ESSAYS ON LABOR INFORMALITY IN DEVELOPING COUNTRIES

Andrea Otero-Cortes

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2018

Approved by:

Klara Peter

Charles Becker

Luca Flabbi

Ju Hyun Kim

Helen Tauchen

ABSTRACT

ANDREA OTERO-CORTES: ESSAYS ON LABOR INFORMALITY IN DEVELOPING COUNTRIES.
(Under the direction of Klara Peter)

My dissertation empirically studies the heterogeneous effects of informality on different labor outcomes and the role of labor regulation in affecting the informality rate in different countries.

The first chapter estimates the heterogeneous returns of informality in Russia using a Marginal Treatment Effect model (MTE) and regional variation in the degree of enforcement of the current labor regulation. The results indicate that self-selection accounts for a relatively large fraction of the difference in wages between formal and informal sector with the wage gap falling from 6.4% to 2.5%, ceteris paribus, and the earnings gap become negative. The findings support a "comparative advantage" hypothesis, as workers are self-selecting into the sector that better rewards their skills. There is also evidence of significant heterogeneity in the size of the formal-informal wage gap depending on individual unobserved cost of being formal. The study also finds significant sector differences in non-pecuniary labor market outcomes, as formal workers have a higher likelihood of receiving benefits such as supplemental medical insurance and paid vacation, and they report to be more satisfied with their jobs, but at the same time formal workers are more concerned about job loss, not finding a job if they get laid off, and they are less satisfied with their pay.

The second chapter estimates the marginal treatment effect of informality on wage rates in Brazil and compares those results to the Russian case in order to study if informality behaves differently in countries with different economic and social characteristics. The primary data source is the Pesquisa Nacional por Amostra de Domicilios (PNAD) for 2015, which is a household survey with information about workforce indicators, marital status and socio-economic characteristics. We use a combination of regional data on economic and geographic indicators at the state level and institutional data on labor inspections for identification. The results indicate that informality in Brazil responds to comparative advantage. Thus, workers self-select into the type of jobs that

better reward their skills. Formal workers do not have, on average, higher wages than informal workers, ceteris paribus, but there is large and significant heterogeneity in the returns to formality. Therefore, for workers with very low costs of being formal, formality offers significantly higher wage premiums.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my parents and my sister, Laura, who have been my biggest cheerleaders throughout this journey and my safe place whenever things seemed difficult (which was quite often!). I would not have been able to finish my doctoral studies without your constant support, encouragement, and unconditional love.

I am forever grateful to my advisor, Klara Peter, for her help and support since the first day I started working on this project, and for all the numerous hours she has spent guiding me, teaching me how to be more efficient at coding, and proof-reading everything I wrote during the past years. I also want to say thanks to her son for proof-reading this dissertation as well, because I am sure teenagers have better plans to do than to read papers about labor economics. I am incredibly grateful to my committee members: Luca Flabbi, Ju Hyun Kim, Helen Tauchen, and Charles Becker for their valuable feedback and assistance. Their insightful remarks and suggestions have significantly improved this dissertation. I would also like to extend my gratitude to the participants in the UNC-Chapel Hill Applied Microeconomics Workshop for helpful comments and advice.

I would also like to thank my friends. The old ones, for keeping up with me despite my lack of time and not being present in so many important moments of your lives ?because of school". Thanks for being there for me and for making me feel that things are still the same now that I am back in your life. The new ones, the ones that became family during the past 5 years, I will always be indebted to you for your kindness, love, and support. You guys made my days in Chapel Hill brighter, more fulfilling, and definitely fun.

This dissertation would not have been possible without the financial support of The Central Bank of Colombia and The Institute for the Study of the Americas at the University of North Carolina at Chapel Hill. The results and conclusions in this dissertation are my own and do not indicate concurrence by the Central Bank of Colombia and its Board of Directors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

HETEROGENOUS EFFECTS OF INFORMALITY: AN APPLICATION TO LABOR REGULATION POLICY IN RUSSIA

1.1 Introduction

Informality is a prevalent phenomenon in developing countries and transition economies (Perry et al. 2007; Meghir et al. 2015). In Latin countries, informal employment accounts for up to 40% of the labor force (Maloney, 2004; Meghir et al. 2015, DANE, 2016). For transition economies in Eastern European nations, informality represents about 20% to 30% of the labor force (Gimpelson and Kapeliushnikov, 2014). In particular, for Russia, as it will be shown below, 22% of all jobs in the economy are informal between 2009 and 2014. But despite its "popularity" in some regions of the world[1], informality is still an under-studied topic, especially in transition economies, due to problems associated to its illegal nature like the lack of data, and the complexity of the mechanisms that can cause it such as minimum wage policies, tax system, labor regulation, among others (Lehman, 2014).

Therefore, Russia provides an interesting setting to study informality as its economy has very different characteristics from traditionally large informal economies, such as in Latin America. Russia is the largest transition economy in the world since the dissolution of the USSR in 1991. Within a short period of time, a large informal sector emerged with companies avoiding paying taxes and social security contributions and rising unregistered self-employment of individuals. In Russia, a new type of informality arose in the form of "envelope payments", which are the part of the wages received in cash. The advantage of this arrangement is that there is no official registration of the transaction, and no taxes are paid on the wages. This type of informality is unique to Russia and is widespread even in manufacturing.

[1]The informality rate was, on average, 55% for Latin America in 2016 (OECD iLibrary, 2016). For Sub-Saharan African countries, the informality rate for 2014 was above 60 per cent (ILO, 2015)

Informal jobs are commonly thought to be low paying bad jobs, where low-skilled workers end up working. It is true that job informality narrows the tax base as informal workers do not declare their real earnings and informal workers could be at higher risk of poverty as they lack social protection and tend to earn lower wages and more unstable earnings (Bobba et al, 2018; La Porta and Schleifer, 2014; Slonimczyk and Gimpelson, 2015; Levy, 2008).

Then, how can we explain that some workers voluntarily choose to be informal if it is apparently a bad decision? From the point of view of a worker, informality may be costly as they would have to partially pay for the benefits associated with formal jobs, so it is not always the case that the benefits of informality are higher than the costs associated with it (Maloney, 2004). Thus, we need to study which factors are driving some individuals into informal jobs as a means to better understand the determinants of informality and its consequences. So we need to have better tools to study this problem from a policy perspective.

This lead us to the research questions of this paper: What are the returns to informality? Should the returns be measured solely on the basis of wages or do workers consider other variables when making decisions about the type of job contract they want? And, are these returns heterogeneous? By answering these questions, we will be able to get a better picture of the informal labor market in Russia and understand if informality is a choice or a curse imposed by a segmented labor market.

Estimating the returns to informality can be troublesome. First of all, there is no consensus about how to measure informality, and most of the definitions used in the literature depend on the availability of data. In this regard, the International Labor Organization, ILO, has tried to come up with a definition broad enough to cover all the different shades of informality. However, applying this broad definition has proven to be difficult empirically as in most cases there is no data with the level of detail required in order to apply that definition (i.e., workers earnings, the type of labor contract they signed, the job benefits they received, and workers preferences about different types of labor contracts).

The measures of informality most commonly used are those based on access to contributory social security systems, based on firm size, and based on legalistic measures (such as workers who are not registered in the labor office or do not have a worker's card)[2]. The problem of such measures is that they often offer an incomplete picture of informality.

For example, when informality is measured based on the access to social security benefits, it is often ambiguous, as there are many different benefits[3], workers may have access to some of them and not others, and/or the researcher may only have data on a few of them, as it happens in our case. So how do we determine who is informal? What if they have access to 3 out of 5 benefits? Thus, this measure does not offer a clear determination of who is informal and who is not.

In regard to informality based on firm size, it does not offer a precise measure, as not all small firms hire informal workers and not all large firms hire all their workers formally. For example, in Russia, 70% of the informal workers work in large firms, while only 30% work in small firms with 5 or less employees. When it comes to legalistic measures, they are less biased than the previous two, but these measures fail to capture subtle forms of informality such as the envelope payments.

Therefore, we use two measures in this paper. The first one is a legalistic measure according to which formal workers are those who are officially registered at the firm[4]. The second measure is an innovation over what has been previously done in the field, as it combines the legalistic aspect of informality explained before with a tax measure that requires declaring labor income. Thus, the second measure considers as "formal" those workers who are officially registered and taxes are paid on the entire amount of their salary.

The second issue with estimating the returns to informality is that the returns are not only limited to wage outcomes. The decision to be formal might be affected by other non-pecuniary labor market outcomes such as having supplementary medical insurance, fringe benefits, the likelihood of losing one's job, and others. Therefore, focusing solely on wages is not accurate, as individuals often make work decisions based on both wage and non-wage aspects of a potential job. On this

---

[2]For the first measure, see Arias and Khamis,(2008); Pratap and Quintin, (2006); and Bobba, Flabbi, and Levy,(2018). For the second measure, see Tannuri-Pianto, Pianto, and Arias, (2004). For informality based on a legalistic measures see Almeida and Carneiro, (2012); Meghir et al., (2015)

[3]i.e. access to healthcare, pensions, paid vacation time, paid maternity leave, paid sick days, among other

[4]This measure is based on the law which determines that every single worker must have an annotation in their labor book with a reference to the labor contract they currently hold. In Russia, a labor book is a document which records individuals' employment history.

subject, we examine the formal-informal gap in different non-pecuniary job-related outcomes and find that this gap is substantial.

The third issue with estimating the returns to informality is that informality status is not randomly determined. It cannot be a covariate in the wage equation without solving the selection issue first. So, there is a need for utilizing econometric methods that account for selection. In addition to that, the returns to informality may be heterogenous on both observable and unobservable characteristics of the individual, which means that informality does not pay off the same for all workers. Computing only average effects provides an incomplete picture of the phenomenon and neglects the distribution of the treatment effect. To address this issue, we use an augmented Roy Model framework and estimate the distributional effect of informality on several outcomes.

The fourth issue is finding proper exclusion restrictions for identification. The identification strategy relies on the use of shifters of the costs of being formal, such as the share of individuals in the community who work for the government as these jobs tend to be formal and the degree of enforcement of the labor code in the federal district (equivalent to state) measured through the ratio of inspectors per 1,000 economic entities and an interaction term between the ratio of inspectors and the distance to the nearest labor inspection office in the district.

In summary, this paper contributes to the literature in four aspects. First, it proposes a more comprehensive measure of informality, which is based on a legalistic definition plus tax compliance. Second, it focuses on pecuniary outcomes such as wage rate and monthly earnings, but also on non-pecuniary outcomes such as job satisfaction, health insurance, paid vacation time, and others. Third, it recovers the marginal treatment effect of informality for a transition economy finding significant heterogenous returns, which has not been done before. Finally, it uses a unique regional database on labor enforcement, which includes the distance to the labor inspection offices and the number of labor inspectors. So we digitized seven years of reports from the Federal Inspection on Labor.

The paper finds that the formal-informal wage gap on wage rate ranges from -60% for those with very high unobserved cost of being formal to 70% for those with an unobserved low cost. The average treatment effect of formality on wage rate is 2.5%, but it is negative for monthly earnings. There is evidence of comparative advantage, so individuals self-select into the informal sector based on higher expected gains. Additionally, formal workers have a higher likelihood of receiving supplemental benefits and report to be more satisfied with their jobs, but at the same time formal workers are more concerned about job loss, not finding a job if they get laid off and are less satisfied with their pay. This suggests that workers take into consideration other job characteristics besides payment when deciding in which sector they want to work.

The remainder of this paper is organized as follows. Section 2 shows the relevant literature review about informality, comparative advantage/segmentation testing, and labor regulation. Section 3 explains the institutional setting of the labor market in Russia and how it is enforced the labor code. Section 4 describes the data used in this paper, how variables were constructed, and analyzes summary statistics. Section 5 explains the empirical framework and discusses the identification strategy. Section 6 presents the results. Section 7 includes robustness checks and a discussion of the results, and section 8 concludes.

1.2 Literature Review

This section introduces the literature review relevant to this study. The review is divided into two groups: papers on comparative advantage versus segmentation testing and papers on the role of labor market institutions and regulation on informality.

1.2.1 Segmentation and Comparative Advantage Hypotheses

There are two traditional views that explain the appearance of informality in labor markets that are in equilibrium: the segmentation or a dualist labor market hypothesis, and the competitive view or comparative advantage hypothesis.

Perry et al. (2007) clearly explain the two different views that coexist about informality. On one side, the segmented labor markets hypothesis or exclusion view, which is the most popular one, claims that there are labor market entry barriers and rigidities, such as minimum wages or tax laws, that restrict the access to the formal labor market to those individuals with lower productivity, as firms cannot afford paying to them what the law requires. Under this view, informality is involuntary as workers cannot find jobs in the formal sector that offer state-mandated benefits.

On the other side, there is the choice view, which is aligned with Hirschman's (1970) idea about "firms and workers choosing their optimal level of consumption of social security and other state mandates depending on their valuation of the net benefits associated with the formality and the capability of the state to enforce the law". Thus, under this scenario, which takes into account comparative advantage considerations, informal jobs reflect workers' implicit choices given their preferences (e.g. desire for job flexibility), skills (entrepreneurial skills or ability to do networking), the cost and benefits of formality, and the availability of other means of social protection (Perry et al., 2007).

Therefore, Perry et al. (2007) conclude that both views rather than being mutually exclusive are complements, especially due to the fact that the informal sector is distinctly heterogenous as was stated before. Additionally, individuals may interact with the state in some dimensions and not others (i.e. a person may want to use the public health system but does not want to contribute to the state's pension fund), an informal worker may be excluded from the point of view of the access to the public health system but this person also chose not to contribute to pension funds. So, in this case, informality responds to both views. In the authors' words: "there is a continuum in the relative importance of exclusion and exit among individual workers and firms within countries".

Since 1950s, with the introduction of the Lewis Model of development, in which there are two sectors in the labor market in developing economies: one developed sector that owns the capital and another one of subsistence that has to rent capital and offers labor, the question about whether those "other jobs" are a consequence of rationing out of "good" labor market or a choice has been of interest of theorist and empirical economists (Lewis, 1954; Piore, 1979; Dickens and Lang, 1985).

Magnac (1991) presents an empirical approach for testing for the "segmentation hypothesis" in the labor market applied to the case of married women in Colombia. The author shows that using OLS is not correct when testing for segmentation based on differences in the wage functions in each sector because of the self-selection into sectors problem. For that reason, he uses an extended Roy model to allow for four sectors: formal sector, informal sector, unemployed, and non-participation, which could be represented by a multivariate generalized Tobit model with three dependent variables: wages, sector choice, and participation. The model was estimated using a maximum likelihood methods (MLE).

Such model assumes that selection happens only based on observable characteristics and does not allow for correlation between equations. The informal sector only includes self-employed workers and the formal sector includes all employees/blue-collar workers. Wages are computed as monthly income divided by number of hours worked and the potential wage function includes education, experience and experience squared, husband's income, number of children 0 to 1 years old, number of inactive young and adult women in the city, and other family income. The paper concludes that there is evidence of a competitive labor market structure instead of a segmented one.

Pratap and Quintin (2006) test for segmentation of the labor market in Argentina by estimating a wage function corrected by selection using a propensity score. Selection model is estimated based on observable characteristics and they use a dummy variable that takes the value of 1 if the person had other family member employed in the formal sector as an exclusion restriction.

Under their definition of informality, any worker who receives both pension and unemployment insurance is considered formal, everyone else is considered informal. Wages are computed as monthly income divided by 4.33 times weekly hours worked. The authors find that the formal-informal premium gap remain after controlling for individual and firm observable characteristics under a parametric model, but when they estimate the model semi-parametrically using propensity score methods in order to solve the selection issue, the premium becomes small and insignificant or negative.Thus, the authors conclude that the view that labor markets are segmented in developing nations such as Argentina is not correct. Our results are similar to this findings, but the estimation strategy and identification is different as we use a more robust model that allows us to account for individual heterogeneity and add exclusion restrictions that capture enforcement as a cost shifter for individuals when deciding which sector to join.

Arias and Khamis (2008) test both the comparative advantage hypothesis and segmentation in labor markets in Argentina using marginal treatment effects.The authors divide their sample into three comparison groups: dependent salaried work (formal) versus self-employment, dependent salaried work (informal) versus self-employment, and formal versus informal salaried work. Their outcome of interest is labor income per hour in the main occupation and their outcome equation follows a standard Mincer specification. The identification strategy for the first two comparison groups relies on the inclusion of the variable "workers intrinsic preference for working in a dependent relationship" in the selection equation and for the last group the exclusion restrictions were "the number of inspected workers at the province of residence" and "having the spouse of other relatives employed in the formal salaried sector".

The authors did not find any significant differences between the earnings of formal salaried workers and the self-employed individuals once they account for selection, which is consistent with the comparative advantage hypothesis. But when comparing formal and informal salaried workers, they do find that informal salaried work carries significant earnings penalties. There is a negative selection bias into formal salaried work relative to informal salaried work and only a slightly positive sorting based on expected earnings gains. So those results are more consistent with labor market segmentation.

1.2.2 Labor Market Institutions and Regulation

There is evidence that informality may arise as a result of different institutions that provide disincentives to formal work. Those institutions, such as labor taxation, employment protection legislation, social security benefits and unemployment insurance, and weak enforcement of the labor code, make individuals more likely to operate informally in the formal sector by accepting an informal job or operate directly in the informal sector (Pagés et al., 2014).

From the point of view of a worker, being formally hired implies that they will have to start contributing to social security even though they may not value those benefits enough to want to pay for them. Additionally, workers may have to pay higher personal income taxes due to a higher gross formal wage and for some workers, it would also imply losing some social assistance benefits that are income-tested. On the employer side, if a firm hires a formal employee, this means they would not only have to contribute to pension funds, health funds, and unemployment insurance funds, but they would also have to generate enough formal revenues to cover such expenses (Koettl and Weber, 2014).

Lehman and Muravyev (2014) do a cross-country study using a novel panel data that includes countries from transition economies in Eastern and Central Europe and Latin American countries, to do a cross-country analysis about the impact of a set of determinants of informality. Their research question is whether employment protection legislation, the tax wedge, the unemployment benefit level and duration and union density affect the size of the informal economy in both regions. In words of the authors, there should be strong policies that aim to increase the size of the formal economic activity and employment, as informality promotes inefficiency (in the use of the limited resources for production in a country) and inequality (as part of the workforce and firms do not pay their share of taxes, so the rest of the contributing society assume their share of the tax burden). Their results suggest that the more regulated a labor market is (more stringent employment protection legislation) and the tax-wedge, the higher the size of the informal sector. Although this paper talks about how tighter labor regulation can have negative impact on informality, it does not account for the role of enforcement that can counterbalance such effect.

Almeida and Carneiro (2012) study the impact that labor inspections in Brazil have on the size of informality at the municipality level. They show that although enforcement of labor regulations in the formal sector is thought to drive workers to informality because they increase the costs of formal labor, it is also true that labor inspections may enforce compliance with mandated benefits which are highly valued by workers, and potentially increase the attractiveness of the formal sector. Summarizing their results, they find that in locations with frequent inspections, workers pay for mandated benefits by receiving lower wages, but minimum wage policies prevents downward adjustment at the bottom of the wage distribution. Then, formal jobs that pay the low wages around the minimum wage become attractive to some informal workers, inducing them to want to move to formality.

Moving to general equilibrium models, Meghir et al. (2015) show, using an equilibrium wage-posting model with heterogeneous firms for the Brazilian case, that there is evidence of compensating differentials in the wage schemes offered by firms to both formal and informal workers. They also find that tightening enforcement does not increase unemployment and increases wages, total output and welfare by enabling better allocation of workers to higher productivity jobs and improving competition in the formal labor market, and low skill workers are found in both formal and informal jobs, which is evidence against segmentation. Their definition of informality is similar to ours as they consider formal those employees who have a signed worker's card and informal those who do not have it or are self-employed. Even though their paper is about firms, their results are a good motivation for this paper as enforcement is acting through increasing competition in the labor market but not through increasing unemployment.

Bobba et al. (2018) develop a search-matching bargaining model for Mexico with endogenous schooling in which potential employers decide the formality status of the job offered and bargain with workers over wages. The authors find that formal employees have on average higher wages than informal employees, but the two wage distributions overlap over a large portion of their support, and additionally, they find that social policies have different effects on informality depending on wether the individual is self-employed or an informal employee.Their definition of informality covers self-employed workers and employees who are not registered by the firms they work for into the social security system. Identification comes from schooling specific vacancy rate, economy-wide labor shares,and the staggered implementation of a non-contributory health insurance called *Seguro Popular*.

1.3 Labor Market Regulation and Enforcement in Russia

Russia's labor regulation seems, at least on paper, to be very stringent when it comes to protect workers' rights, but the enforcement of the regulation is not as strong as the law itself (Gimpelson and Kapeliushnikov, 2014). This section introduces the main characteristics of the labor regulation in Russia, the most common law breaks and how it is enforced the labor code.

1.3.1 Labor Regulation

The Federal Service for Labor and Employment, Rostrud, is the federal body in charge of the supervision and control of labor conditions, employment, social protection of the population, among other benefits and rights established by the legislation of the Russian Federation to workers and embedded in the Russian Labor Code, which was put in place in 2002. In 2013, the Article 25 of the Federal Law No. 426-FZ, "On Special Assessment of Working Conditions", indicated that Rostrud and its territorial bodies (state labor inspectorates) exercise state control compliance with the mandatory requirements.

By law, all employment contracts should be written and all workers need to have a labor book, which keeps a record of the work history of the worker. If the contract is not written and the employee is admitted to work, the employer needs to provide the employee with a written contract within 3 days since the employee started working; if not the employer could be fined.

Additionally, employers have to comply with other rules such as to provide at least 28 calendar days of paid vacation, pay for maternity leave up to 140 days, provide a safe working environment for its employees, comply with minimum salary requirements, contribute to social security for the employee, comply with the maximum number of overtime hours per year (120 hours). Failure to comply with the labor regulation can carry administrative fines of up to RUB200,000 (equivalent to US$3,500) and disqualification or suspension of activities for up to 90 days, depending on the severity of the offense (ILO, Labor Code of the Russian Federation of 31 December 2001).

With regards to resignations and dismissals, if an employee wants to quit her job, she can do so by notifying her employer two weeks in advance; no reasons need to be stated. On the other side, the Labor Code only includes 10 legal reasons for firing an employee and the employer needs to clearly document the misbehavior of the employee with relation to those 10 reasons and also prove that the employee was notified and warned about his poor performance or misconduct before issuing an internal termination order, which has to be done at least 2 weeks in advance before terminating the labor contract. In addition to that, severance payments equivalent to two months of work need to be paid to the dismissed employee.

