigma}_s}} n_s, z^{\frac{1+\tilde{\sigma}_i}{\tilde{\sigma}_i}\frac{\tilde{\sigma}_u}{1-\tilde{\sigma}_u}} n_u\right)}{\partial n_i} dz \tag{17}$$

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

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

Finally, note that, although we have assumed the same bargaining power for all workers, it is immediate to extend it to the more general case with type-specific bargaining power. This would lead to an expression similar to 17, but with $\tilde{\sigma}_i = \frac{\sigma_i(1+\tau_i)}{\sigma_i(1+\tau_i)+(1-\sigma_i)a_i}$. Similarly, extending the solution to more than two types of workers would be trivial, requiring essentially a change in notation. See CMW, in particular how they define the matrix $\mathbf{NA}_i(z)$.

Minimum Wages and Wage Bargaining

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

Figure D.1 shows an example of how wages can be affected by the minimum wage according to firm size. For small values of n_s and n_u, marginal productivities are high and bargained wages are above the minimum wage. As the quantity of either type of worker increases, it is possible that marginal productivities decrease so much that the minimum wage binds. For high values of both of inputs, it is possible that all wages equal the minimum wage. In this example, the curves are upward sloping because there is complementarity between labor types ($\frac{\partial^2 F^z(n_s,n_u)}{\partial n_s \partial n_u} > 0$). They would be straight or downward sloping if that cross derivative was null or negative, respectively.

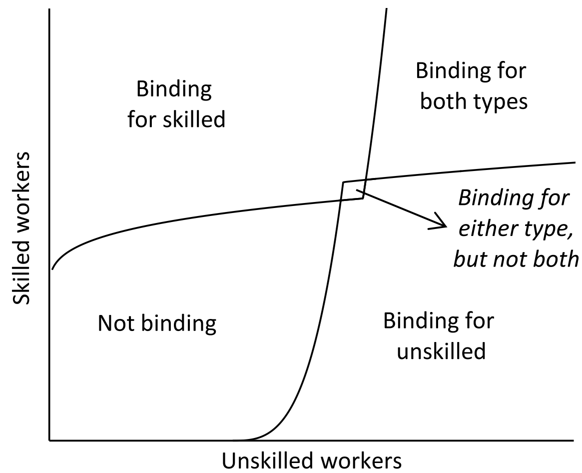

Figure D.1 – Minimum Wage Status According to Firm Size

It is also possible that, for certain values of (n_s, n_u), there is multiplicity of wages satisfying the bargaining conditions: either type of worker might receive the minimum wage, but not both. This pathology is caused by discontinuities in the marginal value of workers which we discuss below. In our applications, there is no possibility that the minimum wage binds for the skilled, no matter how many workers of this type are hired. The reason is that the first term in the wage equation 15, related to the reservation wage, is strictly greater than the minimum wage in all simulations. Hence, we are not concerned about this multiplicity problem.

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

To find the correct skilled wage function in this case, we observe that the differential equation 10 simplifies to:

$$c_i w_s(n_s, n_u) = (1-\sigma)(rU_i - b_i) + \sigma \left[F_s(n_s, n_u) - (1+\tau_s) n_s \frac{\partial w_s(n_s, n_u)}{\partial n_s} \right] \tag{18}$$

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 12. The solution is analogous:

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

Note that skilled wages are still a function of the number of both skilled and unskilled workers, but not the same function as before. When the cross derivative of the production function $\frac{\partial^2 F^z(n_s,n_u)}{\partial n_s \partial n_u}$ is positive, as in our quantitative exercises, then we should expect this new wage function to be strictly greater than the unconstrained one for the same values of n_s and n_u. The reason is that, in the unconstrained case, hiring an additional skilled worker leads to an increase in unskilled wages due to the effect in the unskilled marginal productivities, which