1.3.2 Violations of the Labor Code

The three most common ways of violating the Labor Code when it comes to informal employment are the following. First, firms can hire an employee without providing her with a work contract or agreement. This way the person works for the firm but she does not have an entry in her labor book about this job and does not belong to the work roster of the company. The firm does not have to pay to her any social insurance or give her any benefits such as maternity leave or vacation time.

The second one is the so-called envelope payments, which consists of an employee who apparently is officially registered in a firm, but receives part of her wage unofficially in an envelope, so there is no record about that money being paid to the worker as part of her salary. In this case the employee does have a written contract and may have all the benefits of a formal job, but her labor contract states that her salary is equal to the minimum wage although she is paid a higher wage off the books. This form of informality benefits the employer, as the firm only needs to pay social security contributions on the officially declared salary and in case of having to fire an employee, then the severance payment is lower. The employee benefits by not having to pay income taxes on the envelope payments as there is no record of those payments.

Third way includes all the other type of workers who are not in a firm and do not comply with the law. For example, individuals working for a private person under any job agreement, self-employed individuals who did not registered their entrepreneur activities, and people who claim to not be working but engage in occasional economic activity and do not report those earnings as labor income.

Additionally, there could be violations in terms of the rights of formally hired employees such as not providing them with a safe work environment, not giving them vacation time as established by the law, making them work longer hours, among other. Thus we will try to capture those violations as an outcome of interest rather than as a measure of informality as we consider them to be a consequence of an informal labor agreement.

1.3.3 Enforcement and Labor Inspections

The Rostrud has Federal Labor Inspectors (FLI) in 82 of the 85 federal subjects in which Russia is divided. As Gimpelson et al. (2009) describe, the main objective of labor inspectors is to monitor and enforce the labor code in all its aspects, including hiring, firing, wage payments, and safety of the workers. Labor inspectors can conduct inspections on any firm or on any employer at any time, with or without notice. All firms are obliged to comply with the requests of the labor inspectors and provide them with any documentation or information that the inspector requires to conduct her investigation. After a violation is found, the inspector notifies the employer, requests to rectify the situation, and may impose fines on the employer or executives of the firm. If the employer does not cooperate with the requests of the inspector or does not comply with the penalty imposed, then the case is taken to the local court or to the prosecutor's office.

Following Carneiro and Almeida (2012), the main empirical challenge of identifying the impact of enforcement on formality is that enforcement is not randomly distributed across cities. This means that we could expect stronger enforcement in areas with a higher number of reports of violations or in areas with better institutions. Thus, in order to avoid potential endogeneity issues, the authors suggest studying the technology of enforcement and controling for it. As we already know how enforcement works and we find the technology to be similar to the one in Brazil, we will use Carneiro and Almeida's (2012) strategy in order to identify our model.

The technology of inspections in Russia is as it follows: The FLI may decide to send inspectors to check a firm either randomly or based on workers' complaints or after a prosecutor's office call. Inspectors have to travel from their base office to the firm in which there is a suspected violation. If several complaints are received from workers in the same industry, then inspectors may audit all firms in that industry. Also, there are some periodical inspections, which are done every two years.

Therefore, as we can see, inspections depend on two main inputs: number of inspectors in each FLI and distance from the FLI to each particular city.

There is a trade-off between distance between a city and the nearest inspection office and the likelihood of having an inspection. Thus, the farther a city is located from the nearest inspection office, the lower the probability of being inspected. But the distance constrain becomes less important when there are more inspectors at the FLI, as the cost of sending an inspector to a place far away is reduced if there are other inspectors in the office who can keep doing inspections while the other inspector is traveling.

Thus, we need to control for such relationship between distance and likelihood of being inspected as it is crucial in determining the amount of inspections that can be done in a set period of time. This is captured in our work by including an interaction term between a measure of the distance to the nearest inspection office and the number of inspectors per 1,000 economic entities in the area, as it will be explained in Section 5.

1.4 Data Description

This study draws on the Russian Longitudinal Monitoring Survey-Higher School of Economics (RLMS), which is a household panel survey that collects data on labor and health outcomes. The survey is representative for the country. Additionally, the RLMS has a "settlement module" that allows to recover information such as the population size of the settlement and distance from the settlement to the city in which the labor inspection office is located. For this study, only data from 2009 to 2016 is used, as in this period it is possible to obtain official information about the number of labor inspectors, which is used to construct the different measures of enforcement used in this paper.

The RMLS has 130,167 observations-year in total from 2009 to 2016. The sample used is restricted to individuals who are 20 to 59 years old, who worked at least 5 hours in the reference month and who are currently working. The final estimation sample varies in size depending on the definition of informality and the estimation method used, but it is in between 50,000 and 62,000 observations.

Through an extensive work of digitizing seven years worth of data contained in the "Reports on Implementation and Effectiveness in Federal State Oversight of Compliance with Labor Legislation and Other Regulatory Legal Acts", that are released by the Russian government every year, we were able to link regional data on labor regulation with individual level data form RMLS. Additionally, we compiled data on other economic indicators at the regional level such as the unemployment rate and consumer price index. Detailed explanation of the data and its sources can be found in the data appendix.

1.4.1 Definition of informality

All employed individuals in RLMS can be classified into two groups: formal workers and informal workers. Thus in this paper, we will use two definitions of informality that are based on the legalistic aspect of it.

Under definition 1, formal workers are those who are officially registered as employees at a firm. Thus formal workers must answer "yes" to the questions "Does this job belong to an enterprise or organization?" and "Are you on a work roster, written work agreement, or work contract?". To be classified as informal, a worker must (i) be working at a firm but not officially registered or (ii) not be working at a firm but possibly working for a private individual or self-employed. Figure 1.1 shows the classification tree based on the survey.

Under definition 2, formal workers are those who are officially registered as employees at a firm and their entire salaries are paid officially (i.e. taxes were payed on the entire salaries) and those workers who do not work at a firm but paid taxes on their labor incomes. Informal workers are (i) those who work at a firm but are not officially registered regardless of whether they pay taxes or not, and (ii) those do not who work at a firm and do not pay taxes on their labor incomes.

For the period of interest, 2009 to 2016, observations come from 32 of the 85 federal subjects of Russia. On average, there are 7,500 unique observations per year. The informality rate in the sample, using definition 1, is 16%, which is slightly lower than the informality rate observed in the country for the same period of time. With definition 2, the sample informality rate of 20.9% is similar to the officially published informality rate of 22% (Table 1.1). The number of observations is different across definitions because fewer individuals responded the question about paying taxes on their labor incomes.

For 2016, for example, under Definition 1, of the 1,190 informal workers, 39.2% are informal employees at a firm. Under Definition 2, of the 1,482 informal workers, 72.1% of them are informal employees at a firm. This distribution is fairly constant across the years included in the sample.

1.4.2 Outcome variables

The main two outcome variables are the log of hourly wage rate and the log of monthly earnings. Both outcomes are from the main job, after-tax and converted to real terms by using the CPI base December 2016 (Table 1.2). A detailed description of how we constructed the variables may be found in Appendix A.1.

Additionally, we evaluate the impact of informality on other outcomes. These outcomes reflect the access the workers have to some non-pecuniary benefits and workers' general well-being, such as the firm's paying for supplementary medical insurance, eligibility for unemployment benefits if the worker loses her job, paid vacation time, satisfaction with the job contract and pay, job stability, and confidence about finding a new job if laid off (Table 1.3).

Using the sample created by definition 2, on average, 6.3% of the workers have supplemental insurance paid by their employers, 12.6% would receive unemployment benefits if they lost their jobs, and 71.5% have had paid vacations in the last 12 months. When it comes to satisfaction in the workplace, 66.5% report being satisfied with their jobs overall, 64.4% report being satisfied with their work contracts, and 35.8% report being satisfied with their pay. On a less positive note, 71.6% are concerned about losing their current jobs and only 42.1% are confident they would find a new job if laid off.

1.4.3 Labor regulation variables

The data for the labor force indicators are from the Federal Inspections on Labor (Rostrud), which provided information about the number of economic entities per federal subject and the number of labor inspectors for 2009, 2011 to 2016, and along with the number of economic entities per labor inspectors for 2010.

On average, there were 43 labor inspectors at each federal subject between 2009 and 2016. The number of inspectors has been declining over time, although the number of economic entities has remained fairly constant during the same period of time. No new labor inspection offices were created during the time period of interest (Table 1.4 and 1.5).

For example, the number of inspectors in Moscow city declined from 163 in 2009 to 99 in 2016 for an average of 118 over the time period. Similarly, in Saint Petersburg, the number of inspectors fell from 95 to 73 over the same time interval.

Klara Peter created the variables for the distance to the nearest labor inspection office. For cities with a labor inspection office, the distance is measured from the city border to the center of the city and for cities with no inspection offices, the distance is from the city's center to the city center where the nearest inspection office is located. Figure 1.2 shows the kernel distribution of the log distance to the nearest inspection office by formality status. This distribution is bimodal, with two peaks at around 20 km and 200 km, but both formal and informal workers seem to face similar distances to labor offices, on average.

1.4.4 Descriptive Statistics

Descriptive statistics are reported based on the measure of informality by definition 2. Results using definition 1 are similar. Table 1.6 shows that individuals with formal and informal jobs seem to have different demographic and socioeconomic characteristics that we need to control for. In the sample, women tend to be more formal than men as it is commonly found in the literature about informality in transition economies when using legalistic measures for informality (Lehman and Zaiceva, 2015). Also, people in informal jobs are younger, more likely to be single, and with less years of education and also with less educated parents. Although there is some variation within the informal group, as self-employed individuals tend to be more educated than the rest of the individuals in the same category.

City size does not seem to play an important role in determining the likelihood of being informal under any of the definitions used in this paper and the data even hints that larger intermediate cities are more suitable for informal jobs. On the other end, population size at the site where the respondent lives does seem to matter. Formal workers tend to live in areas with larger population, which are also larger and richer cities. Unemployment rate in the region and share of government employment n the community also seem to matter, although the differences are not striking.

In terms of earnings and working hours, real labor earnings per month are virtually identical for formal and informal workers, but the real wage rate is 8.5% higher for formal workers and they also report, on average, 18 fewer hours of work per month than informal workers. These differences may hint at differences in productivity as formal workers spend fewer hours at work but earn more per hour.

1.5 Econometric Framework

This section introduces the empirical methodology used in the paper. First, it discusses a basic potential outcomes model and then, it introduces a refinement of the model, which is the marginal treatment effects model used for the final estimation.

1.5.1 Basic Potential Outcomes Model

Let Y be the observed outcome of interest, the log real wage rate at main job or the log real monthly earnings, but for simplicity we will talk about the real wage rate but the model is also estimated using earnings. Assume that there are two types of occupations indexed by two labor market sectors: formal (treated state) and informal (untreated state). Let D represent the binary treatment of interest: being formal. Define Y_1 as the potential outcome of an individual in the treated state (D=1), and define Y_0 as the potential outcome of an individual in the untreated state (D=0), such that Y_1 represents the potential wage rate of an individual who works formally, and Y_1 represents the potential wage rate of someone who works informally.

The observed outcome is therefore:

$$Y = (1 - D)Y_0 + DY_1. \tag{1.1}$$

The outcomes for the formal and informal sector are assumed to be linear functions of socioeconomic and demographic characteristics (X) such as schooling, age, parents' education, and regional controls such as regional unemployment rate and the distance to the closest labor inspection office, and the error terms U_0 and U_1 are independent of X and the selection rule.

$$Y_1 = X'\beta_1 + U_1 \tag{1.2}$$

$$Y_0 = X'\beta_0 + U_0, \tag{1.3}$$

The rule for selection to treatment, in this case to have a formal job, can be separated into two components: an observable part, $Z\gamma$, which contains the same variables as X and some exclusion restrictions, and an unobservable component, V,

$$D = \begin{cases} 1 & if \quad Z\gamma + V > 0 \\ 0 & if \quad Z\gamma + V \leqslant 0. \end{cases} \tag{1.4}$$

1.5.2 Switching Regression with Selection on Observables: Regression Adjustment and Inverse Probability Weighting

In order to estimate the previous switching regression model, there are some assumptions that are required for the tractability of the model. If we believe that there is conditional independence between the treatment assignment and the potential outcomes, which means that there are no unobservable factors that could affect at the same time the assignment to formality and the potential outcomes of both formal and informal workers, then this will translate to assuming that $\sigma_{0V} = 0$ and $\sigma_{1V} = 0$ in the variance-covariance matrix.

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{10} & \sigma_{0V} = 0 \\ \sigma_{10} & \sigma_1^2 & \sigma_{1V} = 0 \\ \sigma_{0V} = 0 & \sigma_{1V} = 0 & 1. \end{pmatrix} \tag{1.5}$$

This assumption says that selection into formality is only based on observable characteristics. Thus, the model could be easily estimated by using regression adjustment methods (RA), inverse probability weighting (IPW), a doubly-robust estimator (IPWRA) or matching methods, as all of them rely on selection on observables. We also need to assume that the data is i.i.d. In this paper, we will explore the first three methods that will be briefly discussed.

Regression Adjustment (RA): The RA estimator uses sample means to estimate treatment effects by using a regression model to predict potential outcomes adjusted for covariates. This means that we can construct counterfactual or unobserved potential outcomes for both treated and untreated observations based on other covariates in the model and then estimate two separate regressions on the treated and untreated groups, including the counterfactuals, and then compare the means of each group (treated group that includes observed and counterfactual outcomes and untreated group which also includes observed and counterfactual outcomes).

Doubly Robust Method- Inverse probability weighting and regression adjustment model (IP-WRA): This method combines the outcome modelling strategy of RA previously discussed and uses an inverse probability weighting (IPW) strategy in order to account for treatment selection. The IPW method computes weights that are the reciprocal of the likelihood of participating in the treatment, which in this case is having a formal job. Then when estimating the effect of formality on the outcomes of interest, we use weighted means rather than simple unweighted means to disentangle the effects of treatment and other confounders. This estimator is doubly robust as only one of the two equations (outcome or treatment) need to be correctly specified in order to get a consistent estimator.

But what happens when there are unobserved factors, like preferences, unobserved skills, and omitted variables, that affect both selection into formality and outcomes such as wages and earnings? In those cases, in order to recover a consistent estimator for the effect of formality on the outcome of interest, we need to correct for the self-selection into treatment and allow for the presence of correlation between V, and the error terms, U_1 and U_0. Additionally, if we believe that the returns to formality and informality vary based on observable and unobservable characteristics of the individual, as it was stated in the introduction, then traditional selection methods will not suffice as it is important to capture the selection on gains in the empirical model by recovering not only mean effects, but the whole distribution of the effects.

1.5.3 Endogenous Switching Regression: Marginal Treatment Effects Model

The marginal treatment effects model does allow for a more flexible variance-covariance matrix as it does not impose any restrictions on the values that σ_{0V} and σ_{1V} can take. It also allows us to recover the distribution of the effect of formality for all the values of V.

Following Carneiro et al. (2011), this estimation method is based on the generalized Roy Model.The decision rule of an individual i to work formally or informally is characterized by a latent variable model in which D equals one for individuals who work formally and zero for individuals who work informally and V represents the unobserved marginal cost of being formal. (Willis and Rosen, 1979). Thus, the selection equation is as follows:

$$D = 1(D^* > 0), \tag{1.6}$$

and $D^* = Z\gamma - V$.

Under this framework, V could be interpreted as the benefit of having a more flexible job or being in a independent working relationship when the individual has strong entrepreneurial skills, among others.

Notice that (X, Z) is observed, but (U_0, U_1, V) is not. Therefore, we need assumptions on the unobserved variables in order to make the model tractable. We assume that V is a continuous random variable with a strictly increasing distribution function F_V and (U_0, U_1, V) is statistically independent of Z given X. The vector Z contains observable individual and family characteristics that affect the decision to work formally or informally along with excluded variables that affect the decision to be formal but do not directly affect earnings or wages. The inclusion of these variables in the selection equation but not the outcome equations is what allows us to get identification.

Therefore, the decision rule can be written as:

$$D = 1(Z'\gamma > V). \tag{1.7}$$

Let $P(Z)$ denote the probability of work formally (D=1) conditional on Z=z, such that $P(Z) = Pr(D = 1|Z = z) = F_V(\mu_D(Z))$. We keep conditioning on X, but to make notation easier it is omitted from now on. Now define $U_P = F_V(V)$, which is uniformly distributed by construction. This transformation is useful because different values of U_P correspond to different quintiles of V.

Rewriting Equation 1.7 using the transformation of the error term and P(Z), we get:

$$D = 1(P(Z) > U_P). \tag{1.8}$$

Now we can rewrite Equation 1.1 as:

$$\begin{aligned} Y = (1-D)Y_0 + DY_1 &= D(\mu_1(X) + U_1) + (1-D)(\mu_0(X) + U_0) \\ &= D(X'\beta_1 + U_1) + (1-D)(X'\beta_0 + U_0) \\ &= X'\beta_0 + D((X'\beta_1 - X'\beta_0) + D(U_1 - U_0) + U_0. \end{aligned} \tag{1.9}$$

Assuming that $\mu_1(X)$ and $\mu_0(X)$ also have a linear representation such that $\mu_j(X) = X\beta_j$.

The conditional expectation of Y given X=x and P(Z)=p is:

$$\begin{aligned} &E(Y|X = x, P(Z) = p) \\ &= E(Y_0|X = x, P(Z) = p) + E(Y_1 - Y_0|X = x, D = 1, P(Z) = p)p \\ &= X'\beta_0 + (X'\beta_1 - X'\beta_0)p + \int_0^p E[(U_1 - U_0)|X = x, U_s = u_s]du_s. \end{aligned} \tag{1.10}$$

When estimating (1.10), we need to consider three cases. As Belskaya, Peter, and Posso (2014) explain, the potential results could be: (i) if the unobserved terms are homogeneous, that is $U_0 = U_1 = \bar{U}$ for all individuals, then the last term of Equation 1.10 cancels out; (ii) the unobserved terms are heterogeneous but mean independent of high-school decisions, that is $E(U_1 - U_0|X = x, U_s = u_s) = E(U_1 - U_0)$, then the last term of Equation 1.10 cancels out; and (iii) if the unobserved terms are heterogeneous and correlated with V (the error term from the selection equation), then the last term of Equation 1.10 cannot be ignored, because it reflects "selection on gains".

In order for a classic instrumental variables approach to be valid we must assume that individuals sort randomly into formal and informal sectors. If instead, an individual considers the relative benefits of each sector to her, then we must use an estimation method that is consistent with the so-called "selection on gains".

Under the potential outcomes framework defined by equation 1.1-1.3 and the selection equation 1.8, the switching regression model assumes that the error terms of the three equations follow a multivariate normal distribution, $(U_0, U_1, V) \sim N(0, \Sigma)$, and the vector (U_0, U_1, V) is independent from (X, Z). The variance of V is normalized to 1, such that $\sigma_V^2 = 1$, and the covariance between U_0 and U_1 cannot be recovered given that we never observe both outcomes simultaneously. Therefore σ_{10} is not identified. The variance-covariance matrix in this case is:

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{10} & \sigma_{0V} \\ \sigma_{10} & \sigma_1^2 & \sigma_{1V} \\ \sigma_{0V} & \sigma_{1V} & 1. \end{pmatrix}$$

Following Lokshin and Sajai (2004), the model can be efficiently estimated by using the full-information Maximum Likelihood method to jointly estimate both the outcome equation and the decision rule. The loglikelihood function of the model in this case would be:

$$\begin{aligned} ln(L) = \sum_i (D\omega_i[ln(F(\eta_{1i})) + ln(\frac{f(\frac{U_1}{\sigma_1})}{\sigma_1})]+ \\ (1-D)\omega_i[ln(1-F(\eta_{0i})) + ln\left\{\frac{f(\frac{U_0}{\sigma_0})}{\sigma_0}\right\}]). \end{aligned} \tag{1.11}$$

Where:

F: Cumulative normal distribution

f: Normal density distribution

ω_i: Optional weighting for observation i

$\eta_{ji} = \frac{Z\gamma + \rho_j(\frac{U_j}{\sigma_j})}{\sqrt{1-\rho_j^2}}$ for j=0,1 and $\rho_j = \frac{\sigma_{jV}^2}{\sigma_V \sigma_j}$ are the correlation coefficients.

In order to estimate (1.11), we need a transformation of the correlation coefficients and standard deviations to guarantee that the correlation is between -1 and 1 and the standard deviation is always positive. This is done in a way that it is easy to recover the true parameters of the model. For the case of the standard deviations, $ln(\sigma_j)$ is used instead of using σ_j. For the correlations, the Fischer's transformation is the standard: $atanh(\rho_j) = \frac{1}{2}(\frac{(1+\rho_j)}{(1-\rho_j)})$.

The MTE methodology does not assume that the returns of formality are the same for everyone, therefore it accounts for selection on gains. Following Carneiro et al. (2011) , this model assumes that agents know the gross return on earnings of having each type of job. This means that individuals know $\Delta = Y_1 - Y_0 = (X'\beta_1 - X'\beta_0) + (U_1 - U_0)$ per each i.

In the third case analyzed before what is happening is that individuals who are identical on their set of X's may make different decisions about which type of employment to get, influenced by their unobserved component V in the selection equation. As a result of this feature, the returns of working formally or informally on wages, for observationally identical individuals, will depend upon a constant component $(X'\beta_1 - X'\beta_0)$ and an individual-specific component $E(U_1 - U_0|X = x, U_s = u_s)$.

If we differentiate Equation 1.10 with respect to p, we get the MTE:

$$MTE(x,p) = \frac{(\partial E(Y|X = x, P(Z) = p))}{\partial p} = (X'\beta_1 - X'\beta_0) + E(U_1 - U_0|X = x, U_s = u_s). \tag{1.12}$$

The last term of Equation (1.12) can be estimated in a parametric version and in a semi-parametric version, both versions can be estimated using polynomials of different orders or not. For this version of the paper, I will use a parametric approach.

Parametric Normal Model

Under the parametric framework, the model has the same set of assumptions as the Normal Switching Regression model explained before. Using the same multivariate normal parameterization, Equation (1.12) can be expressed as:

$$\begin{aligned} MTE(x,u_s) &= X'(\beta_1-\beta_0)+E(U_1-U_0|U_s=u_s) \\ &= X'(\beta_1-\beta_0)+E(U_1-U_0|V=\Phi^{-1}(U_S)) \\ &= X'(\beta_1-\beta_0)+(\sigma_{1V}-\sigma_{0V})\Phi^{-1}(U_S). \end{aligned} \quad (1.13)$$

The parameters $(\beta_1, \beta_0, \sigma_{1V}, \sigma_{0V})$ and their standard errors can be estimated by maximum likelihood methods. The most common ways of estimating this model under normality assumptions in order to recover the parameters of interest are: (i) Following Lokshin and Sajai (2004), who specified the loglikelihood function that we already showed or (ii) Following Maddala (1983), who proposes a linear regression model augmented by a binary endogenous treatment variable and assumes that $\beta_1 = \beta_0$ and $\sigma_0^2 = \sigma_1^2$. This paper follows Lokshin and Sajai (2004) approach given that it imposes less restrictions on the model.

1.5.4 Hypotheses to be tested

Following Magnac (1991), there are two hypotheses that can be tested under the switching regression framework. On one side, the segmented labor markets hypothesis claims that access to the formal labor market is restricted by minimum wages, tax laws, and other labor regulations, thus lower productivity workers are rationed out of the formal sector and can only find jobs in the informal sector. If true, we should observe that: $cov(U_1, V) > 0$ and $cov(U_0, V) > 0$. On the other side, the comparative advantage hypothesis says that informal jobs reflect workers' implicit choices given their preferences, skills, the cost and benefits of formality, and the availability of other means of social protection (Perry et al. 2007). If true, we should observe that: $cov(U_1, V) > 0$ and $cov(U_0, V) < 0$.

1.5.5 Identification

Theoretically, the parameters of interest in the econometric model, $(\beta_1, \beta_0, \sigma_{1V}, \sigma_{0V})$, could be identified from non-linearities in the selection equation, which in this case they exist. But if the econometrician would like to have more consistent estimates, there is a need for exclusion restrictions in the selection equation (1.8). This means that the selection equation should contain at least one variable that is not included in the outcome equation that affects the decision to be formal or informal, but does not affect wage rate directly besides its effect through the decision rule (being formal or not).

Our identification strategy is based on the use of variation in the degree of enforcement at the regional level (see Appendix A for statistics about the number of penalties imposed by the inspectors and the amount of money collected through fines) and the share of public employment at the community level as shifters of the decision to be formal or informal. But enforcement may not be randomly located across regions as we expect to have more inspectors in areas where there are more violations of the labor law or in richer cities with a larger budget[5], and both issues may be correlated with our outcomes of interest (wage rate and earnings). Thus, we need to control for the technology of enforcement that we described before, and account for the trade-off between the likelihood of being inspected and the distance to the nearest labor inspection office. This can be done by including the number of labor inspectors per 1,000 firms in the federal subject and an interaction term between distance and the ratio of inspectors given that distance is less of a constrain in areas with a higher number of labor inspectors (Almeida and Carneiro, 2012).

In order for the ratio of labor inspectors and the interaction of that variable with our distance measure to be valid as exclusion restrictions, they must be correlated with formality status but they cannot be correlated with the wage rate or monthly earnings of the individuals, conditional on the other covariates included in the model. We argue that our exclusion restrictions do not have a direct impact on the outcome of interest but only through formality status of the individuals, as: (i) we control for distance to the nearest labor inspection office, which we think is the variable that may be correlated with earnings and wages as informal jobs may be offered by firms located far from the inspection area or informal self-employed individuals may prefer less centric locations in order to be off the radar; (ii) we control for the main determinants of wages at the individual

[5]For example, the Central region, where Moscow and St. Petersburg are located, is the region with the highest number of fines imposed by inspectors and penalties

level such as education, education of the parents, sex, and age; and, (iii) we control for regional level characteristics that may impact labor markets such as population size, unemployment rate, and regional fixed effects that capture the wealth of the city, among others aspects[6].

The share of the individuals in the community who have a job in the public sector works as an exclusion restriction a la Carneiro, Heckman, and Vytlacil (2011) and others, as the availability of these jobs is correlated with formality status given that jobs in the public sector are mostly formal (in our sample, 43.6% of the individuals hold a job in the public sector, but only 2.7% of them work informally under definition 2), but it is not correlated with the outcomes of interest as we already control for the variables that may influence the labor market such as unemployment rate, population size and regional fixed effects[7]. Additionally, wages in the public sector are, on average, 10RUB lower than average wages. Thus, if the share of public employment in the community has any effect on the local labor market, this effect would be negative as wages in that sector are lower than the average, and concerns about our estimates being upward biased should be mitigated. We will go in depth about this concern in the robustness checks section.

1.6 Results

This section shows results based on an endogenous switching regression with selection on observables and MTE methods using the log real hourly wage rate, log real monthly earnings, and other non-wage labor market outcomes as the outcomes of interest. Results reported in this section use the preferred definition of informality (definition 2) as results under this definition are more conservative than when using definition 1.

1.6.1 Switching Regression with Selection Based on Observables

Results from RA estimates

The results presented here include RA estimates of informality using definition 2 on log wage rate, log monthly earnings, and other non-wage labor outcomes. Tables in the main text use only definition 2. Results using definition 1 can be found in Appendices A.3-A.7.

---

[6] All the covariates included in the X vector

[7] See Cameron S. V. and J. Heckman, (1998, 2001) and the papers they cite for other studies that used this exclusion restriction

Tables 1.7 and 1.8 show two types of estimates: unweighted and weighted. The weighted estimates attempt to correct for the selection into employment participation and use the inverse of the predicted probability of being employed as a weight. The Probit selection model that we use in order to compute the weights is presented in Appendix A.2. Since the differences between the two sets of estimates are negligible, we will describe only the results for the unweighted estimations.

Results from the unweighted RA estimations in Tables 1.7 and 1.8 indicate that the ATE of formality on the wage rate under definition 2 is 6.2% and the ATE on monthly earnings is 1.2%, respectively. This means that after controlling for selection on observables there is a positive and statistically significant wage gap between formal and informal workers. The earnings gap is also positive and statistically significant but small in magnitude. Results using definition 1, which can be found in the appendices, are slightly larger than the ones we presented. The formal-informal wage gap is 6.7% and the earnings gap is 2.4%.

Results from Table 1.9 show that formal workers are, on average, 5.3 percentage points (p.p.) more likely to have supplemental health insurance paid by the firms. They are also 4.4 p.p. more likely to receive unemployment insurance if they lost their jobs in the following period, and 30.9 p.p. more likely to have had a paid vacation in the last 12 months. Thus, formal workers do have a higher likelihood of enjoying employee benefits than informal employees. Formal workers are also 7.8 p.p. and 8.9 p.p. more likely to be satisfied, on average, with their jobs and job contracts. On the other side, formal workers are 1.5 p.p. less likely to be satisfied with their pay, 2.4 p.p. more concerned about losing their jobs, and 4.4 p.p. less confident in finding jobs if laid off, which reflects that although formal workers do not feel their wages are higher than if they were informal, they are happy with their choice as they get other non-wage benefits that compensate for the lower pay. Results for Definition 1 are provided in Appendix A5-A-7.

Results from IPWRA estimates

The IPWRA estimates were obtained from the sample using the informality definition 2 on log wage rate, log monthly earnings, and other non-pecuniary labor outcomes. The results are consistent with what we found under the RA estimation alone, although slightly smaller. As Table 1.10 and Table 1.11 show, the average treatment effect of formality on the log wage rate is 5.9%

and there is no effect on log earnings. Under RA results there was small positive effect on earnings and the effect on wage rate was 6.4%. This hints that the sign of the selection bias is positive as the results become smaller when we control for treatment assignment. The results using informality definition 1 can be found in Appendix A.8-A.9, but they are consistent with the estimates from the RA model. The formal-informal wage gap is 6.6% and the earnings gap is 2.2%, which are slightly higher than what we found when using informality definition 2.

Results for other outcomes are summarized in Table 1.12, but the complete results are included in Appendices A.10-A.13. In summary, formal workers are 5.2 p.p. more likely of having supplemental medical insurance paid by their employer, 4.6 p.p. more likely to receive unemployment benefits if they lose their job in the following period, and 30.8 p.p. more likely of having paid vacation time. Additionally, formal workers are 9 p.p. and 7.8 p.p. more likely to report being satisfied with their jobs and their work contract, respectively, but as it was shown before, formal workers are also more likely to be less satisfied with their job pay (-1.6 p.p.), more concerned about losing their job (-2.4 p.p.) and less confident in finding a new job if they get laid off (-4.3 p.p.).

Hence, both sets of estimates, RA and IPWRA, show evidence that when estimating the effect of formality on labor outcomes, it is important to look at other outcomes that are non-pecuniary, as workers could value them enough to be willing to earn a lower wage but receive other benefits in compensation.

1.6.2 Marginal Treatment Effects

Table 1.13 results for log wage rate show that highly educated individuals, who are married are more likely to be formal. Having a higher ratio of labor inspectors per 1,000 economic entities and a higher share of members in the community who work at the government also increase the likelihood of being formal. On the other side, living in Moscow or other regional centers reduces the likelihood of being formal compared to living in villages and the interaction term between the ratio of inspectors and distance to the inspection office does the same, which means that when we keep the number of inspectors constant and increase the distance, individuals are less likely to be formal.

Figure 1.3 reports the marginal treatment effect of formality on log wage rate evaluated at the mean value of all the covariates and allowing for U_P between 0 to 1 interval the MTE has a negative slope, which reflects the fact that individuals with a higher propensity to be formal are those who are getting the larger gains from working formally. The estimates also show that those individuals with a very low propensity of working formally can be negatively affected by having these types of jobs. Therefore, those individuals with a high unobservable cost of being formal, V, are better off working informally.

When looking at the covariance between the equations,we find that $cov(U_1, V) = -0.254$ and $cov(U_0, V) = -0.038$. These findings support the comparative advantage hypothesis proposed by Magnac (1991), which means that formal and informal workers self-select into the sectors in which they believe themselves to have a comparative advantage given their sets of abilities.

The average impact of formality on log wage rate is 2.5%, which is lower than what was reported under the RA and IPWRA models.The impact ranges from -80% to 60% for those with very high cost of being formal and very low cost, respectively. These results are consistent with the literature that stresses the importance of taking heterogeneity into account, as we could see there are some individuals who benefit from informality and some who are better off by working formally.

Table 1.14 includes the results of the MTE model for log monthly earnings. The results are very similar to the ones reported in Table 1.13. But they differ in the size of the ATE of formality on monthly earnings as in this case is -3.6%. There is also significant heterogeneity as it can be seen on Figure 1.4. The slope of the MTE is negative and magnitude and size of the covariances, $cov(U_1, V) = -0.19$ and $cov(U_0 = -0.014$, also favor a comparative advantage hypothesis, which is consistent with what we found when analyzing the impact of formality on wage rate.

When the results from Table 1.13 and Table 1.14 with the results from the RA and IPWRA model, we find that the wage gap between formal and informal workers is significantly reduced when controling for selection based on observable and unobservable characteristics and the earnings gap becomes negative. The first one goes from 6.4% to 2.5% and the latter goes from 1.4% to -3.6%.

Unfortunately, we cannot compute the effect of informality on other non-pecuniary outcomes under the same MTE framework used here as the current model is only developed for continuous outcomes, but it is future work to be done.

1.7 Robustness Checks

1.7.1 Robustness Checks

This section presents more estimations and results that aim to answer questions about the validity of the results. It includes estimations just for men in prime age to abstract from labor market participation issues and estimations without including Moscow to abstract from concerns about the results being driven by the main city in the country.

Selection into labor force participation

As in the MTE models previously estimated there is no correction for selection into the labor force and there is evidence that women have lower participation rates than men, then there could be concerns about the validity of the results. Table 15 shows participation rates for both men and women during the period of interest indicating that there is evidence of important differences by gender. Thus, the same models on log wage rate and log monthly earnings were estimated in a sample only containing men 20 to 59 years old, which are considered to be in prime working age. The results reported in Table 16 and 17 are robust to the change in the composition of the sample.

Table 1.16 shows the results for the MTE on log wage rate for only men. The ATE indicates that men working formally have wage rates 15% higher than those working informally and that there is evidence of comparative advantage based on the covariances of the selection equation and potential outcome. Figure 1.5 sheds light on the heterogeneity of the effect, as it could be seen that men with a very low cost of being formal highly benefit from it, getting wage rates up to 100% higher. On the other side, men who have a high cost of formality in terms of unobservables could also have significant loses of around -80% if they were to work formally.

Table 1.17 shows the results for the MTE on log monthly earnings only for men. Although the ATE of formality is positive, at 1.9%, it is similar to what was estimated under the RA and IPWRA estimators, which did not happen for the wage rate. Most of the differences between formal and informal workers are coming through the number of worked hours, as formal workers earn higher wage rates, which is especially true for men, but work less hours. Therefore, they may end up having monthly earnings that are not as high as expected or even lower than informal workers as they work less hours.

Results could be driven by the presence of Moscow in the data

In order to rule out that the results are driven due to the data including such a large and rich city as Moscow, which could be very different from the rest of the country, the model was re-estimated without Moscow city. Results are robust to the exclusion of Moscow from the estimation sample. As we can see in Table 1.18 and Table 1.19, increasing the number of labor inspectors increases the likelihood of formality, the interaction term has a negative sign which reflects the fact that if we keep the number of inspectors constant and increase the distance of the labor inspection offices, then there is more likelihood of informality. The ATE of formality on the log wage is 5.8% and on log monthly earnings is zero. The results for both outcomes indicate a slightly higher average effect of formality on the log wage and log monthly earnings without Moscow than in the whole sample.

Public sector wages

As a large share (47%) of the individuals in the sample have a job in the public sector (or firm owned by the government), there could be a concern about wages in this sector being higher than the average, thus biasing the estimates on formality upward. First, we provide descriptive statistics of average wage in the public sector by year and region, finding that wages in public sector jobs are lower on average than wages in other jobs (Table 1.20). Then, in Tables 1.21 - 1.22, we show that our results for both outcomes of interest are robust to controlling for average public sector wages by region and year. The coefficient of average public sector wage in the selection equation for formality is zero, which can be expected as we controlled for an ample set of characteristics that could have an effect on regional labor markets. In addition to that, the estimated ATE of formality on the log wage rate is 2.8% and the ATE of formality on earnings is -3.2%, which confirms that average public sector wage by year and region does not change our baseline results significantly. The slope of the MTE is negative and the covariances have the expected sign.

1.8 Concluding Remarks

This paper contributes to the literature in several ways. First, it presents a more comprehensive measure of informality, which is based on a legal definition plus tax compliance. Second, it focuses on pecuniary outcomes such as the wage rate and monthly earnings, but also on non-pecuniary outcomes such as job satisfaction, health insurance, paid vacation time, and others. Third, it recovers the marginal treatment effect of informality for a transition economy finding significant heterogenous returns, which has not been done before. Finally, it uses a unique regional database on labor enforcement, which includes the distance to the labor inspection offices and the number of labor inspectors. So we digitized seven years of reports from the Federal Inspection on Labor.

The results found in this paper show that the wage gap between formal and informal workers is significantly reduced when controling for selection based on observable and unobservable characteristics and the earnings gap becomes negative. The wage gap goes from 6.4%, when not controling for selection based on unobservable characteristics, to 2.5%. The earnings gap goes from 1.4% to -3.6%. We also find that there is significant heterogeneity in the formal-informal wage gap, as the effect of formality ranges from -60% for those with very high unobserved cost of being formal to 70% for those with an unobserved low cost.

There is evidence of comparative advantage, so individuals self-select into the formal or informal sector based on higher expected gains. Additionally, formal workers have a higher likelihood of receiving supplemental benefits and report to be more satisfied with their jobs, but at the same time formal workers are more concerned about job loss, not finding a job if they get laid off and are less satisfied with their pay. This suggests that workers take into consideration other job characteristics besides payment when deciding in which sector they want to work.

Lastly, the enforcement of the labor regulations through the use of labor inspectors has a positive impact in making individuals more likely to become formal as they increase the cost associated with informality. The share of individuals in the community who work for the government also plays a role in shifting individuals to formality, as jobs in the public sector tend to be formal and we believe it is easier to get a formal job if there is a close person who already has one.

Table 1.1: Proportion of formal and informal workers in Russia by definition 1 and 2

	Definition 1			Definition 2		
Year	Formal	Informal	Total	Formal	Informal	Total
2009	0.848	0.152	5,800	0.778	0.222	4,564
2010	0.854	0.146	8,837	0.796	0.204	7,040
2011	0.843	0.157	8,893	0.775	0.225	6,837
2012	0.828	0.172	9,081	0.807	0.193	7,273
2013	0.828	0.172	8,810	0.794	0.206	7,016
2014	0.823	0.177	7,335	0.796	0.204	5,977
2015	0.842	0.158	7,030	0.789	0.211	6,613
2016	0.833	0.167	7,143	0.782	0.218	6,804
Total	0.837	0.163	62,929	0.790	0.210	52,124

Table 1.2: Average hourly real wage rate and monthly earnings in rubles for the sample

Year	Wage rate	Standard Dev.	Monthly earnings	Standard Dev.
2009	131.5	99.8	23,315.8	18,193.9
2010	131.3	99.5	23,813.3	19,549.7
2011	138.4	103.4	24,942.4	19,497.5
2012	149.8	116.5	27,046.4	22,089.4
2013	154.1	113	27,752.7	21,406.8
2014	156.3	117.6	28,093.6	21,509.8
2015	142.4	101	25,338.4	18,085.4
2016	140.6	97.4	25,136.6	18,357.8

Note: For reference purposes, the exchange rate at December 2015 was 1 USD= 57.86 Rubles. The monthly minimum wage in Russia in January 2016 was RUB 7,537 (or US$130.26)

Table 1.3: Other non-pecuniary outcomes of interest based on sample of definition 2

Variable	**Mean**	**No.of observations**
Has supplementary insurance paid by the firm	0.063	51,989
Receives unemployment benefits if loses job in t+1	0.126	2,114
Had paid vacation in the last 12 months	0.715	48,033
Satisfied with job	0.665	51,709
Satisfied with work contract	0.644	51,614
Satisfied with pay	0.358	51,517
Not concerned about chance of job loss	0.284	51,816
Confident in finding a job if laid off	0.421	47,227

Note: All the outcomes are binary as the potential answers are yes/no. Similar results are found when using definition 1.

Table 1.4: Inspectors per 1,000 economic entities per year

Year	**Ratio of Inspectors**
2009	0.561
2010	0.43
2011	0.485
2012	0.404
2013	0.344
2014	0.359
2015	0.37
2016	0.337

Table 1.5: Inspectors per 1,000 economic entities per okrug

Okrug	**Ratio of Inspectors**
Central	0.386
North West	0.454
Sotuh	0.396
Volga	0.459
Urals	0.402
Siberia	0.341
Far East	0.452

Table 1.6: Descriptive statistics

Variable	Formal	Informal	t-test
Female=1	0.54	0.44	-18.22
Age at the time of survey	39.7	37.1	-23.03
	(10.7)	(10.3)	
Years of schooling	12.6	11.7	-37.59
	(2.2)	(2.2)	
Married=1	0.61	0.52	-18.8
Schooling level of self			
Primary or lower (omitted cat.)	0.07	0.12	18.85
Secondary=1	0.31	0.44	26.3
Upper vocational=1	0.24	0.2	-9.33
Higher education=1	0.36	0.21	-29.17
Schooling level of parents			
Secondary or lower (omitted cat.)	0.4	0.41	1.66
Upper vocational=1	0.23	0.21	-3.9
Higher education=1	0.2	0.17	-4.8
Missing=1	0.15	0.18	7.6
Urban Status			
Moscow (omitted cat.)	0.08	0.07	-3.12
Regional center=1	0.33	0.41	15.56
Other city=1	0.36	0.32	-10.09
Real monthly labor earnings (In Rubles)	26,322	25,923	-1.79
	(20,612)	(19,270)	
Real wage rate (in Rubles)	147.9	136.2	-11.9
	(109.8)	(102)	
Usual hours worked per month	183.23	201.77	32.94
	(47.0)	(63.3)	
Population	1,365,480	1,353,607	-8.5
	(3,187,766)	(3,024,007)	
Number of inspectors per 1000 economic entities	0.39	0.39	1.88
	(0.25)	(0.27)	
Distance to labor inspection office	118.2	79.44	-2.73
	(160.8)	(97.2)	
Share of government	0.41	0.39	-16.28
employment in community	(0.11)	(0.12)	
Unemployment rate in the region	5.67	5.87	8.33
	(2.22)	(2.33)	
Has supplementary insurance paid by firm	0.074	0.015	-21.66
Receives unempl. benefits if loses job in t+1	0.157	0.109	-3.01
Had paid vacation in the last 12 months	0.781	0.418	-70.28
Satisfied with job	0.687	0.574	-21.88
Satisfied with work contract	0.665	0.557	-20.62
Satisfied with pay	0.353	0.355	0.344
Not concerned about chance of job loss	0.278	0.306	5.61
Confident in finding a job if laid off	0.404	0.489	14.3

Standard deviation in parenthesis.
Constant prices using CPI base 2016.

The monthly minimum wage in Russia in January 2016 was RUB$7,537 (or US$130.26).