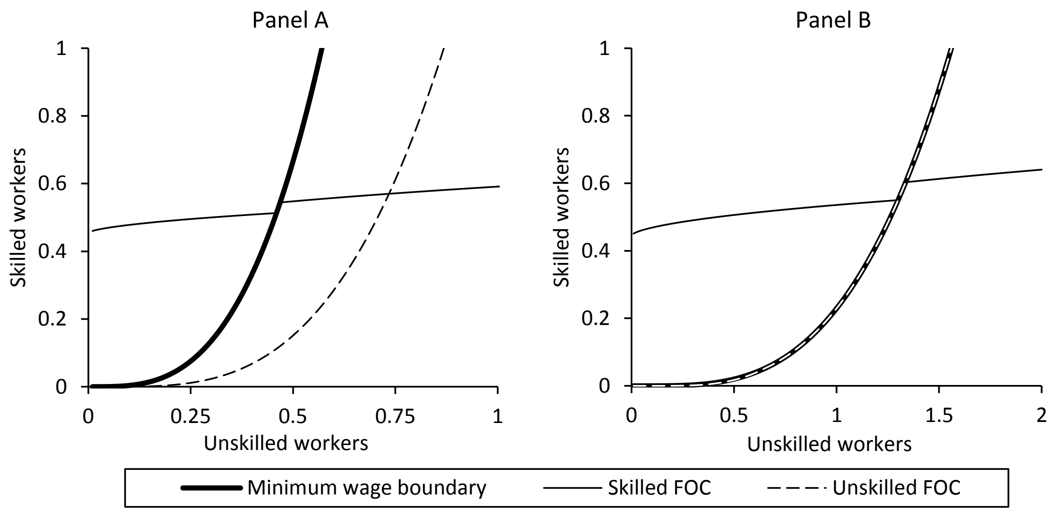

Figure D.2 – Problem of the Firm and Minimum Wages

reduces the surplus being bargained over (from the point of view of the firm and the single skilled worker with whom it is bargaining). This "negative" effect does not exist (at the margin) when the minimum wage binds: the surplus is bigger, and so are bargained wages. Note that this implies a discontinuity in the wage function at the points that separate the regions where the minimum wage is or is not binding.

Minimum Wages and the Solution to the Problem of the Firm

Finally, we discuss how the existence of the minimum wage might change the problem of choosing the optimal firm size. The discontinuity in the wage function, discussed above, is caused by discrete changes in the net marginal value of workers $J_i^{for}(\cdot)$ (see equation 2) 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 D.2, we show how the minimum wage can affect the problem of the firm. In Panel A, we illustrate the problem of a formal firm with average productivity ($z = 1$) in our baseline calibration. The heavy solid line marks the transition between a non-binding and a binding minimum wage for the unskilled workers – that is, it is the vertical line in figure D.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, $J_s^{for}(n_s, n_u)$, is equal to the expected search cost $\frac{(r+s^{for})\xi}{q(\theta_s)}$. Above this line, there are too many skilled workers, which drives down their marginal productivity and makes the marginal value less than the search cost. The same reasoning is valid for the dashed line: to the right of it, the marginal value of unskilled workers is less than the expected search cost, and the converse is true to the left of the line. As before, the upward slope of all curves comes from complementarity between labor inputs.

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

Panel B describes a case in which there is no solution to the problem of the firm because of the discontinuities associated with the minimum wage. It follows from a change in the baseline model that increases overall productivity (parameter A in the quantitative experiments section), making the minimum wage binding by a smaller margin. The difference between Panel B and Panel A is that the discrete fall in the marginal value of the unskilled workers causes it to drop from a number strictly greater than the expected search costs to another strictly less than it. As a consequence, there is no point in the graph in which the unskilled first order condition is satisfied. The skilled first order condition is not satisfied either at the intersection of the three lines.

In such situation, the firm would strategically choose a point to the left of that intersection (where the minimum wage does not bind), since bargained wages for skilled workers would be discontinuously lower than immediately to the right of the intersection. There is no similar discontinuity in unskilled wage because it cannot drop below the minimum wage, and thus unskilled wages are approximately equal on both sides of the boundary. In our numerical applications, the optimal firm size in those situations is chosen by finding the point (n_s^*, n_u^*) that satisfies the first order condition for skilled workers and lies immediately to the left of the discontinuity.[17]

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

Appendix E: Numerical Procedures

In this appendix, we describe the numerical procedures required to solve for an equilibrium and to perform the minimum distance estimation presented in our quantitative exercises. The descriptions below include general overviews of the procedures as well as specific operational details. Figure D.3 provides a sketch of how each step of the numerical implementation of the paper relate to each other. In the first subsection of this Appendix, we describe the procedure to solve for equilibrium, comprising the three "lower" levels of the hierarchy described in Figure D.3. Next, we discuss the estimation procedure.