Table 1.7: RA Results using informality definition 2 and log wage rate as outcome variable

Variables	**Informal**	**Formal**	**Informal W**	**Formal W**
Dep Var: Log wage rate	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	-0.260***	-0.312***	-0.252***	-0.306***
	(0.010)	(0.005)	(0.011)	(0.005)
Age	0.042***	0.040***	0.039***	0.038***
	(0.004)	(0.002)	(0.004)	(0.002)
Age squared	-0.060***	-0.052***	-0.057***	-0.050***
	(0.005)	(0.002)	(0.005)	(0.002)
Married (=1)	0.099***	0.037***	0.095***	0.038***
	(0.011)	(0.005)	(0.011)	(0.005)
Schooling categories				
High School	0.076***	0.076***	0.076***	0.078***
	(0.016)	(0.010)	(0.016)	(0.011)
Technical/Vocational	0.208***	0.175***	0.207***	0.182***
	(0.018)	(0.011)	(0.018)	(0.011)
College or more	0.412***	0.464***	0.410***	0.470***
	(0.018)	(0.011)	(0.019)	(0.011)
Log population site	0.024***	0.026***	0.025***	0.027***
	(0.006)	(0.003)	(0.006)	(0.003)
Log distance inspection	-0.041***	-0.018***	-0.040***	-0.021***
	(0.007)	(0.003)	(0.008)	(0.003)
Unemployment rate at region	-0.068***	-0.031***	-0.068***	-0.032***
	(0.004)	(0.002)	(0.004)	(0.002)
Constant	4.140***	3.793***	4.200***	3.830***
	(0.102)	(0.048)	(0.105)	(0.050)
ATE (Formal=1 vs Informal=0)	0.062***		0.064***	
	(0.006)		(0.006)	
Observations	52,124	52,124	52,124	52,124

Dependent variable: Log real hourly wage rate.
W stands for weighted estimates.
Clustered standard errors at the individual level in parentheses.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 1.8: RA Results using informality definition 2 and log monthly earnings as outcome variable

Variables	**Informal**	**Formal**	**Informal W**	**Formal W**
Dep Var: Log monthly earnings	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	-0.374***	-0.413***	-0.369***	-0.409***
	(0.011)	(0.005)	(0.012)	(0.006)
Age	0.061***	0.048***	0.059***	0.047***
	(0.004)	(0.002)	(0.004)	(0.002)
Age squared	-0.085***	-0.063***	-0.083***	-0.062***
	(0.005)	(0.003)	(0.006)	(0.003)
Married (=1)	0.093***	0.037***	0.091***	0.038***
	(0.011)	(0.006)	(0.012)	(0.006)
Schooling categories				
High School	0.091***	0.065***	0.085***	0.066***
	(0.018)	(0.011)	(0.018)	(0.011)
Technical/Vocational	0.184***	0.147***	0.185***	0.154***
	(0.020)	(0.011)	(0.020)	(0.012)
College or more	0.365***	0.408***	0.364***	0.414***
	(0.020)	(0.011)	(0.021)	(0.012)
Log population site	0.015**	0.028***	0.020***	0.030***
	(0.006)	(0.003)	(0.006)	(0.003)
Log distance inspection	-0.049***	-0.018***	-0.048***	-0.020***
	(0.008)	(0.004)	(0.008)	(0.004)
Unemployment rate at region	-0.068***	-0.030***	-0.068***	-0.030***
	(0.004)	(0.002)	(0.005)	(0.002)
Constant	9.239***	8.852***	9.246***	8.870***
	(0.110)	(0.051)	(0.115)	(0.052)
ATE (Formal=1 vs Informal=0)	0.012**		0.014*	
	(0.007)		(0.007)	
Observations	50,747	50,747	50,747	50,747

Dependent variable: Log real monthly earnings.
W stands for weighted estimates.
Clustered standard errors at the individual level in parentheses.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 1.9: RA Results using informality definition 2 and other non-wage labor market outcomes

Dependent Variable	**ATE (Formal=1)**	**No. Obs.**
Has supplementary insurance paid by the firm	0.053***	51,989
Receives unemployment benefits if loses job in t+1	0.044***	2,114
Had paid vacation in the last 12 months	0.309***	48,033
Satisfied with job	0.089***	51,709
Satisfied with work contract	0.078***	51,614
Satisfied with pay	-0.015***	51,517
Not concerned about chance of job loss	-0.024***	51,816
Confident in finding a job if laid off	-0.044***	47,227

The table reports the ATE of formality on each of the dependent variables. The outcome equation was modeled as a probit.
Regression tables in Appendix.
Clustered standard errors at the individual level in parentheses.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 1.10: IPWRA Results using informality definition 2 and log wage rate as outcome variable

Variables	**Informal [D=0]**	**Formal [D=1]**	**Selection Eq.**
Dep Var: Log wage rate	Coef./SE	Coef./SE	Coef./SE
Female	-0.253***	-0.315***	0.146***
	(0.012)	(0.005)	(0.013)
Age	0.050***	0.039***	-0.021***
	(0.005)	(0.002)	(0.005)
Age squared	-0.069***	-0.051***	0.041***
	(0.006)	(0.002)	(0.006)
Married (=1)	0.069***	0.041***	0.204***
	(0.013)	(0.005)	(0.014)
Schooling categories			
High School	0.064***	0.075***	0.085***
	(0.016)	(0.010)	(0.023)
Technical/Vocational	0.198***	0.171***	0.355***
	(0.019)	(0.011)	(0.025)
College or more	0.418***	0.456***	0.610***
	(0.020)	(0.011)	(0.025)
Log population site	0.026***	0.025***	0.010
	(0.007)	(0.003)	(0.008)
Log distance inspection	-0.041***	-0.021***	0.147***
	(0.009)	(0.003)	(0.012)
Unemployment rate at region	-0.057***	-0.033***	-0.006
	(0.005)	(0.002)	(0.005)
Share of public employment in community			0.710***
			(0.063)
Inspectors per 1,000 entities			0.179**
			(0.085)
Distance x Ratio Inspectors			-0.065***
			(0.021)
Constant	3.898***	3.845***	-0.332**
	(0.127)	(0.048)	(0.138)
ATE (Formal=1 vs Informal=0)	0.059***		
	(0.006)		
Observations	52,124	52,124	52,124

Dependent variable: Log real hourly wage rate.
Clustered standard errors at the individual level in parentheses.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 1.11: IPWRA Results using informality definition 2 and log monthly earnings as outcome variable

Variables	**Informal [D=0]**	**Formal [D=1]**	**Selection Eq.**
Dep Var: Log earnings	Coef./SE	Coef./SE	Coef./SE
Female	-0.375***	-0.414***	0.149***
	(0.013)	(0.005)	(0.014)
Age	0.065***	0.048***	-0.022***
	(0.005)	(0.002)	(0.005)
Age squared	-0.088***	-0.063***	0.042***
	(0.006)	(0.003)	(0.006)
Married (=1)	0.059***	0.042***	0.197***
	(0.013)	(0.006)	(0.014)
Schooling categories			
High School	0.082***	0.063***	0.085***
	(0.019)	(0.011)	(0.023)
Technical/Vocational	0.180***	0.143***	0.356***
	(0.021)	(0.011)	(0.025)
College or more	0.377***	0.400***	0.606***
	(0.022)	(0.011)	(0.026)
Log population site	0.013*	0.026***	0.015**
	(0.007)	(0.003)	(0.008)
Log distance inspection	-0.049***	-0.020***	0.151***
	(0.009)	(0.004)	(0.012)
Unemployment rate at region	-0.056***	-0.032***	-0.006
	(0.006)	(0.002)	(0.005)
Share public employment in community			0.723***
			(0.064)
Inspectors per 1000 entities			0.172**
			(0.086)
Distance x Ratio Inspectors			-0.065***
			(0.021)
Constant	9.088***	8.901***	-0.356**
	(0.132)	(0.051)	(0.140)
ATE (Formal=1 vs Informal=0)	0.009		
	(0.007)		
Observations	50,747	50,747	50,747

Dependent variable: Log real monthly earnings.
Clustered standard errors at the individual level in parentheses.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 1.12: IPWRA Results using informality definition 2 and other non-wage labor market outcomes

Dependent Variable	**ATE**	**No. Obs.**
Has supplementary insurance paid by the firm	0.052***	51,989
Receives unemployment benefits if loses job in t+1	0.046***	2,114
Had paid vacation in the last 12 months	0.308***	48,033
Satisfied with job	0.090***	51,709
Satisfied with work contract	0.078***	51,614
Satisfied with pay	-0.016***	51,517
Not concerned about chance of job loss	-0.024***	51,816
Confident in finding a job if laid off	-0.043***	47,227

The table reports the ATE of formality on each of the dependent variables. The outcome equation was modeled as a probit. Regression tables in Appendix.

Clustered standard errors at the individual level in parentheses.

The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 1.13: Outcome and selection equation: Log real hourly wage rate

Variables	**Formal [D=1]**	**Informal [D=0]**	**Selection Eq.]**
Female	-0.298***	-0.265***	0.150***
	(0.005)	(0.013)	(0.013)
Age	0.038***	0.043***	-0.021***
	(0.002)	(0.004)	(0.005)
Age squared	-0.048***	-0.061***	0.042***
	(0.002)	(0.005)	(0.006)
Married (=1)	0.059***	0.093***	0.195***
	(0.006)	(0.014)	(0.013)
Schooling categories (omited: primary or less)			
High School	0.088***	0.074***	0.083***
	(0.010)	(0.016)	(0.023)
Technical/Vocational	0.217***	0.199***	0.346***
	(0.011)	(0.024)	(0.025)
College or more	0.530***	0.394***	0.616***
	(0.011)	(0.034)	(0.025)
Log population site	0.025***	0.024***	0.016**
	(0.003)	(0.006)	(0.007)
Log distance inspection	-0.006*	-0.045***	0.163***
	(0.004)	(0.010)	(0.012)
Unemployment rate at region	-0.033***	-0.068***	-0.002
	(0.002)	(0.004)	(0.005)
Share public employment in community			0.840***
			(0.062)
Inspectors per 1000 entities			0.340***
			(0.086)
DistancexRatioInspectors			-0.090***
			(0.021)
Constant	3.632***	4.106***	-0.558***
	(0.049)	(0.113)	(0.138)
σ_{DV}	0.254***	-0.038***	
σ_{1V} - σ_{0V}	-0.292***		
	(0.046)		
ATE	0.025***		
	(0.007)		
Number of Observations	52,124		

Dependent variable: Log real hourly wage rate.
Standard errors computed using the delta method.
The model includes education of the parents, urban status, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 1.14: Outcome and selection equation: Log monthly earnings

Variables	**Formal [D=1]**	**Informal [D=0]**	**Selection Eq.]**
Female	-0.402***	-0.376***	0.147***
	(0.006)	(0.013)	(0.013)
Age	0.047***	0.061***	-0.023***
	(0.002)	(0.004)	(0.005)
Age squared	-0.060***	-0.085***	0.043***
	(0.003)	(0.006)	(0.006)
Married (=1)	0.053***	0.091***	0.190***
	(0.006)	(0.014)	(0.014)
Schooling categories			
High School	0.075***	0.090***	0.084***
	(0.011)	(0.018)	(0.023)
Technical/Vocational	0.179***	0.181***	0.350***
	(0.012)	(0.025)	(0.025)
College or more	0.456***	0.358***	0.611***
	(0.012)	(0.035)	(0.025)
Log population site	0.027***	0.015**	0.021***
	(0.003)	(0.006)	(0.007)
Log distance inspection	-0.009**	-0.051***	0.169***
	(0.004)	(0.010)	(0.012)
Unemoloyment rate at region	-0.031***	-0.068***	-0.003
	(0.002)	(0.004)	(0.005)
Share public employment in community			0.865***
			(0.065)
Inspectors per 1000 entities			0.343***
			(0.088)
DistancexRatioInspectors			-0.099***
			(0.021)
Constant	8.732***	9.227***	-0.582***
	(0.052)	(0.118)	(0.140)
σ_{DV}	0.190***	-0.014***	
σ_{1V} - σ_{0V}	-0.205***		
	(0.053)		
ATE	-0.036***		
	(0.077)		
Number of Observations	50,747		

Dependent variable: Log real monthly earnings.
Standard errors computed using the delta method.
The model includes education of the parents, urban status, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 1.15: Labor force participation by gender for 2009-2016

Year	**Female (%)**	**Male (%)**
2009	58.8	70.4
2010	56.5	70.7
2011	56.9	71.1
2012	56.9	71.2
2013	56.7	71.4
2014	56.7	71.6
2015	56.7	71.9
2016	56.9	72.1

Source: International Labour Organization, ILOSTAT database. Published by: The World Bank.

Table 1.16: Outcome and selection equation: Log wage rate- only men

Variables	Formal [D=1]	Informal [D=0]	Selection Eq.]
Age	0.045***	0.046***	-0.036***
	(0.003)	(0.005)	(0.007)
Age squared	-0.060***	-0.067***	0.061***
	(0.004)	(0.007)	(0.009)
Married (=1)	0.143***	0.131***	0.229***
	(0.009)	(0.017)	(0.020)
Schooling categories (omited: primary or less)			
High School	0.080***	0.072***	0.070**
	(0.013)	(0.020)	(0.029)
Technical/Vocational	0.208***	0.202***	0.350***
	(0.015)	(0.028)	(0.033)
College or more	0.449***	0.281***	0.588***
	(0.015)	(0.035)	(0.033)
Log population site	0.022***	0.032***	0.019*
	(0.004)	(0.008)	(0.010)
Log distance inspection	-0.019***	-0.053***	0.153***
	(0.005)	(0.012)	(0.017)
Unemployment rate at region	-0.037***	-0.074***	-0.018**
	(0.003)	(0.005)	(0.007)
Share public employment in community			1.062***
			(0.087)
Inspectors per 1000 entities			0.141
			(0.116)
DistancexRatioInspectors			-0.055*
			(0.028)
Constant	3.643***	4.044***	-0.276
	(0.071)	(0.149)	(0.192)
σ_{DV}	0.2285	-0.1069	
σ_{1V} - σ_{0V}	-0.336***		
	(0.044)		
ATE	0.151***		
	(0.005)		
Number of Observations	24,638		

Dependent variable: Log real hourly wage rate.
The estimation sample includes only males in prime age.
Robust standard errors computed using the Delta Method.
The model includes education of the parents, urban status, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 1.17: Outcome and selection equation: Log monthly earnings- only men

Variables	**Formal [D=1]**	**Informal [D=0]**	**Selection Eq.]**
Age	0.054***	0.062***	-0.038***
	(0.003)	(0.006)	(0.007)
Age squared	-0.074***	-0.088***	0.063***
	(0.004)	(0.008)	(0.009)
Married (=1)	0.172***	0.172***	0.220***
	(0.009)	(0.018)	(0.020)
Schooling categories			
High School	0.079***	0.096***	0.074**
	(0.014)	(0.022)	(0.029)
Technical/Vocational	0.178***	0.191***	0.350***
	(0.016)	(0.030)	(0.034)
College or more	0.399***	0.258***	0.585***
	(0.016)	(0.037)	(0.034)
Log population site	0.020***	0.013	0.024**
	(0.004)	(0.008)	(0.010)
Log distance inspection	-0.010*	-0.055***	0.167***
	(0.006)	(0.013)	(0.017)
Unemployment rate at region	-0.031***	-0.076***	-0.020***
	(0.003)	(0.006)	(0.007)
Share public employment in community			1.065***
			(0.090)
Inspectors per 1000 entities			0.184
			(0.120)
DistancexRatioInspectors			-0.070**
			(0.029)
Constant	8.636***	9.315***	-0.315
	(0.076)	(0.160)	(0.196)
σ_{DV}	0.2285	-0.1069	
σ_{1V} - σ_{0V}	-0.256***		
	(0.051)		
ATE	0.019***		
	(0.006)		
Number of Observations	23,829		

Dependent variable: Log real monthly earnings.
The estimation sample includes only males in prime age.
Robust standard errors computed using the Delta Method.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 1.18: Outcome and selection equation: Log real hourly wage- without Moscow

Variables	**Formal [D=1]**	**Informal [D=0]**	**Selection Eq.]**
Female	-0.298***	-0.266***	0.150***
	(0.005)	(0.013)	(0.013)
Age	0.039***	0.043***	-0.022***
	(0.002)	(0.004)	(0.005)
Age squared	-0.049***	-0.062***	0.042***
	(0.002)	(0.005)	(0.006)
Married (=1)	0.057***	0.087***	0.196***
	(0.006)	(0.014)	(0.013)
Schooling categories			
High School	0.089***	0.070***	0.083***
	(0.010)	(0.017)	(0.023)
Technical/Vocational	0.217***	0.189***	0.346***
	(0.011)	(0.024)	(0.025)
College or more	0.531***	0.381***	0.616***
	(0.011)	(0.034)	(0.025)
Log population site	0.064***	0.058***	0.004
	(0.001)	(0.003)	(0.004)
Log distance inspection	-0.010***	-0.048***	0.161***
	(0.004)	(0.010)	(0.012)
Unemployment rate at region	-0.030***	-0.064***	-0.004
	(0.002)	(0.004)	(0.005)
Share public employment in community			0.810***
			(0.062)
Inspectors per 1000 entities			0.321***
			(0.086)
DistancexRatioInspectors			-0.084***
			(0.021)
Constant	3.387***	3.865***	-0.457***
	(0.047)	(0.109)	(0.131)
σ_{DV}	0.254	-0.0619	
σ_{1V} - σ_{0V}	-0.316***		
	(0.044)		
ATE	0.058***		
	(0.007)		
Number of Observations	52,124		

Dependent variable: Log real hourly wage rate.
The estimation sample does not include observations from Moscow.
Robust standard errors computed using the Delta Method.
The model includes education of the parents, urban status, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 1.19: Outcome and selection equation: Log monthly earnings- without Moscow

Variables	**Formal [D=1]**	**Informal [D=0]**	**Selection Eq.]**
Female	-0.402***	-0.377***	0.147***
	(0.006)	(0.013)	(0.013)
Age	0.047***	0.062***	-0.023***
	(0.002)	(0.004)	(0.005)
Age squared	-0.061***	-0.086***	0.043***
	(0.003)	(0.005)	(0.006)
Married (=1)	0.051***	0.086***	0.191***
	(0.006)	(0.014)	(0.014)
Schooling categories			
High School	0.075***	0.087***	0.084***
	(0.011)	(0.018)	(0.023)
Technical/Vocational	0.178***	0.173***	0.351***
	(0.012)	(0.024)	(0.025)
College or more	0.457***	0.347***	0.611***
	(0.012)	(0.032)	(0.025)
Log population site	0.068***	0.047***	0.005
	(0.001)	(0.003)	(0.004)
Log distance inspection	-0.012***	-0.054***	0.167***
	(0.004)	(0.010)	(0.012)
Unemployment rate at region	-0.028***	-0.065***	-0.005
	(0.002)	(0.004)	(0.005)
Share public employment in community			0.833***
			(0.064)
Inspectors per 1000 entities			0.323***
			(0.088)
DistancexRatioInspectors			-0.093***
			(0.021)
Constant	8.480***	8.993***	-0.455***
	(0.049)	(0.111)	(0.133)
σ	0.1873	-0.0355	
σ_{1V} - σ_{0V}	-0.223***		
	(0.047)		
ATE	-0.004		
	(0.005)		
Number of Observations	50,747		

Dependent variable: Log real monthly earnings.
The estimation sample does not include observations from Moscow.
Robust standard errors computed using the Delta Method.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 1.20: Average public sector wages by year and region (mean and standard deviation below)

Year/Region	Central	North Western	South	Volga	Urals	Siberia	Far East	Total
2009	147.9	168.5	102.9	105	96.5	109.4	121.2	123.8
	102	103.9	68.4	69.8	57.9	64	73.8	85.8
2010	142.8	150.4	102.1	104.1	99.5	104.7	120.3	120
	103.3	86	65.2	61.9	57.6	61.1	73.2	81.5
2011	152.2	174.1	100	103.3	105.6	116.6	130.8	127.1
	108.5	108	59.2	62.9	61.5	76.4	78.5	89.2
2012	172.6	184.3	110.8	118.8	113.1	120.8	138	140.1
	126.4	126.2	61.2	76.6	70.3	75	86.2	101.1
2013	173.4	199.8	116.6	126.9	127.9	131.8	151.6	147
	115.7	130.5	71.6	84.6	88.1	85.7	103	102.3
2014	181.4	206.1	121.1	124.8	123.3	131.4	145.2	148.4
	111.4	119.5	71.6	74.7	76	78.1	95.5	96.2
2015	169.1	190.9	108	114.9	111.7	117.8	143	136.3
	116.5	108.9	64.3	75.9	62.7	71.2	84	93.8
2016	158.2	182.6	105.2	115.5	120.1	125	137.1	133.3
	103.5	93.9	63.5	59.8	68	73	71.2	83
Total	161.8	181.6	108.4	114.4	112	119.8	135.8	134.5
	112.3	111.9	65.9	71.8	69.1	74.3	84.2	92.6

Table 1.21: Outcome and selection equation: Log real hourly wage- average public sector wage

Variables	Formal	Informal	Selection Model]
Female	-0.298***	-0.265***	0.150***
	(0.005)	(0.013)	(0.013)
Age	0.038***	0.043***	-0.021***
	(0.002)	(0.004)	(0.005)
Age squared	-0.048***	-0.061***	0.042***
	(0.002)	(0.005)	(0.006)
Married (=1)	0.059***	0.093***	0.195***
	(0.006)	(0.014)	(0.013)
Schooling categories (omitted: primary or less)			
High School	0.089***	0.074***	0.083***
	(0.010)	(0.016)	(0.023)
Technical/Vocational	0.217***	0.198***	0.346***
	(0.011)	(0.024)	(0.025)
College or more	0.530***	0.393***	0.616***
	(0.011)	(0.034)	(0.025)
Log population site	0.025***	0.024***	0.015**
	(0.003)	(0.006)	(0.007)
Log distance inspection	-0.006*	-0.045***	0.163***
	(0.004)	(0.010)	(0.012)
Unemployment rate at region	-0.033***	-0.068***	-0.003
	(0.002)	(0.004)	(0.005)
Average public sector wage	0.002***	0.003*	0.000
	(0.001)	(0.001)	(0.002)
Share public employment in community			0.839***
			(0.062)
Inspectors per 1000 entities			0.337***
			(0.086)
DistancexRatioInspectors			-0.089***
			(0.021)
Constant	3.330***	3.738***	-0.625**
	(0.107)	(0.222)	(0.281)
σ	0.253	-0.04	
σ_{1V} - σ_{0V}	-0.293***		
	0.045		
ATE	0.0285***		
	0.0074		
Number of Observations	52,124		

Dependent variable: Log real monthly earnings.
The estimation sample does not include observations from Moscow.
Robust standard errors computed using the Delta Method.
The model includes education of the parents, urban status, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 1.22: Outcome and selection equation: Log monthly earnings- average public sector wage

Variables	**Formal**	**Informal**	**Selection Model]**
Female	-0.402***	-0.376***	0.147***
	(0.006)	(0.013)	(0.013)
Age	0.047***	0.061***	-0.023***
	(0.002)	(0.004)	(0.005)
Age squared	-0.060***	-0.085***	0.043***
	(0.003)	(0.006)	(0.006)
Married (=1)	0.053***	0.090***	0.190***
	(0.006)	(0.014)	(0.014)
Schooling categories (omited: primary or less)			
High School	0.075***	0.090***	0.084***
	(0.011)	(0.018)	(0.023)
Technical/Vocational	0.179***	0.180***	0.350***
	(0.012)	(0.025)	(0.025)
College or more	0.457***	0.357***	0.611***
	(0.012)	(0.034)	(0.025)
Log population site	0.027***	0.014**	0.021***
	(0.003)	(0.006)	(0.007)
Log distance inspection	-0.009**	-0.052***	0.169***
	(-0.004)	(-0.010)	(-0.012)
Unemployment rate at region	-0.031***	-0.068***	-0.003
	(-0.002)	(-0.004)	(-0.005)
Average public sector wage	0.002***	0.003**	0.000
	(-0.001)	(-0.001)	(-0.002)
Share public employment in community			0.864***
			(0.065)
Inspectors per 1000 entities			0.339***
			(0.088)
DistancexRatioInspectors			-0.098***
			(0.021)
Constant	8.382***	8.750***	-0.585**
	(0.112)	(0.236)	(0.286)
σ	0.19	-0.0163	
σ_{1V} - σ_{0V}	-0.206***		
	0.051		
ATE	-0.0326***		
	0.00739		
Number of Observations	50,747		

Dependent variable: Log real monthly earnings.
The estimation sample does not include observations from Moscow.
Robust standard errors computed using the Delta Method.
The model includes education of the parents, urban status, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Figure 1.1: Russian Longitudinal Monitoring Survey respondent tree

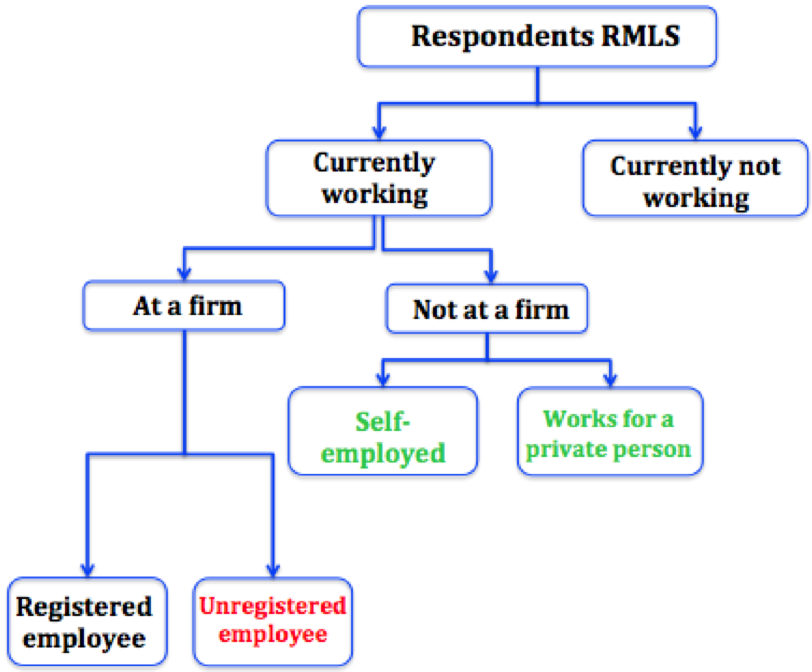


Figure 1.2: Kernel density estimate of the (log) distance to the nearest inspection office by formality status of the worker

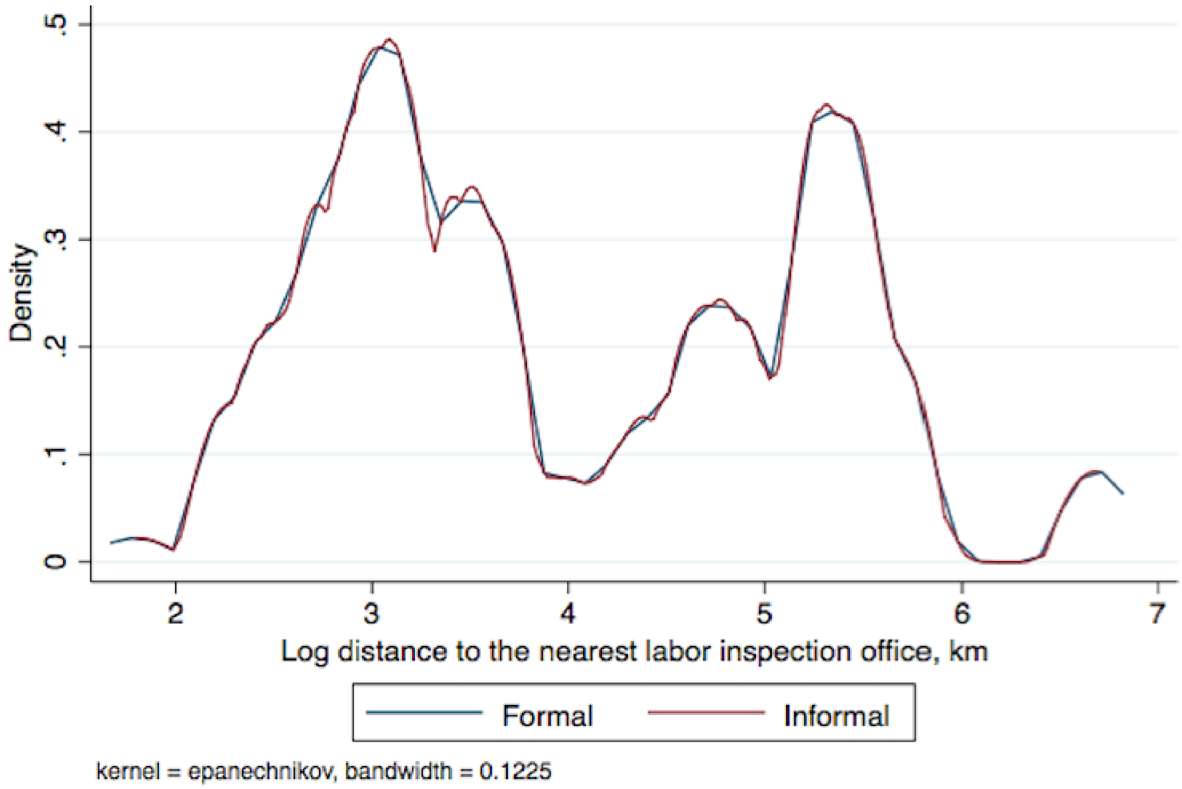


Figure 1.3: Estimated marginal treatment effects - log real hourly wage rate

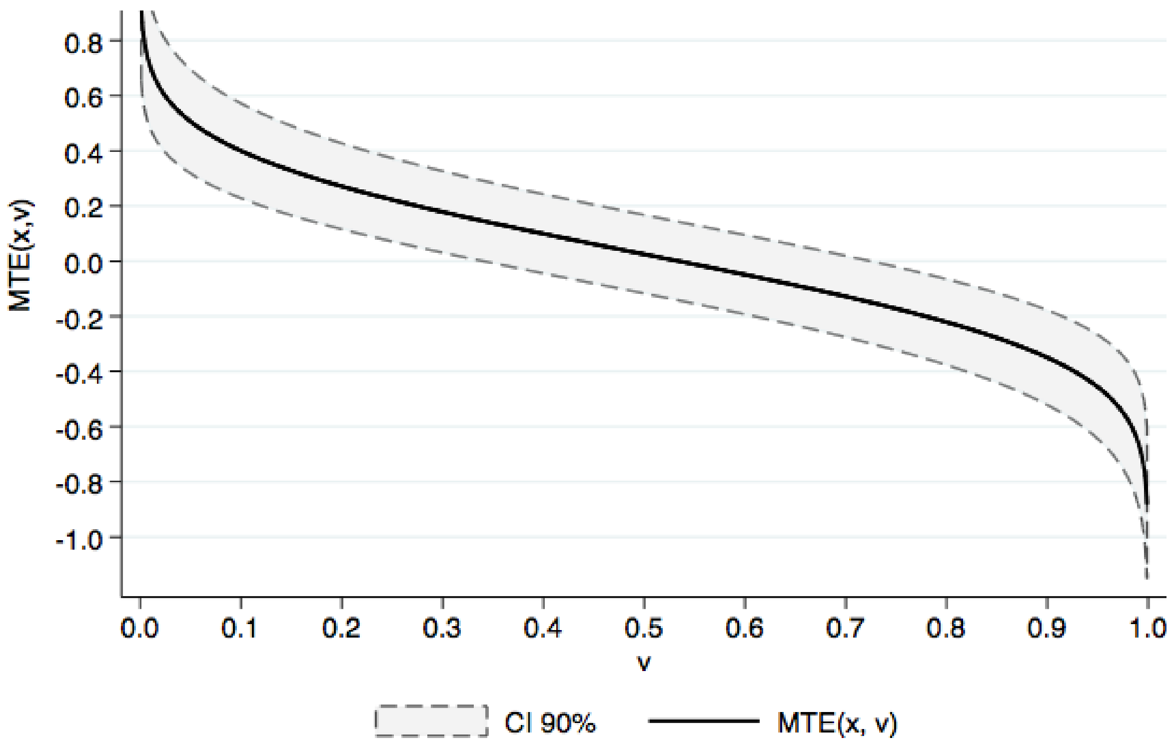


Figure 1.4: Estimated marginal treatment effects - log monthly earnings

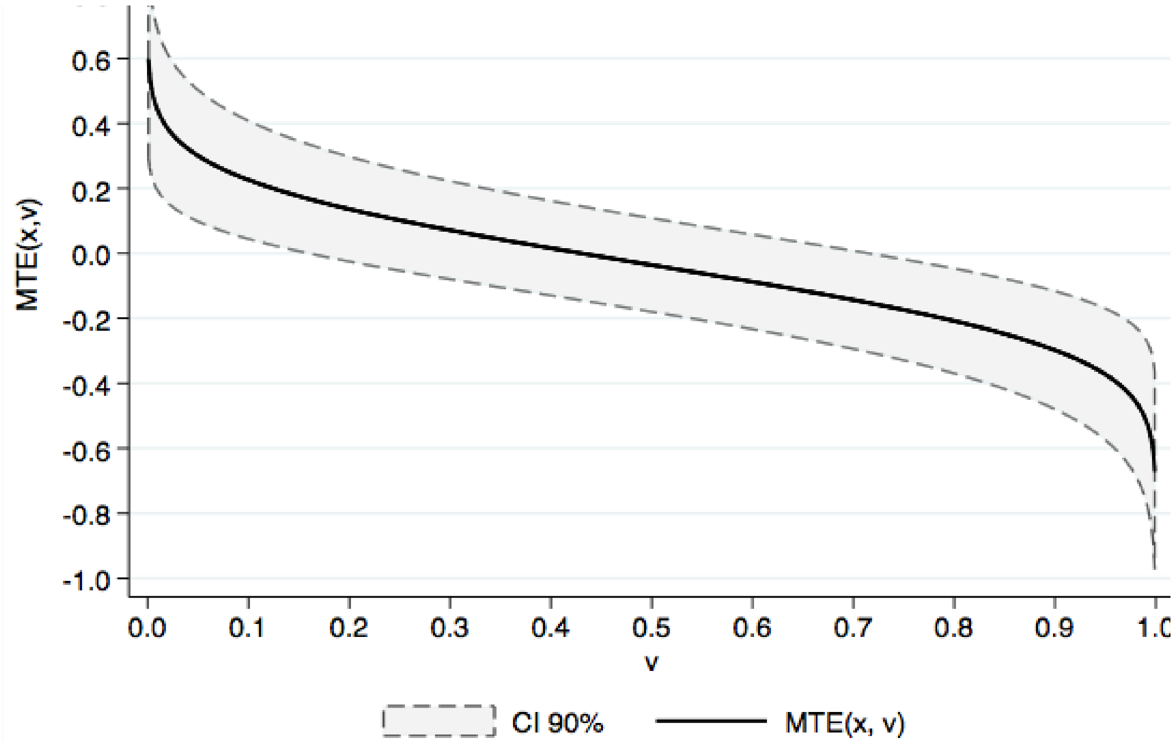


Figure 1.5: Estimated marginal treatment effects on log real hourly wage rate - only men

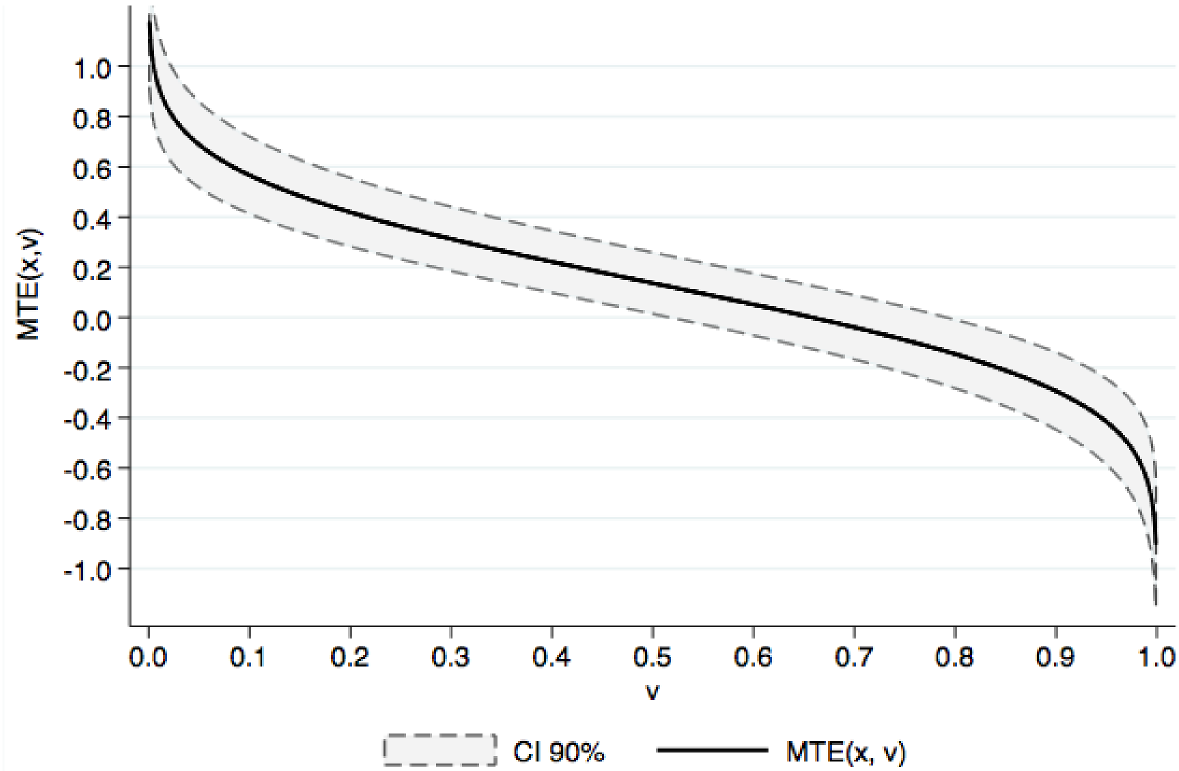


Figure 1.6: Estimated marginal treatment effects on log monthly earnings- only men

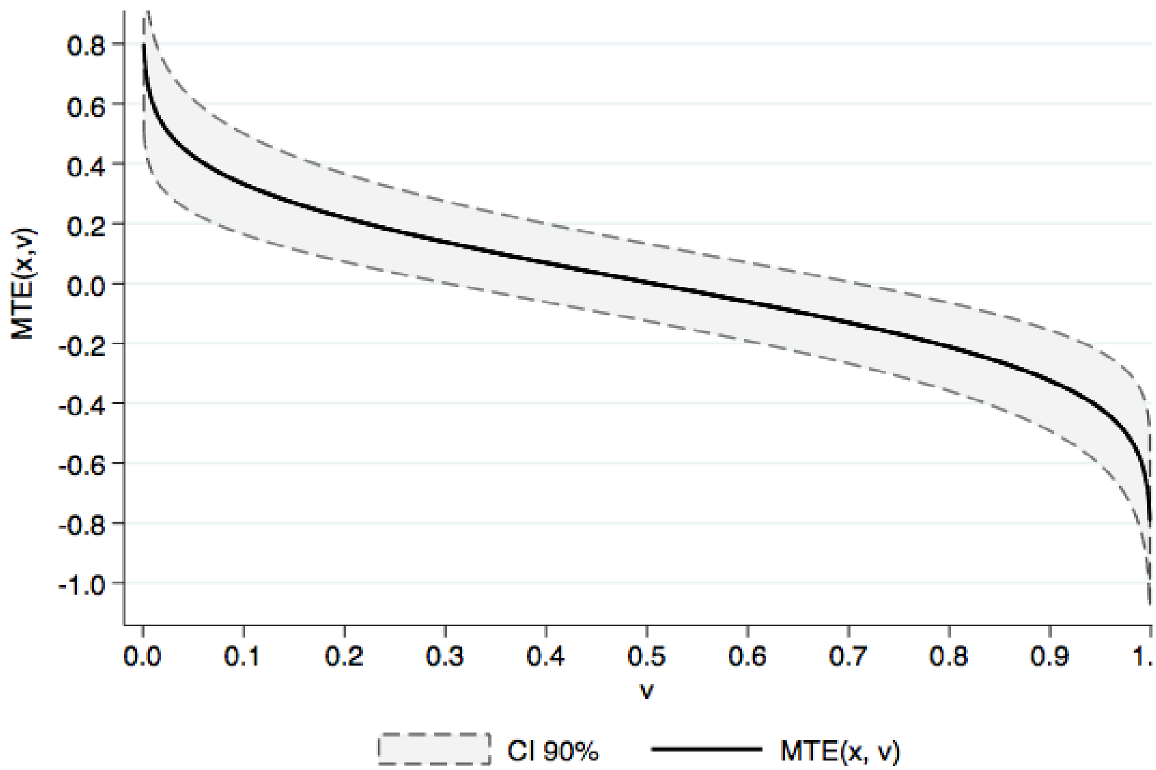


Figure 1.7: Estimated marginal treatment effects for log real hourly wage rate - without Moscow

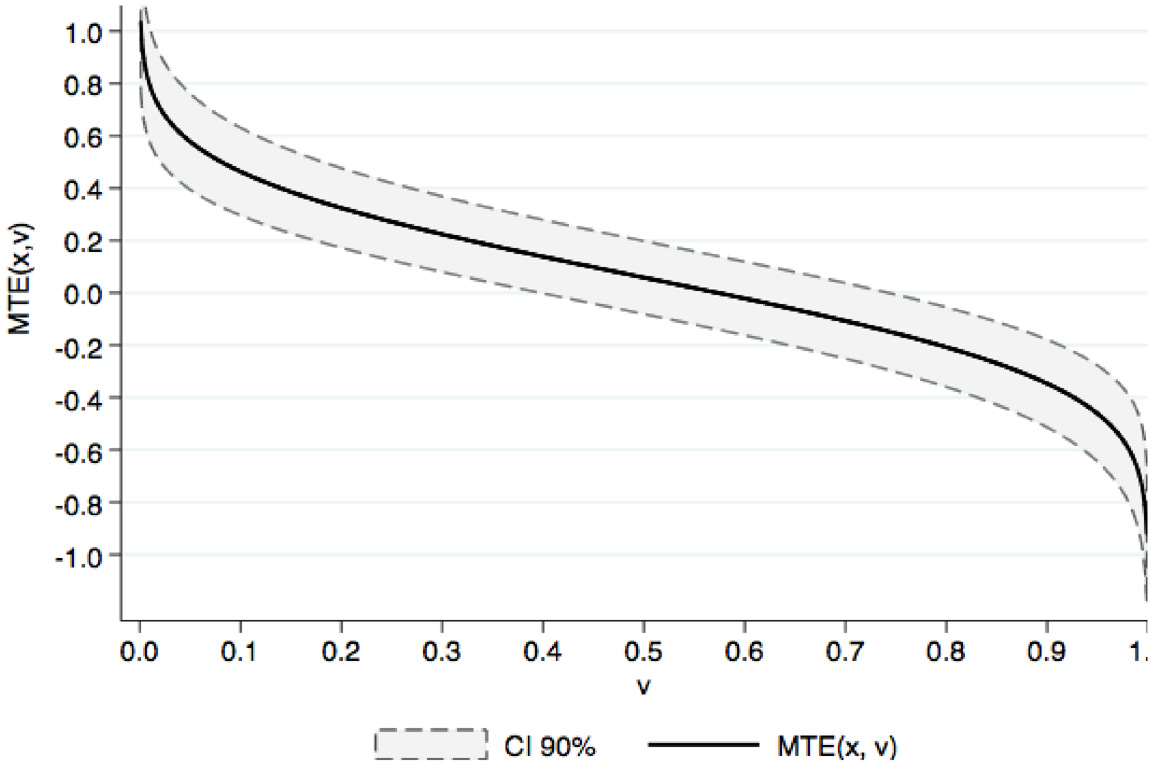


Figure 1.8: Estimated marginal treatment effects for log monthly earnings - without Moscow

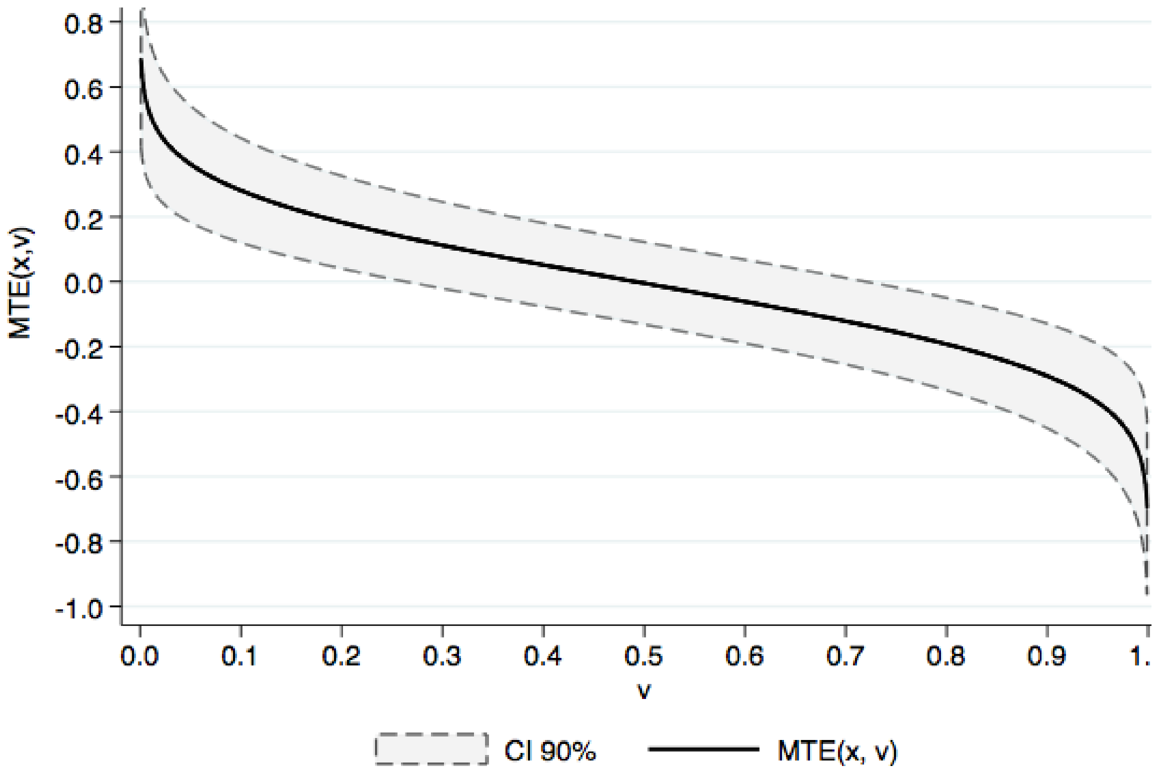


Figure 1.9: Estimated marginal treatment effects for log real hourly wage rate - average public sector wage