[17]In practical terms, our algorithm first tries to solve the problem of the firm using a derivative-based method. If it cannot find the solution, it solves the system given by the skilled FOC and the equation $w_u^{for}(n_s, n_u) = \bar{w}$ (the solid line in figure D.2). After solving this system, the algorithm checks if the solution is such small deviations in the number of unskilled workers make $J_u^{for}(n_s, n_u) - \frac{(r+s^{for})\xi}{q(\theta_u)}$ change sign. If not, n_u is increased or decreased, depending on the sign of $J_u^{for}(n_s, n_u) - \frac{(r+s^{for})\xi}{q(\theta_u)}$, until the condition just described is satisfied. The optimal solution is the smallest value of n_u (in a finite grid with intervals given by the numerical tolerance) such that $J_u^{for}(n_s, n_u) - \frac{(r+s^{for})\xi}{q(\theta_u)} > 0$, as a strategic firm would choose.

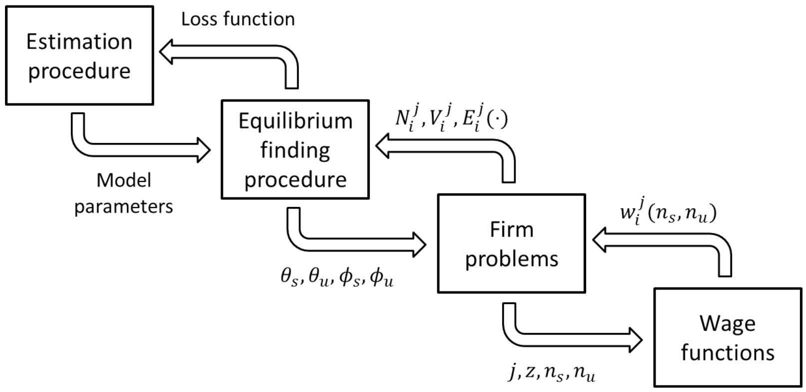

Figure D.3 – Hierarchy for the numerical procedures.

Solving for the Equilibrium

Solving the model numerically is equivalent to finding values for θ_s, θ_u, U_s and U_u that solve equations 7 and 8 up to a desired numerical precision – that is, that set a residual term implied by the given equations to less than a tolerance value. Two observations make this computation easier. First, from 8, one can see that U_i is a function of only θ_i and parameters of the model when the minimum wage does not bind for workers of type i. This suggests a procedure to minimize the dimension of the problem: guess whether the minimum wage binds for each type of worker, and calculate U_i directly from θ_i if the minimum wage is not binding for type i, instead of using it as a choice variable.

In the estimation procedure described in the next subsection, since the target economy is one where the minimum wage binds only for unskilled workers, we impose this restriction in the equilibrium procedure to reduce computational time.[18] This approach is not problematic if there is a neighborhood of the parameter space around the optimal point where the minimum wage is always binding for the unskilled only. However, when running the counterfactual exercises, we allow for any combination of minimum wage status. To do so, we sequentially solve the model given one of potentially four assumptions about minimum wages, until a solution such the assumption holds is found. Checking if the assumption holds is simply a matter of checking whether freely bargained wages are above or below the minimum wage.

The second observation that helps with the computation of an equilibrium is that, when the minimum wage binds for workers i, using ϕ_i as a choice variable is easier than choosing U_i directly. This is because ϕ_i is a dimensionless ratio, bound by 0 and 1. The corresponding value of U_i is obtained from ϕ_i and θ_i, using equation 8.. So, if the minimum wage is guessed to bind only for unskilled workers, for instance, then the problem of finding an equilibrium is to choose θ_s, θ_u and ϕ_u that set three residual terms to zero.

To calculate the residual terms associated with a given choice of θ_i and ϕ_i, we solve the problem of all firms, aggregate all employment and vacancy decisions, and then calculate the relative differences between the choice variables and the corresponding values implied by the aggregates.[19] In the model, the distribution of firm produc-

[18] As discussed later in that section, it also makes the loss function less prone to discontinuities.

[19] We use relative measures of the difference between the LHS and the RHS of equations 7 and 8 as the residuals to be set to zero in the optimization procedure. The LHS is the value of θ_i or rU_i (implied by ϕ_i) that is the "guess" taken as given when solving the firm problem, and the RHS is calculated using aggregates obtained after solving the problem of the firm for all firms and interpolating the results. The procedure is not substantially affected by the specific functional form of the residual (e.g. log(RHS/LHS) or RHS/LHS-1).

tivities, $G(z)$, is continuous, but since a closed form solution is not available we need to discretize it in the numerical procedure. To reduce computational time while still maintaining "smoothness" down to the desired tolerance levels, we use a interpolation procedure. Namely, we solve the model for a relatively small number of firm types (20) and then interpolate quantities like labor demand and profits over a much finer grid of productivity levels (100,000) using cubic splines. The quality of the approximation can be tested by solving the problem for in-between levels of productivity and comparing the solution to the interpolated value. We found the interpolation to be very reliable, which is expected given the smoothness of $G(z)$ and the continuity of the production function on z.[20] [21]

Given that $G(z)$ is a Generalized Pareto distribution in our quantitative exercises, we need to truncate it at the top in order for the interpolation procedure to work. We choose 30 as the upper bound (for reference, remember that the distribution is constrained to have mean 1). At the estimated value for the distribution in the quantitative exercises, the mass above that threshold for the non-truncated distribution is less than 0.001%. The results are not sensitive to changes in the threshold (conditional on re-estimating the model if the changes are relatively large).[22]

When solving the problem of an individual firm type, we use a standard optimization procedure to solve the first order conditions (or for the strategic solution described at the end of Appendix D) taking the chosen values of θ_i and U_i as given. This involves using a numerical integration procedure for the integral terms in the expressions for wages and their derivatives. We used a trapezoidal rule with 1,000 trapezoids in a uniform grid. The firm's compliance decision is determined by comparing formal and informal profits for each level of productivity z, after the interpolation is done.

Finally, by integrating vacancies and employment along the discrete distribution of productivity, we can calculate what are the implied tightnesses, θ_i, and reservation wage, rU_i, using equations 7 and 8. Note that this computation is not possible if the initial guess for θ_i is too low, since it can lead to levels of employment greater than the measure of the workforce. In this case, a larger initial value for θ_i should be provided. After we find the θ_i and ϕ_i that solve the equilibrium equations, we can verify whether the initial guess for which minimum wages bind is correct. If so, an equilibrium has been found. If not, a different guess must be tried.[23]

Given that θ_i is a ratio and must be greater than zero, we use $z_i = \log(\theta_i)$ as the choice variable in the optimization procedure instead of θ_i itself, thus eliminating the need for constrained optimization.

[20] The model provides a direct test for the quality of the interpolation. First, calculate U_i using individual wages for all 100,000 interpolated firms and the first expression in 8 . Then, compare this value to the results found assuming that the FOC holds for all firms (the other expressions). To the extent that the interpolated values do not necessarily solve the FOC, there might be a discrepancy between these two ways to calculate U_i. In our baseline calibration, the relative difference is at most 0.14%.

[21] Even though the discrete nature of the firm distribution makes the problem non-smooth, derivative-based methods usually work well if the initial guess is close enough to the solution. It is important, though, to use a relatively large change in the choice variables when calculating the numerical derivatives, compared to the number of atoms in the firm productivity distribution. If the change in parameters is too small, then it's unlikely that any of the "marginal atoms" of the distribution will shift its compliance decision, even when the choice variable is relevant for that decision. In this case, the effect of changing the choice variables might be biased by not taking into account extensive margin effects. See the "A note on tolerance levels" subsection below.