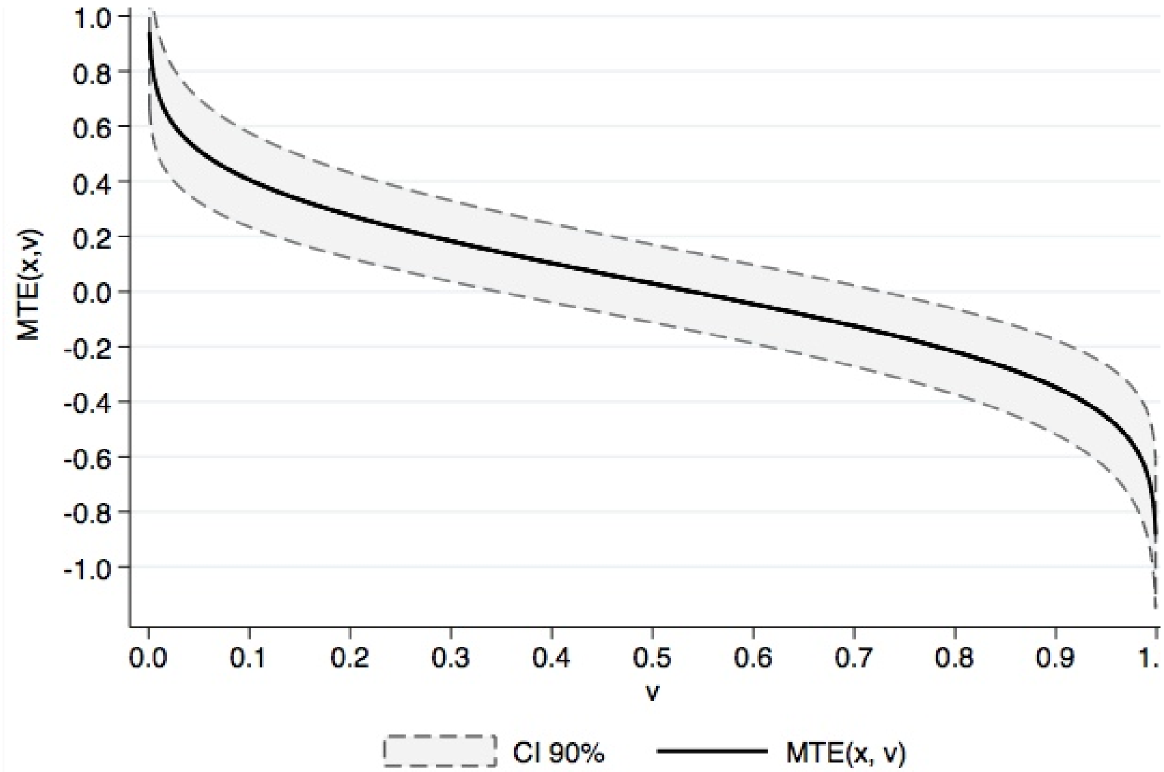
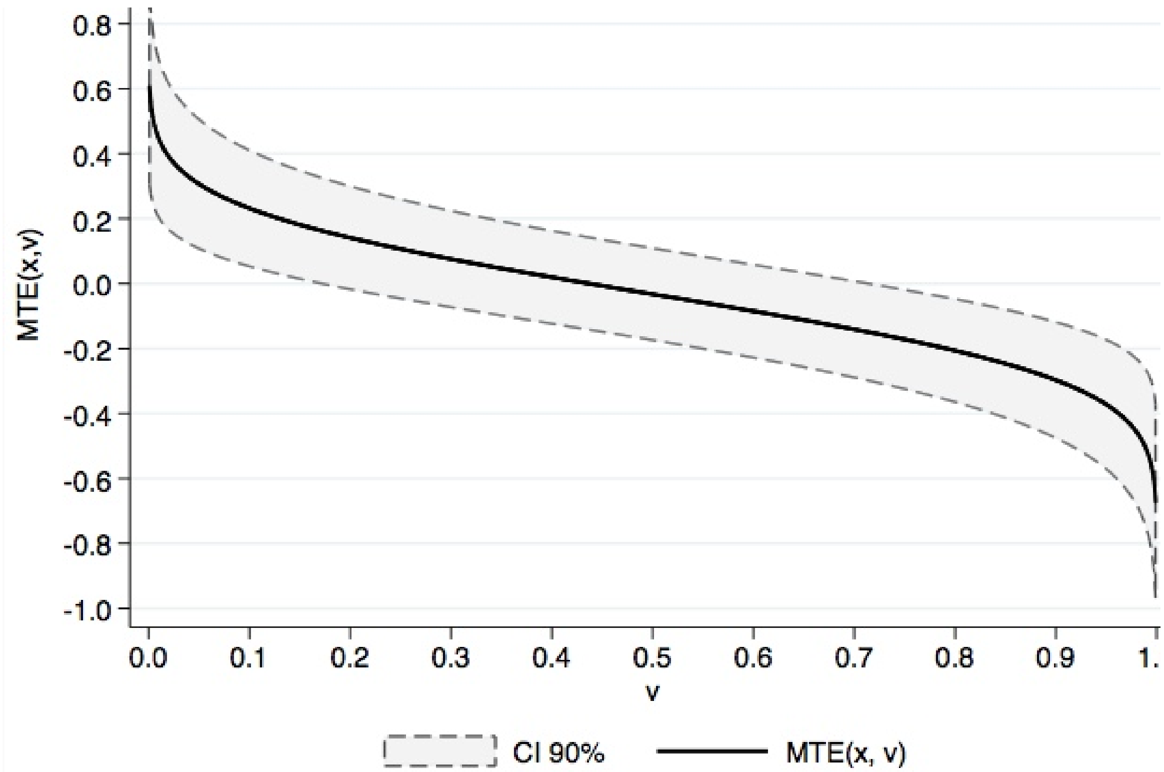


Figure 1.10: Estimated marginal treatment effects for log monthly earnings - average public sector wage



CHAPTER 2

HETEROGENEOUS RETURNS OF INFORMALITY: A COMPARATIVE STUDY BETWEEN BRAZIL AND RUSSIA

2.1 Introduction

Brazil, a country with one of the fastest growing economies during the last decade and part of famously-known BRIC countries, faces nowadays one of the highest informality rates in Latin America. Approximately 40% of the GDP and 35% of employees are informal. Therefore informality plays a big role in the local economy (Ulyssea, 2017). This is a common feature of developing countries, but Brazil stands out as a good case study given the large size of their economy and potential to be considered a developed country.

Although, between 2002 and 2010, informality was in fact decreasing and formal employment growing, currently there are more self-employed and informal employees than formal employees. Thus, the question about whether these workers are informal because they have to or because they want to still remains open. This paper aims to shed light on such issue by measuring the impact that informality have on wages and evaluating if workers choose to be informal voluntarily or as a last resource option. Additionally, a comparison between Brazil and Russia will be established in terms of the effect of labor regulation and the impact of informality on wages.

When it comes to informal jobs, these are considered to be precarious and of bad quality. Informal workers are also seen as individuals who are rationed out of the formal labor market and are stuck in informal jobs, because formal jobs are perceived as the desirable outcome. But when we look at the data on informality around the world, the picture is not quite like that. Many workers choose to be informal because they want to, not because they could not get a formal job (Maloney, 2004; Perry et al. 2007, Levy, 2007; Lehman and Muravyev, 2014). Of course this does not deny the existence of some informality as a result of segmentation, but it provides anecdotic evidence about the existence of voluntary informality.

As there are different degrees of informality, this paper will focus only on informality from the worker's side. A person is classified as informal if she does not have a signed worker's card when working as an employee or maid or if she does not have an official registration as a self-employed or entrepreneur (CNPJ). This restrictions, of course, leave outside the paper other forms of violation of the labor code such as paying to a worker less than the minimum wage or not providing full-time workers with vacation time or maternity leave, among other common forms of violation. But at the same time, this paper provides a good base estimate that helps us understand better the benefits and costs of having an informal job and what drives individuals into informality.

A sizable share of literature studies informality from the point of view of the firm in the Brazilian context and, in particular, the characteristics of informal firms, but research on the worker side is more limited as it is usually hard to estimate a partial equilibrium model in the absence of a natural experiment that introduces exogenous variation in the decision to be formal or informal. Comparative studies between countries are also rare, as they require detailed information about the labor codes of each of the countries of interest and data that allows the researcher to capture informality.

Thus, the contributions of this paper are twofold. First, it uses a unique data base on labor enforcement as part of the identification strategy, which includes information about number of labor inspectors in 2015, number of labor inspection offices in 2015, and other regional characteristics of the state that may affect enforcement. And second, it recovers the marginal treatment effect of formality on wage rate, which shows that although the ATE ranges from 0 to 22% depending on the sample used (males and females or only males), there is a lot of variation of such effect within the sample.

The paper is divided as follows. Section 2 reviews the state of the art in the literature about segmentation versus comparative advantage hypothesis and also about informality in Brazil, providing a good background about the local labor market and what strategies have other authors used. Section 3 describes very thoroughly the Brazilian labor market code, its legal implications when it comes to violation of the law about formality, and how it is enforced. Section 4 describes the data and shows descriptive statistics. Section 5 contains the empirical model we use and discusses the identification strategy. Section 6 includes results and Section 7 shows robustness checks that support our findings. And Section 7 concludes.

2.2 Literature Review

2.2.1 Segmentation versus Comparative Advantage

Perry et al. (2007) introduce the two views of informality that are predominant in labor markets. First, there exists the exclusion view or segmented labor markets hypothesis, in which informal workers are excluded from the formal labor market through entry barriers, such as minimum wages or tax laws, that restrict the access to formal jobs to only highly productive individuals. Thus, lower productivity workers are rationed out of the high tier labor market, where the good, formal jobs are available. Second, the comparative advantage hypothesis or voluntary informality view says that informal workers choose those jobs after considering the costs and benefits associated to each type of jobs. In this case, informality is an optimal choice for those workers. Our paper does not assume any prior about the origins of informality and tests empirically if the Brazilian labor market is segmented or not.

Botelho and Ponczek (2011) evalute the hypothesis of segmentation versus comparative advantage for the Brazilian case by using a fixed-effect framework as the authors do not have access to any valid instruments in order to account for selection into sectors. When applying this methodology, the authors use a rotating panel data for the main six metropolitan regions of the country in which households are interviewed continuously for four months, then left out of the sample of eight months, and reinterviewed again one last time for four more months. They find an average wage differential between formal and informal workers to be 7.8%, which they take as a small degree of segmentation. This paper fails to model selection into formality/informality, which, as Magnac (19991) established, is a key element when testing for segmentation. Besides that feature, they only include employees at a firm, which leaves out an important group of informal workers. We try to account of the shortcomings found in this paper by using a more robust estimation method and including self-employed and entrepreneurs.

Arias and Khamis (2008) approach is the closest to ours as they test for comparative advantage hypothesis and segmentation in labor markets in Argentina using marginal treatment effects. In their work, identification relies on the inclusion of the variable “workers intrinsic preference for working in a dependent relationship” in the selection equation and for the last group the exclusion restrictions were “the number of inspected workers at the province of residence” and “having the spouse of other relatives employed in the formal salaried sector”. The authors did not find any significant differences between the earnings of formal salaried workers and the self-employed individuals once they account for selection, which is consistent with the comparative advantage hypothesis. But when comparing formal and informal salaried workers, they do find that informal salaried work carries significant earnings penalties. So those results are more consistent with labor market segmentation. The main difference in what we propose comes from the identification strategy as we use regional variation in the degree of enforcement of the labor code as an exclusion restriction.

2.2.2 Regulation and Enforcement

Cardoso (2016) does an detailed review of the policies the Brazilian government has put in place in the last 20 years to fight informality. He explains the program Simples Nacional, which was created in 2006, and it is perceived as one of the two main programs targeting informal workers. In this program, taxes and bureaucratic procedures for small and medium businesses and micro-entrepreneurs are reduced. The program introduced an average tax reduction of 40% as it unified federal, state, and municipal level taxes into one single payment and made it easier to open a business.The second program, Microempreendedor Individual (MEI), was created in 2008, and it reduced taxes even more for micro-entrepreneurs who formalized their economic activity. Cardoso points out that the problem with public policy designed to combat informality is that it is often built upon incomplete diagnose of informality and its role in the structuring of the economic and social relations in a country, which is in line with what other Brazilian authors have found when using economic modeling.

Almeida and Carneiro (2012) study for Brazil the impact that labor inspections have on the size of informality at the municipality level. They find that enforcement has two contradictory effects: on one side, tighter enforcement may increase formal labor costs and drive workers to informality through a decrease in formal labor demand and less self-employment, but on the other side, workers may value all the benefits they get when firms comply with the labor code, so enforcement could make the formal sector more appealing if workers are interested in the benefits of formality. Their strategy to deal with the endogeneity of enforcement is to use regional variation in the degree of enforcement, represented by the ratio of log inspections per firm in the city as a source of identification, and control for distance measured in driving time between cities and the closest labor inspectors office, as they show that the inspections technology is very primitive and relies on the usage of cars to drive anywhere they get complaints in the state. So using distance to the closest inspection office seems to capture the cost of doing more inspections.

Our paper's approach is also close to Almeida and Carneiro (2012) as we also use inspectors data as a measure of enforcement, but the main differenc is that we use individual level data as one of the interests of this paper is to test for comparative advantage versus segmentation hypothesis. Additionally, as this paper also aims to recover heterogeneous effects, using individual level data is necessary in order to study how individual unobservable costs to be formal make returns to formality/informality vary between individuals with similar observable characteristics.

Viollaz (2016) analyzes how changes in the enforcement of labor regulations, measured through labor inspections, impact on the compliance level depending on the firm size. The author finds for Peru that firms can reduce their size to benefit from lower fines and less stringent regulations, so at the end there is little effect of better enforcement on the compliance level. The author uses an instrumental variables approach in order to account for the endogeneity of inspections. Her strategy proposes as a measure of the arrival cost of labor inspectors, measured as a combination of the the extension of the region road network (national plus regional roads) in kilometers divided by the region territory and the number of per capita crossing vehicles in the road network in every region by year. There are different definitions of informality used in this paper: if the worker is enrolled in the pension system, if the worker earns the minimum wage, if the worker works more than the maximum number of hours allowed by the law. The main differences with our approach is that our main interest is the impact of informality on wages, Additionally, we use a legalistic definition of informality and not a broader definition in which all the violations of the labor code are included as informality. Third, the institutional setting in Brazil is different than what happens in Peru, as the next section explains. In Brazil there is no different treatment for those who are caught violating the labor code. The administrative process is the same regardless of the number of employees in the firm or if the person is self-employed or works as a maid.

Meghir et al. (2015) use an equilibrium wage-posting model with heterogeneous firms and find evidence of compensating differentials when comparing informal firms' wage schemes to the wages paid by formal firms of equal productivity. Contrary to what Almeida and Carneiro (2012) showed, this paper finds that tightening enforcement does not increase unemployment and increases wages, total output and welfare by enabling better allocation of workers to higher productivity jobs and improving competition in the formal labor market. Their definition of informality considers as informal those employees who do not have a signed workers' card and those who are self-employed. This definition is different as the one proposed in this paper as they assume that all self-employed workers are informal and as we show in the next section that is not the case. Self-employed workers are mandated by law to be registered as self-employed and should contribute to social insurance, and many of them comply with the law. Additionally, their main focus was on the firm side and this paper targets workers.

Soares and Haanwinckel (2016) also develop a search and matching model of informal labor markets with worker and firm heterogeneity, intra-firm bargaining with imperfect substitutability across types of workers, and labor market regulation. Their model can replicate empirical facts observed in the data such as a reduction in informality among salaried workers of around 10 percentage points between 2003 and 2012, from an initial level of 30% while the minimum wage increased by 61% in real terms. The authors argue that the country experienced an important economic transformation during 2003-2012 based on a substantial increase in average years of schooling and TFP, which could have had their own equilibrium effects on informality. Thus their paper allows for heterogeneous labor supply and decreasing returns to scale in order to capture the previously mentioned features of the labor market. Their definition of informality only includes employees who do not have a signed workers' card.

On equilibrium, firms and workers self-select into the formal/informal sector as the compensating differentials theory predicts. Firms do not want to comply with labor regulation, but non-compliance is too costly for large firms as they can be caught. Workers want to receive employment benefits, but may be willing to accept informal jobs and leave unemployment for a sufficiently high wage. Minimum wages can also distort labor market allocations as if the minimum wage is binding for unskilled workers, they strictly prefer to have a formal job but are willing to accept an informal job in equilibrium in order to avoid unemployment. In this equilibrium, the formal wage premium decreases in the skill level, becoming negative for skilled individuals.

Ulyssea (2017) develops an equilibrium model where heterogeneous firms exploit two margins of informality: the extensive margin, in which firms do not register their business, and the intensive margin, in which firms hire workers “off the books”. The author uses Brazilian data to calibrate the model and finds that often firm and labor informality can move on different direction as a response to a unique policy to promote formality. For example, a policy such as reducing the firm’s entry cost to the formal sector, as Simples Nacional, induces firms to become formal, but then these newly created firms hire a large share of informal workers, so at the end there is zero effect on informal employment. On the other side, increasing enforcement of labor regulation reduces informality among workers but it increases informality among firms. Therefore it is very important to study the effect of policies on the extensive margin, but also on the intensive margin as the effects of apparently good policies can be counterproductive. Their definition of informality is only for employees and it includes those who do not have a signed workers’ card, in a similar fashion to what we use.

2.3 Labor Market Regulation and Enforcement in Brazil

2.3.1 Labor Market Regulation

The social security system in Brazil has three components: health, social insurance (previdência social), and social assistance. The health and social assistance components are not contributory as they are financed through general taxation, so all Brazilians have access to them. The social insurance part is mostly a contributory insurance. It includes benefits such as pension for those who reach the target age (60 years old for women and 65 for men) or those who reach the target number of years contributing to the system (30 years for women and 35 for men, regardless of their age), disability pension, death pension, and sickness and maternity benefits. Additionally, having a signed worker's card also give an additional benefit to the worker such as as extra wage payment at the end of the year.

Employees must contribute 8% to 11% of their monthly wage to the social insurance. On the other side, urban employers must contribute every month 20% of the wage paid to their employees and rural employers contribute 2.85% on their billed revenues. And last but not least, self-employed and own account workers earning a minimum wage must contribute 5% or 11% depending if they are covered by one of the special plans for low income individuals, and those who earn more than a minimum wage or are not classified into any of the special plans they must contribute 20% of their earnings (Instituto Nacional do Seguro Social, 2017).

In Brazil, every single individual who works as an employee in any economic activity or works as a maid must have a "Carteira de Trabalho" or workers' card, which is a document that guarantees that the worker has been hired formally and there is a registration about it in the workers roster and accounting books. If the individual is an employee, having a signed workers card guarantees access to all the social insurance benefits.

If the person does not work at a firm or does not have a signed workers card, they can also contribute to the social insurance as own-account workers and they can get access to the same benefits if they are not under one of the two special regimes for low income workers. The difference, though, is that in the latter the worker has contribute up to 20% of their income, and when employees have a signed card they only contribute 8% to 11% depending on their wage.

Additionally, self-employed individuals, entrepreneurs, and contractors (which usually operate as regular employees as it will be discussed later on) must be registered at the Cadastro Nacional de Pessoa Juridica (CNPJ), which is the national registry of entities that pay taxes and social insurance contributions. Not being registered at CNPJ when working as a self-employed or entrepreneur is illegal.

Thus, an employee who does not have a workers' card is considered informal in this paper, but we acknowledge that some individuals may get confused with the definition of "employee" and can probably think that contractors are employees, then we will use an alternative definition of formality that includes individual employees that are only registered at CNPJ.

2.3.2 Violations of the Labor Code

Informality can come in many different flavors in Brazil. In this paper, as discussed earlier, we will focus on a legalistic approach that uses a clear-cut definition for informality. An individual is classified as informal if this person works as an employee in a firm or as a maid and she does not have a signed workers card, or if the person declares to be self-employed or an entrepreneur but does not have a registration in the CNPJ, which basically means constituting a single-person firm.

Thus, such type of violations are very common as it is frequently found that employers hire workers and do not sign their workers' card to avoid paying their portion of the social insurance. This is especially true for maids, given that it is harder for labor inspectors to target houses in which there are maids working. Usually, when a maid is working under informal conditions, if the violation is caught is because the inspectors knew about the irregularity through the maid,who can self-report her poor working condition, but not through random inspections. And it is also true for self-employed individuals as it is hard to keep track of the economic activity of each one of the citizens of the country.

Another source of informality that we want to capture in this paper comes from apparently formal self-employed individuals. In this case, what happens is that firms hire workers under the figure of "contractors", which means that this new worker is actually not an employee of the company so she does not have a signed workers? card but a registration in the CNPJ. This version of hiring per se is not informal, but the problem is that these individuals work as employees, without distinction from regular employees. Thus firms hire them under this figure in order to avoid paying social insurance taxes as in this case, they worker assumes 100% of the cost of the social insurance contribution.

Last but not least, there are other violations to the labor code besides the status of the workers' card or the CNPJ. For example, with regard to the number of hours worked, which is set at 44 hours per week by the Federal Constitution of the country, many employees end up working more hours than the legally established. The minimum wage is also another source of violation as we can see in the data that there were a significant portion of the employees with earnings below the minimum wage (R$ 788 monthly or equivalent to USD$296, or R$3.58 per hour which is equivalent to USD$1.34, using the official exchange rate of January 1st 2015).