[22] Additional details on handling the productivity distribution:

We first obtain a vector of 100,001 values uniformly distributed from 0 through 30. This leads to 100,000 intervals whose bounds are the elements in that vector. Given the shape parameter for $G(z)$, we can calculate the CDF at each of the 100,001 points, and thus the probability mass associated with each of the 100,000 intervals (remembering to normalize so that the probabilities add up to 1). Finally, we calculate the mean of the continuous distribution $G(z)$ conditional on lying within each interval. We use these conditional means as the value of z associated with the interval, for the sake of increased precision (instead of, for instance, using midpoints). Thus, while the bounds of the intervals remain fixed, both the probability mass function and the values of z used by the model change as the shape parameter of the distribution changes. Of course, given the large number of intervals, the changes in the z vector tend to be minor.

[23] Additional details on solving the equilibrium set of equations:

Before starting the derivative-based method, we use a simple heuristics to approximate the solution given the size of the residuals, increasing or decreasing θ_s or θ_u if there is excess demand or supply for that kind of workers, respectivelly. After the residuals are relatively small, the derivative-based method is called. It is possible that the discreteness of the productivity distribution implies non-existence of an equilibrium for a given tolerance level. The choice of the granularity of the discrete productivity distribution must take this problem into account.

Estimation Procedure

The numerical procedure implemented in the previous subsection can be seen as a function mapping from the space of parameters to the space of moments and quantities implied by the model. In the estimation procedure, we focus on 7 parameters, listed in Table 5, and 8 moments, listed in Table 6. Let us denote these parameters as a vector $x \in X \subset \mathbb{R}^7$, the moments calculated from the model as a function $h : X \to \mathbb{R}^8$, and the value of these moments in the population of interest (in terms of logarithms) as π. Under the assumption that there is a unique vector of parameters x_0 such that minimizes $[\pi - h(x)]' W [\pi - h(x)]$, we can obtain an estimate of x_0 by solving the following minimization problem:

$$\hat{x} = \underset{x \in X}{\operatorname{argmax}} \, [\hat{\pi} - h(x)]' W [\hat{\pi} - h(x)] \tag{19}$$

where $\hat{\pi}$ is an estimate for π and W is a symmetric weighting matrix.

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

$$AVAR\left[\sqrt{N}\,(\hat{x} - x_0)\right] = \left[H(x_0)'WH(x_0)\right]^{-1} H(x_0)'WVWH(x_0)\left[H(x_0)'WH(x_0)\right]^{-1} \tag{20}$$

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.

The assumption that $h(x)$ is differentiable does not hold strictly, given that the model's equilibrium is solved numerically by discretizing the distribution of firms. In particular, as marginal firms change discretely into and out of informality, the model outcomes also change discretely. However, given the fine granularity of the firm's distribution (100,000 atoms), we expect our numerical implementation to be a very good approximation of the continuous case. In the standard error estimations, we do not explicitly account for this additional source of imprecision. However, we do verify that results are not sensitive to choosing a larger number of atoms in that distribution.

Another potential source of non-smoothness are transitions into and out of different minimum wage regimes. As explained in the model description, changes in minimum wage can lead to discrete changes in other wages and in employment decisions. We avoid this problem by focusing in the case in which the minimum wage binds for unskilled workers, and disregarding all other cases in the estimation procedure. We also make sure that the initial points satisfy this constraint.

Given this econometric framework, we must complete five tasks in order to estimate x_0: choosing the functional form of the moments, using the data to obtain $\hat{\pi}$ and $\hat{V}$, choosing the weighting matrix, solving the optimization procedure, and calculating the covariance matrix. Below we lay some additional details on each step.

1) Choosing Moments and Functional Forms

The rationale for the specific choice of the 8 target moments is explained in the main text. However, instead of targeting them directly, we define π as the log of these moments. This choice avoids problems related to the scaling of the moments: we focus on relative gaps between the model outputs and the targets, rather than on simple differences.

2) Obtaining $\hat{\pi}$ and $\hat{V}$

The first six of the eight moments listed in Table 6 are calculated directly from the PME survey dataset. To obtain $\hat{\pi}$, we calculate the weighted mean of the desired moments in the sample (with the weights given by survey sampling weights), and then take the log of these means. To obtain the covariance matrix of the moments (in levels), we first estimate each variance or covariance without using weights. Then, we multiply these variances and covariances by a factor $K = n\frac{\sum_i w_i^2}{(\sum_i w_i)^2}$, where w_i is an individual weight, to account for the weighting. This factor needs to be calculated separately for each specific variance or covariance, taking into account only the relevant sample for those variables (e.g. only informal workers when assessing informal wage).

It is also important to note that, given the panel structure of the PME, the observations cannot be assumed to be independent. We take the most conservative approach possible and bundle all observations for the same worker as one, taking a weighted mean of outcomes within individual and adding up the weights (separately for each statistic). This procedure ensures that, while the aggregate means match what a researcher would find by using the sampling weights and pooled data, each worker counts as only one observation for the purposes of calculating the covariance matrix $\hat{V}$.

After the matrix of covariances for the moments is calculated, we pre- and post-multiply it by a diagonal matrix where terms in the diagonal are the inverse of the corresponding $\hat{\pi}$ term. This is just a delta-method adjustment to obtain the covariance matrix for the moments in logs. The result is the 6×6 top-left component of the matrix $\hat{V}$.

The seventh moment is the relative wage difference between the minimum wage and informal wages for unskilled workers, which we proxy by a quantile regression result from Bargain and Kwenda (2011). Specifically, we take the result from Table 3, for informal salaried, in specification 4: Panel, fixed effects quantile regression. The log wage penalty is -0.078 and its standard error is 0.004. We consider this estimate to be uncorrelated with the other moments, so that the non-diagonal terms of the seventh row and column are zero.

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

3) Choosing the Weighting Matrix

The results shown in the paper use the identity matrix as the weighting matrix W. We consider this to be a conservative and intuitive approach, while also sidestepping the need to calculate an additional component to the covariance if the model is mispecified (see Chamberlain (1994)). To verify the sensitivity to this choice, we re-ran the optimization procedure starting from the estimated results but using the optimal weighting matrix $\hat{V}^{-1}$. The resulting estimates were very similar to the ones shown in the paper, with no parameter changing by more than 5%.

4) Solving the Minimization Problem

Once the $h(x)$ function is defined and both $\hat{\pi}$ and W are available, estimating $\hat{x}$ requires solving the problem defined in 19 up to the desired precision. We transform the input parameters to circumvent the need for constrained optimization. We use a logarithmic transformation for parameters that should be positive but unbounded, and a logit transformation for parameters that must lie in the $(0, 1)$ interval.

A different set of starting points should be used, since it cannot be guaranteed that any local minimum is the global minimum. We automate this process by using the genetic algorithm optimization tool from Matlab(R), with a randomly drawn initial population of 50 points. The initial points are drawn uniformly over a specific range of parameters values,[24] designed to contrain the initial points to have (i) a binding minimum wage for unskilled workers, (ii) non-binding minimum wage for unskilled workers, and (iii) both informal and informal firms in equilibrium. After finding the equilibrium for the randomly drawn points, we discard those that do not satisfy any of these three criteria. Even though the initial population is constrained to these bounds, the genetic algorithm can "escape" it through mutations. We use the best point after 10 generations as the starting point (after noting that all points with low values for the loss function appear to be near, suggesting a global minimum).

After the initial point was defined, we used a standard nonlinear minimization procedure (Matlab(R)'s fmincon) to estimate $\hat{x}$.

It is worth noting that the layered structure of the equilibrium calculation makes it computationally demanding. Each evaluation of $h(x)$ requires numerically solving a set of three equations. In turn, each evaluation of this set of equations requires solving the problem of the firm for 20 different types, leading to 20 separate optimization problems. Each firm problem is itself the numerical solution to a system of equations whose computation include solving a number of numerical integrations. All in all, each evaluation of $h(x)$ may easily take a few minutes on a relatively fast computer.[25] The whole estimation procedure can take days, even when using parallel computing in

[24]The specific range for the transformed parameters, along with the corresponding values for the actual parameter values, is as follows:

Parameters	Range			
	Transformed		Not transformed	
A (productivity)	2	2.75	7.39	15.6426
B (technology bias)	0.3	0.9	0.5744	0.7109
α (dec. returns)	-0.5	0.5	0.3775	0.6225
γ (CES param.)	-1.2	-0.5	0.2315	0.3775
C (informality cost)	-2.75	-2	0.0639	0.1353
ξ (search cost)	-0.5	0.5	0.6065	1.6487
T (firm dist. shape)	-2.2	-1.4	0.0998	0.1978

[25]For these tests, we used a computer with an Intel(R) Core(TM) i7-4710HQ processor, large enough RAM, not running in parallel, and using standard optimization procedures in Matlab(R).

the outer optimization problem.