2.3.3 Enforcement

The Brazilian Constitution of 1988 established that the Ministry of Labor must hire labor inspectors (Auditores Fiscais do Trabalho - AFT) in order to execute and organize labor inspectors that guarantee the right to a safe job. Thus, the Secretary of Labor Inspections, which is an office within the Ministry of Labor, is in charge of establishing the guideline for labor inspections in Brazil, formulating social programs to protect workers, and promoting the enforcement and compliance of the labor code. Additionally, the Secretary of Labor Inspections created in 2013 the Escola Nacional da Inspeção do Trabalho, Enit, which is a government-sponsored technical institution that offers on-the-job training for labor inspectors (Enit, 2015).

Inspections take place under two scenarios: complaints from workers to the labor office or random inspections. Inspectors check the status of the workers cards to make sure they are properly signed, registration of the workers in the labor books of the company, and that workers are in a safe environment covered by all the laws included in the labor code.

Given that there is a shortage of inspectors in Brazil, then most of the visits are scheduled after a complaint. When the inspection is done, if the inspector found an actual violation, then an administrative process starts. As Figure 2.1 shows, when the administrative process starts, the employer or worker has 10 days in order to present her defense. Then a designed labor inspector checks the arguments presented by the defense in case there was one and decides the validity of the argument to rule if there should be a fine or not. If the infraction was found to be valid, then the defense has to pay a fine for it. If she pays in the following 10 days, there is a 50% discount in the amount of the fine (Cardoso and Lage, 2005). If the defense does not pay, she can appeal the fine and a new process starts again. If the person was found guilty of the violation, then she has to pay the fine without any discount. If the individual does not pay the fine, the federal government immediately registers this person into the the database of individuals who own money to the government and this action can have serious consequences such as a the person not being able to get a job in the public sector.

The magnitude of the fines related to not having a workers' card or not having a signed workers' card with an entry of the current job is around US$103.9 or equivalent to R$402.53, using the exchange rate of January 1st 2015. This amount of money doubles for every infraction that the inspectors find in a company or for every relapse.

2.4 Data Description

This section describes how the sample used in this study was created, describes the main variables used, and shows descriptive statistics.

This study uses data from Pesquisa Nacional por Amostra de Domicilios (PNAD) for 2015, which is a household survey with information about workforce indicators, migration, marital status and socio-economic characteristics. Additionally, it uses data from Instituto Brasileiro de Geografia e Estatística (IBGE) for regional indicators about GDP, number of firms, area of the states, among others. A full list of variables and its description can be found in the Appendix.

For 2015, 356,904 individuals were surveyed by PNAD. The sample of this study only includes individuals who are between 20 and 60 years old (156,529 observations were dropped), who do not have a job in agricultural activities as their main job (16,340 observations were dropped), and who get a salary or receive a payment in monetary terms for their work (58,878 observations were dropped). Additionally, we only include in the sample individuals who are currently working and can be classified as formal or informal (12,393 observations were dropped), who do not have missing values for their reported earnings (1,436 observations were dropped), who worked at least 20 hours in their main job if they claim to be formal employees or 5 hours if they are not employees (2,696 observations were dropped), who are not in the top and bottom 1% of the earnings distribution (1,759 observations), and who did not have missing covariates (1,676 observations were dropped). The final sample has 105,197 unique observations.

Formality is defined as an employee or maid who has a signed worker's card or a self-employed or entrepreneur who is registered at the CNPJ. On the other side, informality is defined as a maid or employee who does not have a workers card and a self-employed or entrepreneur who is not registered at the CNPJ. Under this definition, "contractors", who actually work as if they were employees of a firm but instead of having a signed worker's card they have CNPJ (which is cheaper to pay for both the employer and the employee), are classified as informal. We will use a a more flexible definition of formality that includes those contractors as formal workers as a robustness check.

2.4.1 Descriptive Statistics

The informality rate in the sample is 37.9% as it can be seen in Table 2.1, which is slightly lower than the informality rate at the country level that was 45% (IPEA). When comparing these numbers with the informality rate in other countries of Latin America, Brazil has an average rate for the region. But if we compare it with the informality rate in developed countries or even with Russia, then Brazil has an informality rate, on average, 20-30 percentage points higher.

Work categories used in this paper only rely on the information provided by respondents when asked about their main job. The main job was defined as the work activity in which the individual spent most of her time during the reference year.

In this regard, 64% of the sampled individuals are self-classified as employees, 23.3% are self-employed, 8.5% work as maids, and 4% are entrepreneurs. In order to be classified as an entrepreneur, the person has to have at least one employee working for themselves.

In general, individuals with formal and informal jobs have different demographic characteristics as it can be seen in Table 2.2. There is slightly more women working informally than men and informal workers tend to be 2 years older, on average. There's more married people in the formal sector than in the informal sector, which also happens in Russia and could be due to marital sorting, but we do not explore that feature of the data in this paper.

Education differences are very important as formal workers tend to be more educated than informal workers. The biggest difference comes from the percentage of individuals who only have primary school or less, which is tremendously different between the two groups. Racial differences also play an important role in Brazil. Those who self-classified as "White" work mostly as formal workers, but afro-brazilian workers have informal jobs in higher proportion.

As we excluded from the sample those who work in agricultural jobs, then the sample over-represents urban workers, as we excluded the main source of employment in rural areas. Regional differences are also important in Brazil as most of the largest economic centers are located in the Southeast region, such as Sao Paulo and Rio de Janeiro. The North and Northeast region are traditionally poor and they have the largest shares of minority groups.

2.4.2 Labor Market Variables

Wages in Brazil are, on average, low and exhibit very high variance. For 2015, the legal monthly minimum wage was established at R$788 or US$199. This wage applies to workers who work 44 hours per week.

In the sample, the average monthly earnings were R$1,586. But as we can see in Table 2.3, formal workers have monthly earnings that are 40% higher than those of informal workers. There is a mass concentration of individuals around the minimum wage cut-off (vertical line in Figure 2.1). Earnings also differ greatly by educational level, race, and state. For example, self-reported Asian individuals earn, on average, R$2,814, but Afro-Brazilain individuals earn R$1,297 monthly.

Weekly hours worked are higher for formal workers, but informal workers have higher variance. Formal workers report to work 43 hours per week and informal workers report 37 hours per week. It is important to keep in mind that during the data cleaning process, employees that reported working less than 20 hours per week were dropped and self-employed and entrepreneurs who reported working less than 5 hours per week were also dropped.

The wage rate, which is our variable of interest, is also different between the two groups. Formal workers have a higher wage rate on average, which is roughly equivalent to US$2.98 per hour.

Participation rate in Brazil in 2012 was 63.7%, on average, but after the economy started to deteriorate, the participation rate when down 9 percentage points to 56.8%. Differences in participation rate between men and women diminished between 2012 and 2016 (Table 2.4).

Unemployment rate previous to 2015 was in the single-digit units and stable, but since 2015 it started climbing as a result of the economic crisis the country has been experiencing in the past few years. In 2016, the unemployment rate was 12%. (Table 2.5).

2.4.3 Labor regulation variables

Informal individuals tend to be more concentrated in states that are less developed, with less firms, and less inspection offices and inspectors (Table 2.6). On average, 8% of the firms in a state are inspected, but not all these inspections are random, as the inspectors usually only focus on firms that they have gotten complains on.

For 2015, there were 2,466 labor inspectors for the whole country, distributed among the 26 states and the federal district, Brasilia. On average, states have 91 inspectors and 5 offices, but these results are skewed by the presence of he rich states from the Southeast region (Sao Paulo, Minas Gerais, and Rio de Janeiro), in which there were 1,012 inspectors. This is expected to happen as it is in these states in which most of the economic activity of the country happens. Sao Paulo state alone produces more than 30% of the GDP of Brazil.

Additionally, for that year, there were inspected 249.649 firms in the country and 20.286 administrative processes were started after finding violations. As explained before, if the process goes through its legal course and the employer is found guilty, then a fine of US$103.9 is charged per each unregistered worker.

2.5 Econometric Framework

This section introduces the empirical methodology used in the paper. First, it briefly discusses why traditional OLS methods are not appropriate in this scenario. Then, it introduces the marginal treatment effect model used for estimation and discusses the requirements for having identification of the parameters of interest under the MTE model.

2.5.1 OLS

Under OLS, the estimation of the returns of informality, λ, would be unbiased only if informality is not correlated with the error term, ϵ, conditional on X,

$$Y = X'\beta + \lambda D + \epsilon, \tag{2.1}$$

where Y is log wage rate, X are exogenous covariates, D is a binary variable that takes the value of 1 if the individual is formal and 0 if informal and ϵ is an error term. But if informality is not randomly assigned and it depends on the characteristics of the individuals, then the self-selection process should be modeled as the coefficient of interest, λ in this case, is biased as it suffers from "selection bias". Thus, the selection process can be represented by the following equation:

$$D = Z'\gamma + \nu. \tag{2.2}$$

Therefore, we need to take into consideration the selection process into formality and correct for it the outcome equation in order to recover a consistent estimator in the presence of selection. Additionally, if the returns to informality vary based on observable and unobservable characteristics of the individual, as it was stated in the introduction, then traditional selection methods will not suffice as it is important to capture this attribute of the data in the empirical model by recovering not only mean effects, but the whole distribution of the effect of informality on wage rate.

As Heckman and Vytlacil (1999) show that self-selection may arise in two forms: selection based on heterogeneous background and characteristics, which is the "selection bias" problem, and also the ?selection on gains" problem, which is when the people who select into formality are the ones who expect the highest gains from it, so the returns of the treatment are not the same for similar individuals.

2.5.2 Marginal Treatment Effects Model

Let Y be the observed outcome of interest, the log real wage rate at main job. Assume that there are two types of occupations indexed by two labor market sectors: formal (treated state) and informal (untreated state). Let D represent the binary treatment of interest: being formal. Define Y_1 as the potential outcome of an individual in the treated state (D=1), and define Y_0 as the potential outcome of an individual in the untreated state (D=0), such that Y_1 represents the potential wage rate of an individual who works formally, and Y_1 represents the potential wage rate of someone who works informally.

This gives rise to a switching model that can be expressed as the following:

$$Y = (1 - D)Y_0 + DY_1. \tag{2.3}$$

Following Carneiro et al. (2011), this estimation method is based on the generalized Roy Model of occupation choice.The decision rule of an individual i to work formally or informally is characterized by a latent variable model (Willis and Rosen, 1979):

$$Y_1 = X'\beta_1 + U_1 \tag{2.4}$$

$$Y_0 = X'\beta_0 + U_0, \tag{2.5}$$

where X contains sociodemographic characteristics such as schooling, age, parents education, and regional controls.

The decision rule of an individual i for choosing betweeb a formal or an informal job can be characterized by a latent variable model (Willis and Rosen, 1979):

$$D = 1(D^* > 0), \tag{2.6}$$

where $D^* = Z\gamma - V$ and D equals one for individuals who work formally and zero for individuals who work informally. V represents the unobserved marginal cost of being formal. Notice that as V is a cost, it could be interpreted as the cost of having a less flexible job or being in a dependent working relationship when the individual has strong entrepreneurial skills, among others.

Notice that (X, Z) observed, but (U_0, U_1, V) is not. Therefore, we need some assumptions on the unobserved parameters in order to make the model tractable. We assume that V is a continuous random variable with a strictly increasing distribution function F_V and (U_0, U_1, V) is statistically independent of Z given X. Z is a vector that contains observable individual and family characteristics that affect the decision to work formally or informally and it also includes exclusion restrictions, which affect the decision of being formal but not earnings directly. The inclusion of these variables is what allows us to get identification.

Therefore, the decision rule can be written as:

$$D = 1(Z'\gamma > V). \tag{2.7}$$

Let $P(Z)$ denote the probability of work formally (D=1) conditional on Z=z, such that $P(Z) = Pr(D = 1|Z = z) = F_V(\mu_D(Z))$. We keep conditioning on X, but to make notation easier it is omitted from now on. Now define $U_P = F_V(V)$, which is uniformly distributed by construction. This transformation is useful because different values of U_P correspond to different quintiles of V.

Rewriting Equation 2.7 using the transformation of the error term and P(Z), we get:

$$D = 1(P(Z) > U_P). \tag{2.8}$$

Now we can rewrite Equation 2.3 as:

$$\begin{aligned} Y = (1-D)Y_0 + DY_1 &= D(\mu_1(X) + U_1) + (1-D)(\mu_0(X) + U_0) \\ &= D(X'\beta_1 + U_1) + (1-D)(X'\beta_0 + U_0) \\ &= X'\beta_0 + D((X'\beta_1 - X'\beta_0) + D(U_1 - U_0) + U_0 \end{aligned} \tag{2.9}$$

Assuming that $\mu_1(X)$ and $\mu_0(X)$ also have a linear representation such that $\mu_j(X) = X\beta_j$. The conditional expectation of Y given X=x and P(Z)=p is:

$$\begin{aligned} &E(Y|X = x, P(Z) = p) \\ &= E(Y_0|X = x, P(Z) = p) + E(Y_1 - Y_0|X = x, D = 1, P(Z) = p)p \\ &= X'\beta_0 + (X'\beta_1 - X'\beta_0)p + \int_0^p E[(U_1 - U_0)|X = x, U_s = u_s]du_s \end{aligned} \tag{2.10}$$

In order to estimate (2.10) we need to consider three cases. As Belskaya, Peter and Posso (2014) explain, the potential results could be: (i) if the unobserved terms are homogeneous, that is $U_0 = U_1 = \bar{U}$ for all individuals, then the last term of Equation 6 cancels out; (ii) the unobserved terms are heterogeneous but mean independent of high-school decisions, that is $E(U_1 - U_0|X = x, U_s = u_s) = E(U_1 - U_0)$, then the last term of Equation 2.10 cancels out; and (iii) if the unobserved terms are heterogeneous and correlated with V (the error term from the selection equation), then the last term of Equation 2.10 cannot be ignored, because it reflects "selection on gains".

Therefore, if in this framework we were going to use a classic instrumental variables approach, we would assume that individuals do not sort into formal jobs based on their expected gains of having a job of such type. This is yet to be proven because it may be the case that individuals who know they have a preference for jobs without a boss or in which they can control their time, for example, want to have informal jobs or being self-employed. This is called selection on gains: given that the returns to job type are heterogenous across individuals, those who will benefit the most from being formal or informal are more likely to select into that type of job.

- Normal Switching Regression Model

  Under the potential outcomes framework defined by equations 2.4-2.6, the switching regression model assumes that the error terms of the three equations follow a multivariate normal distribution, such as $(U_0, U_1, V) \sim N(0, \Sigma)$, and (U_0, U_1, V) is independent from (X, Z). The variance of V is normalized to 1, such that $\sigma_V^2 = 1$, and the covariance between U_0 and U_1 cannot be recovered given that we never observe both outcomes simultaneously. Therefore σ_{10} is not identified. The variance-covariance matrix in this case is:

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{10} & \sigma_{0V} \\ \sigma_{10} & \sigma_1^2 & \sigma_{1V} \\ \sigma_{0V} & \sigma_{1V} & 1 \end{pmatrix}$$

  Following Lokshin and Sajai (2004), the model can be efficiently estimated by using the full-information Maximum Likelihood method to jointly estimate both the outcome equation and the decision rule. The loglikelihood function of the model in this case would be:

$$\begin{aligned} ln(L) = \sum_i (D\omega_i[ln(F(\eta_{1i})) + ln(\frac{f(\frac{U_1}{\sigma_1})}{\sigma_1})] + \\ (1-D)\omega_i[ln(1-F(\eta_{0i})) + ln\left\{\frac{f(\frac{U_0}{\sigma_0})}{\sigma_0}\right\}]) \end{aligned} \tag{2.11}$$

  Where:

  F: Cumulative normal distribution

  f: Normal density distribution

ω_i: Optional weighting for observation i

$\eta_{ji} = \frac{Z\gamma+\rho_j(\frac{U_j}{\sigma_j})}{\sqrt{1-\rho_j^2}}$ for j=0,1 and $\rho_j = \frac{\sigma_{jV}^2}{\sigma_V \sigma_j}$ are the correlation coefficients.

In order to estimate (2.11), we need a transformation of the correlation coefficients and standard deviations to guarantee that the correlation is between -1 and 1 and the standard deviation is always positive. This is done in a way that it is easy to recover the true parameters of the model. For the case of the standard deviations, $ln(\sigma_j)$ is used instead of using σ_j. For the correlations, the Fischer's transformation is the standard: $atanh(\rho_j) = \frac{1}{2}(\frac{(1+\rho_j)}{(1-\rho_j)})$.

- Marginal Treatment Effects Model

  The MTE methodology does not assume that the returns of formality are the same for everyone, therefore it accounts for selection on gains.

  Following Carneiro et al. (2011) , this model assumes that agents know the gross return on earnings of having each type of job. This means that individuals know $\Delta = Y_1 - Y_0 = (X'\beta_1 - X'\beta_0) + (U_1 - U_0)$ per each i.

  In the third case analyzed before what is happening is that individuals who are identical on their set of X's may make different decisions about which type of employment to get, influenced by their unobserved component V in the selection equation. As a result of this feature, the returns of working formally or informally on wages, for observationally identical individuals, will depend upon a constant component $(X'\beta_1 - X'\beta_0)$ and an individual-specific component $E(U_1 - U_0|X = x, U_s = u_s)$.

  If we differentiate Equation 2.10 with respect to p, we get the MTE:

$$MTE(x,p) = \frac{(\partial E(Y|X = x, P(Z) = p))}{\partial p} = (X'\beta_1 - X'\beta_0) + E(U_1 - U_0|X = x, U_s = u_s). \tag{2.12}$$

  The last term of Equation (2.12) can be estimated in a parametric version and in a semi-parametric version, both versions can be estimated using polynomials of different orders or not. For this version of the paper, I will use a parametric approach.

Parametric Normal Model

Under the parametric framework, the model has the same set of assumptions as the Normal Switching Regression model explained before. Using the same multivariate normal parameterization, Equation (8) can be expressed as:

$$\begin{aligned} MTE(x, u_s) &= X'(\beta_1 - \beta_0) + E(U_1 - U_0 | U_s = u_s) \\ &= X'(\beta_1 - \beta_0) + E(U_1 - U_0 | V = \Phi^{-1}(U_S)) \\ &= X'(\beta_1 - \beta_0) + (\sigma_{1V} - \sigma_{0V})\Phi^{-1}(U_S) \end{aligned} \tag{2.13}$$

The parameters $(\beta_1, \beta_0, \sigma_{1V}, \sigma_{0V})$ and their standard errors can be estimated by maximum likelihood methods. The most common ways of estimating this model under normality assumptions in order to recover the parameters of interest are: (i) Following Lokshin and Sajai (2004), who specified the loglikelihood function that we already showed or (ii) Following Maddala (1983), who proposes a linear regression model augmented by a binary endogenous treatment variable and assumes that $\beta_1 = \beta_0$ and $\sigma_0^2 = \sigma_1^2$. This paper follows Lokshin and Sajai (2004) approach given that it imposes less restrictions on the model.

2.5.3 Hypotheses to be tested

Following Magnac (1991), there are two hypotheses that can be tested. On one side, the segmented labor markets hypothesis claims that access to the formal labor market is restricted by minimum wages, tax laws, and other labor regulations, thus lower productivity workers are rationed out of the formal sector and can only find jobs in the informal sector. If true, we should observe that: $cov(U_1, V) < 0$ and $cov(U_0, V) < 0$. On the other side, the comparative advantage hypothesis says that informal jobs reflect workers' implicit choices given their preferences, skills, the cost and benefits of formality, and the availability of other means of social protection (Perry et al. 2007). If true, we should observe that: $cov(U_1, V) < 0$ and $cov(U_0, V) > 0$.

2.5.4 Identification

We would like to recover $(\beta_1, \beta_0, \sigma_{1V}, \sigma_{0V})$ as those are the coefficients of interest of our model, given that we want to estimate the returns to formality on wages and study if the Brazilian labor market is segmented or competitive, which is captured through the covariances.

In order to uniquely estimate the effect of formality on wages, the selection equation should contain at least one variable that is not included in the outcome equation, that affects the decision to be formal or not but does not affect wage rate directly besides its effect through the decision rule. As most of the determinants of formality status are also determinants of wages, it is not easy to find valid exclusion restrictions outside of the context of a natural experiment. For example, ideally, we will have a lottery in which some individuals are randomly assigned to formal jobs and other individuals are randomly assigned to informal jobs. This way we will recover the causal effect of informality on wages and not a correlation that could be biased if we believe that there are unobservable characteristics like preferences for independent work or for a flexible schedule, that may determine if someone wants an informal job but also their wage.

As an attempt to find a valid exclusion restriction in a non-experimental setting, we use data on labor inspections provided by the Ministry of Labor and geographic information on state's size as exclusion restrictions. This strategy follows what Almeida and Carneiro (2012) did as it aims to capture the technology of inspections previously described and account for the cost of having to perform inspections in areas that are farther located as it takes more driving time that can be used doing other inspections in closer locations.

The exclusion restrictions used are the log number of inspectors per state, the log urban area of the state in squared kilometers, and an interaction term between the number of inspectors per office at the state level and the urban area measure times 10,000. We argue that the exclusion restrictions proposed in this paper are valid as state size is predetermined, so it cannot be changed as a response to labor market outcomes, and the concern about the number of state inspectors comes from the fact that it may be correlated with wealthiness of the state, therefore we control for state GDP and regional fixed effects in order to capture that. Additionally, we estimate the correlation between state size and state GDP to dissipate any issues about the correlation between them, and we find that it is 0.41, which is at the limit of what can be considered low correlation (Figure 2.3).

2.6 Results

This section presents the results for the OLS and MTE models and it discusses the implications of such results.

2.6.1 OLS

OLS coefficients can be interpreted as a biased Average Treatment on the Treated (ATT), as $OLS = ATT + E[Y_0|D = 1] - E[Y_0|D = 0] = ATT + Selection\ \ Bias$. The results presented in Table 2.8 suggest that formal workers earn, on average, wage rates that are 12.1% higher than informal workers. In this case, we suspect of negative selection bias as the higher the unobserved cost of formality, the less likely a person is going to work formally. Other coefficients in the regression should be interpreted with caution as they could be biased. Their purpose on the model is to help as controls not as the coefficients of interest.