5) Obtaining the Variance of the Estimates

Once $\hat{x}$ is available, one can use the numerical implementation of function $h(x)$ to obtain the numerical Jacobian matrix $H(\hat{x})$. Then, it is a simple matter to compute the covariance matrix of the estimated parameters using equation 20. The finite differences must be taken regarding the original parameter, not the transformed variables from the optimization procedure (alternatively, one can estimate the covariance matrix for the transformed variables and obtain the desired covariance matrix using the delta method).

A Note on Tolerance Levels

Given the layered structure of the procedure, it is important to have a hierarchy of tolerances in the nested optimization problems. The "inner" procedures must use stricter tolerance values than the "outer" ones. Otherwise, numerical approximation errors in the former will lead to systematic errors in the latter. In a similar note, for procedures that use finite differences to calculate numerical derivatives, the finite difference must be of a substantially higher order than the tolerance of "inner" procedures.

In our application, the Classical Minimum Distance minimization problem has a tolerance of 1×10^{-5}, the equilibrium finding procedure has a tolerance of 1×10^{-8}, and the problem of the firm has a tolerance of 1×10^{-13}. The minimum size of the finite difference is set to be equal to the tolerance level of the problem.

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

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

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

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

The micro-region regressions include as demographic controls the shares of the sample in two age categories (30-39 and 40-50) and the log of population (all calculated based on the sample used in the individual-level regressions explained above). To allow for differential trends across local labor markets with different initial conditions, we also control for an interaction of the initial (1991) formality rate (from the individual-level regressions that do not control for schooling) with year dummies. In some specifications, we also control for the shares of employment in 8 broadly defined sectors (agribusiness and extractive industries, excluding mining; mining; manufacture; construction; utilities; retail; services; and government), and for interactions of year fixed-effects with a set of initial (1991) socioeconomic characteristics (schooling, which is the independent variable of interest; average earnings; and employment, which is one of the dependent variables considered). All regressions include micro-region and year fixed-effects and are weighted by the inverse of the standard error of the dependent variable (obtained from the individual level regressions). Standard errors are clustered at the micro-region level.

The results from these regressions are presented in Table D.1. In the table, each coefficient corresponds to a different regression, with the rows indicating different specifications and dependent variables (not controlling and controlling for education in the first stage, and using informality and employment as dependent variables). The columns correspond to different sets of controls, as indicated at the bottom of the table. All coefficients in the table refer to the same independent variable: the fraction of the population in the micro-region with at least 8 years of schooling. We present results not controlling and controlling for education in our "first-stage" to distinguish the individual level association between schooling and formality from the equilibrium effect of the composition of the population on the incidence of formality, conditional on individual schooling.

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

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

Table D.1 – Effect of share of population with at least eight years of schooling on formality and employment.

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

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

equilibrium effects, while less than 40% would be due to the direct relationship between schooling and informality at the individual level.

The magnitude of the estimated effects are between 30% and 40% of the quantitative effects from the comparative statics exercise in column 6 of Table 8. This difference may be due to the lack of a truly exogenous source of identification in our empirical results from this section, to limitations in our definition of skilled workers in this empirical setting, or to different sample (most importantly, the exercise here uses all micro-regions in Brazil, while the calibration was conducted using data from the PME, which includes only the 6 main metropolitan areas in the country).

The results related to employment, shown in the 3^{rd} and 4^{th} rows, are much less robust. Some specifications point to a positive and statistically significant relationship between educational composition and employment, but most results are small in magnitude and not statistically significant. Overall, we do not find a systematic relationship between educational composition of the population and employment rates. One potential explanation is that the utility from unemployment may be different across skilled and unskilled workers – possibly higher for skilled workers, due to higher wealth and savings –, something not considered in the model. This might weaken the correlation between educational composition and employment in the data. In addition, the problems alluded to in the previous paragraph could also be interfering with these employment results.