2.6.2 Marginal Treatment Effects

Based on the hypothesis presented by Magnac (1991), informality in Brazil responds to comparative advantage as workers are selecting themselves into the sector their skills are going to be better rewarded. Table 2.9 and Table 2.10 show the selection equation and wage equation that were jointly estimated in order to compute the marginal treatment effect model. Figure 2.4 shows the full distribution of the MTE over the domain of the unobserved cost of being formal (U). The graph shows the effect of formality when we compare a formal individual against an individual who is indifferent between formality and informality given their unobserved non-pecuniary costs, U.

The ATE of formality is 0.219, which means that formal workers earn, on average, wage rates that are 22% higher than informal workers, but the result is not significant at 1%. The covariances between the wage equations the selection equation are $cov(U_1, V) = -0.6 < 0$ and $cov(U_0, V) = 0.4 > 0$, both are significant at 1%. This confirms the comparative advantage hypothesis as formal workers are the ones who have lower cost of being formal (lower U) and informal workers workers are the ones who have the highest cost of being formal (higher U). Additionally, as $cov(U_1, V) - cov(U_0, V) < 0$, this means that there is selection on gains, as those with the highest gains from the treatment are the ones who are more likely to be formal.

The exclusion restrictions are highly significant. The higher the number of inspectors in the state, the higher the likelihood of being formal. The interaction term between urban area of the state and number of inspectors is negative, reflecting the fact that when we keep the number of inspectors constant and increase the distance, then the likelihood of being informal is higher.

2.7 Robustness Checks

This section includes different robustness checks that support the findings of the previous section.

2.7.1 Marginal Treatment Effects for Only Males

As we may have been concerned about selection into employment, especially for the female labor force, given that their participation in the labor market is significantly lower than the one by males, we estimated the same model but for a reduced sample of only men in primer age. This sample has 59,218 observations.

The results, included in Tables 2.11 and 2.12, are consistent with what we previously found that the ATE is positive but not significant. The covariances have the expected signs that support a comparative advantage hypothesis and the slope of the MTE curve is negative.

2.7.2 Marginal Treatment Effects using only one exclusion restriction: Log number of state inspectors

Tables 2.13 and 2.14 include this specification. The results are consistent with the comparative advantage hypothesis, as they all have the expected signs and significance level. Now the ATE is 0.22 and it is significant at the 1%, which could hint that formal workers get paid a premium.

2.7.3 Marginal Treatment Effect Model using an alternative definition of formality: counting as formal those employees who do not have a worker's card but only CNPJ.

The results of this specification are included in Tables 2.15 and 2.16. There is also evidence of a comparative advantage. The ATE is slightly higher than what was estimated before but it is not significantly different from zero. The exclusion restrictions are highly significant and with the expected signs.

Conclusion

The results found strongly support that informality in Brazil responds to a comparative advantage hypothesis as it is the case in Russia. Thus meaning that workers self-select into the sector in which their skills are going to be better rewarded. On average, the ATE is not statistically different from zero, which implies that formal workers do not have a premium, but there are heterogeneous effects as workers with lower non-pecuniary costs associated to formality do earn very high premiums and workers with very high non-pecuniary costs of formality, actually get hurt by being formal as they could have earned more working informally.

These results could be driven by the recent efforts of the government to formalize historically informal sectors such as household employment (maids, gardeners, baby sitters, among others), as now these employees are contributing to social security and getting the benefits of it, but their wage rates are so low to start with that they end up having lower earnings than before.

Additionally, it is important to notice the effect of government policies such as Simples, which is a program that simplified the tax code for entrepreneurs and self-employed individuals who have revenues below certain threshold, in order to give them incentives to become formal by only acquiring the CNPJ and paying a unified social security tax is lower than what they should contribute. This program has contributed to the reduction of informality from self-employed and small entrepreneurs, but has contributed to increase informality in other ways. First, it promotes the informal hiring of employees as the firm now hires a consultant/free-lance who claims to be self-employed but the person is actually an employee as she is not hired formally with a signed worker's card but only with CNPJ. Two, it provided incentives for entrepreneurs to become formal, so it reduced informality in the extensive margin, but this newly formal employer does not have incentives to hire a formal worker, but an informal one. So it does not reduce informality in the intensive margin.

Table 2.1: Formal and informal workers sample

Worker type	Formal	Informal	Total
Maid	3,153	5,844	8,997
Employee	52,885	14,509	67,394
Entrepreneur	3,559	670	4,229
Self-employed	5,634	18,943	24,577
Total	65,231	39,966	105,197

Table 2.2: Descriptive Statistics I

Variable	Formal	Informal	t-test
Female=1	0.424	0.457	10.36
Age at the time of survey	36.7	38.3	23.88
	(10.7)	(10.3)	
Married=1	0.639	0.559	-25.73
Schooling level			
Primary school or less=1	0.256	0.447	65.3
High school=1	0.477	0.384	-29.53
College=1	0.255	0.162	-35.5
Grad school=1	0.01	0.005	-8.88
Race			
White	0.474	0.359	-36.85
Afrobrazilian	0.103	0.116	6.29
Asian	0.004	0.002	-5.68
Mixed	0.414	0.518	33.05
Indigenous	0.002	0.003	3.01
Urban=1	0.958	0.91	-31.71
Region			
North	0.101	0.189	40.78
Northeast	0.212	0.309	35.4
South	0.21	0.121	
Southeast	0.36	0.271	-30.23
Center	0.122	0.142	9.09
Population at state	15,617,474	12,772,761	-35.09
	(13,443,962)	(11,561,795)	
GDP per capita	30,374.984	26,151.085	-52.55
	(12,811.836)	(12,384.891)	

Note: For reference purposes, the exchange rate at December 31 2015 was USD$1 to R$3.96.

Mean and standard deviation in parenthesis.

Table 2.3: Descriptive Statistics II

Variable	Formal	Informal	t-test
Wage rate	11.6	9.27	-34.13
	(11.35)	(9.65)	
Labor earnings per month	1813.16	1217.06	-68.7
	(1499.22)	(1,114.09)	
Weekly hours worked	42.92	37.33	-85.64
	(7.96)	(13.2)	
Average monthly earnings by state	1,833.32	1,699.13	-63.06
	(338.64)	(338.64)	

Note: For reference purposes, the exchange rate at December 31 2015 was USD$1 to R$3.96.

Mean and standard deviation in parenthesis.

Table 2.4: Labor Force Participation rate for 2012-2016 (%)

Year	Total	Women	Men
2012	63.7	52.5	76
2013	59.4	48.3	71.7
2014	59.2	48.3	71.2
2015	58.5	47.8	70.3
2016	56.8	46.7	67.8

Source: IBGE.

Table 2.5: Unemployment rate for 2011-2016

Year	Unemployment rate
2011	6.0%
2012	7,4%
2013	4,1%
2014	6,8%
2015	8,5%
2016	12%

Source: IBGE.

Table 2.6: Labor Enforcement Statistics

Variable	Formal	Informal	t-test
Number of inspectors per state	177.79	144.36	-41.92
	(130.45)	(117.04)	
Number of inspection offices in the state	11.22	8.93	-42.73
	(8.71)	(7.97)	
Number of firms in the state	450,253.85	325,490.04	-42.69
	(485,275.49)	(415,672.14)	
Number of inspected firms	16,965.19	13,748.68	-42.53
	(12,207.505)	(11,394.87)	
Urban area in squared km to inspectors	905.76	847.56	-29.04
	(310.49)	(323.23)	
Urban area in squared km	1,720.23	1,319.43	-42.99
	(1,538.76)	(1,343.07)	

Source: Ministry of Labor and Employment. Author's calculations.

Standard deviation in parenthesis.

Table 2.7: OLS Results

Variables		
Dep. Var: Wage rate	Coef.	SE
Formality=1	0.123***	
	(0.004)	
Female=1	-0.247***	
	(0.003)	
Age at the time of survey	0.058***	
	(0.001)	
Age squared	-0.001***	
	(0.000)	
Married=1	0.037***	
	(0.004)	
Schooling level (Primary schoor or less=base)		
High school=1	0.201***	
	(0.004)	
College=1	0.688***	
	(0.006)	
Grad school=1	1.290***	
	(0.021)	
Race (white=base)		
Indigenous=1	-0.089***	
	(0.033)	
Afrobrazilian=1	-0.118***	
	(0.006)	
Asian=1	0.142***	
	(0.031)	
Mixed=1	-0.099***	
Urban=1	0.114***	
	(0.007)	
Log state population	-0.017**	
	(0.009)	
Log state urban area in sq. km	0.005	
	(0.010)	
Constant	0.886***	
	(0.093)	
No. of Observations	105,197	

This table shows basic OLS results for the wage equation, in which “Works formally” is included as an independent variable.
Dependent variable: Log real hourly wage rate.
Robust standard errors in parentheses.
Region dummies are included.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 2.8: Selection Equation: Probit model

Variables	Coef./SE
Female=1	-0.212***
	(0.008)
Age at the time of survey	0.057***
	(0.003)
Age squared	-0.001***
	(0.000)
Married=1	0.113***
	(0.008)
Schooling level (Primary schoor or less=base)	
High school=1	0.450***
	(0.009)
College=1	0.744***
	(0.012)
Grad school=1	0.905***
	(0.044)
Race (white=base)	
Indigenous=1	-0.124*
	(0.068)
Afrobrazilian=1	-0.125***
	(0.013)
Asian=1	0.224***
	(0.063)
Mixed=1	-0.113***
	(0.009)
Urban=1	0.332***
	(0.016)
Log number of state inspectors	0.076***
	(0.011)
Log state urban area in sq. km	-0.062***
	(0.010)
Interaction	-0.000
	(0.001)
Constant	-1.148***
	(0.070)
Number of observations	105,197

This table shows the Probit model for the decision of being formal, in which "Works formally" is the dependent variable. The probit model is jointly estimated with the outcome equation.
Dependent variable: Works formally (=1).
Robust standard errors in parentheses.
Region dummies are included.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table 2.9: Wage equation

Variables	**Treated [D=1]**	**Untreated [D=0]**
Female=1	-0.302***	-0.151***
	(0.005)	(0.007)
Age at the time of survey	0.065***	0.040***
	(0.002)	(0.002)
Age squared	-0.001***	-0.000***
	(0.000)	(0.000)
Married=1	0.079***	-0.008
	(0.005)	(0.007)
Schooling level (Primary schoor or less=base)		
High school=1	0.353***	0.048***
	(0.006)	(0.010)
College=1	0.827***	0.405***
	(0.007)	(0.015)
Grad school=1	1.465***	0.884***
	(0.024)	(0.046)
Race (white=base)		
Indigenous=1	-0.122***	-0.065
	(0.044)	(0.056)
Afrobrazilian=1	-0.148***	-0.047***
	(0.008)	(0.012)
Asian=1	0.162***	0.097
	(0.035)	(0.064)
Mixed=1	-0.123***	-0.053***
	(0.005)	(0.008)
Urban=1	0.213***	0.033***
	(0.011)	(0.013)
Log GDP Per Capita	0.156***	0.155***
	(0.008)	(0.016)
Constant	-1.480***	-0.953***
Sigma	-0.641***	0.483***
	(0.003)	(0.018)
Sigma1V-Sigma0V	-1.124***	
	(0.019)	
ATE	0.219	
	(0.204)	
Number of observations	105,197	

This table shows the normal switching regression results for the outcome equation.
Dependent variable: Log real hourly wage rate.
Bootstrap standard errors 50 reps.
Region dummies are included.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 2.10: Selection Equation: Probit model - Only Men

Variables	Coef./SE
Dep. Var: Being formal	
Age at the time of survey	0.064***
	(0.004)
Age squared	-0.001***
	(0.000)
Married=1	0.149***
	(0.011)
Schooling level (Primary schoor or less=base)	
High school=1	0.464***
	(0.012)
College=1	0.707***
	(0.016)
Grad school=1	0.823***
	(0.066)
Race (white=base)	
Indigenous=1	-0.071
	(0.091)
Afrobrazilian=1	-0.105***
	(0.018)
Asian=1	0.185**
	(0.087)
Mixed=1	-0.086***
	(0.012)
Urban=1	0.278***
	(0.021)
Log number of state inspectors	0.094***
	(0.015)
Log state urban area in km2	-0.072***
	(0.013)
Interaction	-0.001
	(0.001)
Constant	-1.302***
	(0.093)
Number of observations	59,218

This table shows the probit model for the decision of being formal, in which "Works formally" is the dependent variable, only for men. The probit model is jointly estimated with the outcome equation.
Dependent variable: Works formally (=1).
Robust standard errors.
Region dummies are included.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 2.11: Wage equation - Only Men

Variables	**Treated [D=1]**	**Untreated [D=0]**
Age at the time of survey	0.076***	0.034***
	(0.002)	(0.003)
Age squared	-0.001***	-0.000***
	(0.000)	(0.000)
Married=1	0.098***	-0.016
	(0.007)	(0.010)
Schooling level (Primary schoor or less=base)		
High school=1	0.359***	0.043***
	(0.008)	(0.012)
College=1	0.819***	0.344***
	(0.010)	(0.019)
Grad school=1	1.432***	0.835***
	(0.035)	(0.067)
Race (white=base)		
Indigenous=1	-0.133**	-0.019
	(0.059)	(0.076)
Afrobrazilian=1	-0.139***	-0.065***
	(0.011)	(0.016)
Asian=1	0.142***	0.055
	(0.048)	(0.089)
Mixed=1	-0.109***	-0.062***
	(0.007)	(0.011)
Urban=1	0.186***	0.028*
	(0.014)	(0.016)
Log GDP Per Capita	0.145***	0.122***
	(0.012)	(0.022)
Constant	-1.604***	-0.507**
	(0.131)	(0.235)
Sigma	-0.640***	0.521***
	(0.004)	(0.021)
Sigma1V-Sigma0V	-1.161***	
	(0.021)	
ATE	0.305	
	(0.232)	
Number of observations	59,218	

This table shows the normal switching regression results for the outcome equation only for men.
Dependent variable: Log real hourly wage rate.
Bootstrap standard errors 50 reps.
Region dummies are included.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 2.12: Selection Equation: Probit model - One exclusion restriction

Variables	**Coef./SE**
Dep. Var: Being formal	
Female=1	-0.212***
	(0.008)
Age at the time of survey	0.057***
	(0.003)
Age squared	-0.001***
	(0.000)
Married=1	0.113***
	(0.008)
Schooling level (Primary schoor or less=base)	
High school=1	0.450***
	(0.009)
College=1	0.744***
	(0.012)
Grad school=1	0.902***
	(0.044)
Race (white=base)	
Indigenous=1	-0.126*
	(0.068)
Afrobrazilian=1	-0.124***
	(0.013)
Asian=1	0.219***
	(0.063)
Mixed=1	-0.113***
	(0.009)
Urban=1	0.334***
	(0.016)
Log number of state inspectors	0.010**
	(0.005)
Constant	-1.270***
	(0.059)
Number of observations	105,197

This table shows the probit results for the selection equation using log number of state inspectors as the only exclusion restriction.
Dependent variable: Works formally (=1).
Bootstrap standard errors 50 reps.
Region dummies are included.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1.$

Table 2.13: Wage equation - One exclusion restriction

Variables	**Treated [D=1]**	**Untreated [D=0]**
Female=1	-0.302***	-0.151***
	(0.005)	(0.007)
Age at the time of survey	0.065***	0.040***
	(0.002)	(0.002)
Age squared	-0.001***	-0.000***
	(0.000)	(0.000)
Married=1	0.079***	-0.008
	(0.005)	(0.007)
Schooling level (Primary schoor or less=base)		
High school=1	0.354***	0.047***
	(0.006)	(0.010)
College=1	0.828***	0.404***
	(0.007)	(0.015)
Grad school=1	1.465***	0.883***
	(0.024)	(0.046)
Race (white=base)		
Indigenous=1	-0.122***	-0.064
	(0.044)	(0.056)
Afrobrazilian=1	-0.148***	-0.047***
	(0.008)	(0.012)
Asian=1	0.162***	0.098
	(0.035)	(0.064)
Mixed=1	-0.123***	-0.053***
	(0.005)	(0.008)
Urban=1	0.213***	0.033**
	(0.011)	(0.013)
Log GDP Per Capita	0.155***	0.160***
	(0.008)	(0.016)
Constant	-1.470***	-1.002***
	(0.093)	(0.172)
Sigma	-0.642***	0.484***
	(0.003)	(0.018)
Sigma1V-Sigma0V	-1.125***	
	(0.019)	
ATE	0.219***	
	(0.027)	
Number of observations	105,197	

This table shows the normal switching regression results for the outcome equation using log number of state inspectors as the only exclusion restriction.
Dependent variable: Log real hourly wage rate.
Bootstrap standard errors 50 reps.
Region dummies are included.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$.

Table 2.14: Selection Equation: Probit model - Alternative definition of formality

Variables	**Coef./SE**
Dep. Var: Being formal	
Female=1	-0.228***
	(0.008)
Age at the time of survey	0.032***
	(0.003)
Age squared	-0.000***
	(0.000)
Married=1	0.105***
	(0.008)
Schooling level (Primary schoor or less=base)	
High school=1	0.455***
	(0.009)
College=1	0.793***
	(0.012)
Grad school=1	0.926***
	(0.047)
Race (white=base)	
Indigenous=1	-0.149**
	(0.069)
Afrobrazilian=1	-0.144***
	(0.014)
Asian=1	0.201***
	(0.067)
Mixed=1	-0.131***
	(0.009)
Urban=1	0.358***
	(0.016)
Log number of state inspectors	0.104***
	(0.012)
Log state urban area in sq. km	-0.076***
	(0.010)
Interaction	0.000
	(0.001)
Constant	-0.522***
	(0.074)
Number of observations	105,197

This table shows the probit results for the selection equation using an alternative definition of formality.
Dependent variable: Works formally (=1).
Bootstrap standard errors 50 reps.
Region dummies are included.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

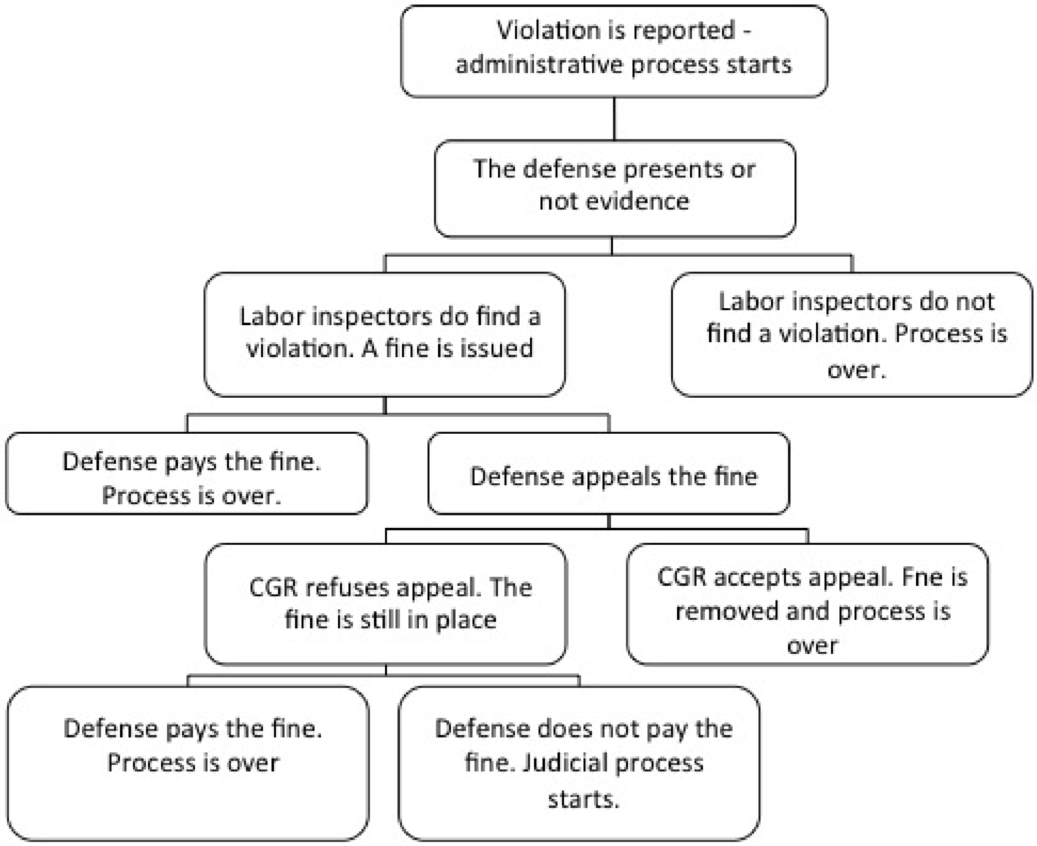
Table 2.15: Wage equation - Alternative definition of formality

Variables	**Treated [D=1]**	**Untreated [D=0]**
Female=1	-0.306***	-0.143***
	(-0.005)	(-0.008)
Age at the time of survey	0.059***	0.046***
	(-0.002)	(-0.003)
Age squared	-0.001***	-0.000***
	(0.000)	(0.000)
Married=1	0.071***	-0.001
	(-0.005)	(-0.008)
Schooling level (Primary schoor or less=base)		
High school=1	0.341***	0.044***
	(0.005)	(0.011)
College=1	0.833***	0.407***
	(0.007)	(0.019)
Grad school=1	1.477***	0.837***
	(0.023)	(0.053)
Race (white=base)		
Indigenous=1	-0.125***	-0.050
	(0.042)	(0.060)
Afrobrazilian=1	-0.150***	-0.040***
	(0.008)	(0.013)
Asian=1	0.169***	0.059
	(0.033)	(0.072)
Mixed=1	-0.126***	-0.049***
	(0.005)	(0.009)
Urban=1	0.210***	0.027*
	(0.010)	(0.014)
Log GDP Per Capita	0.165***	0.142***
	(0.008)	(0.018)
Constant	-1.470***	-1.002***
	(0.093)	(0.172)
Sigma	-0.593***	0.479***
	(0.003)	(0.024)
Sigma1V-Sigma0V	-1.071***	
	(0.024)	
ATE	0.323	
	(0.284)	
Number of observations	105,197	

This table shows the normal switching regression results for the outcome equation using an alternative definition of formality.
Dependent variable: Log real hourly wage rate.
Bootstrap standard errors 50 reps.
Region dummies are included.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Figure 2.1: Process for violations and penalties



Source: Source: Ministerio de Trabalho e Emprego

Figure 2.2: Kernel Distribution Monthly Earnings

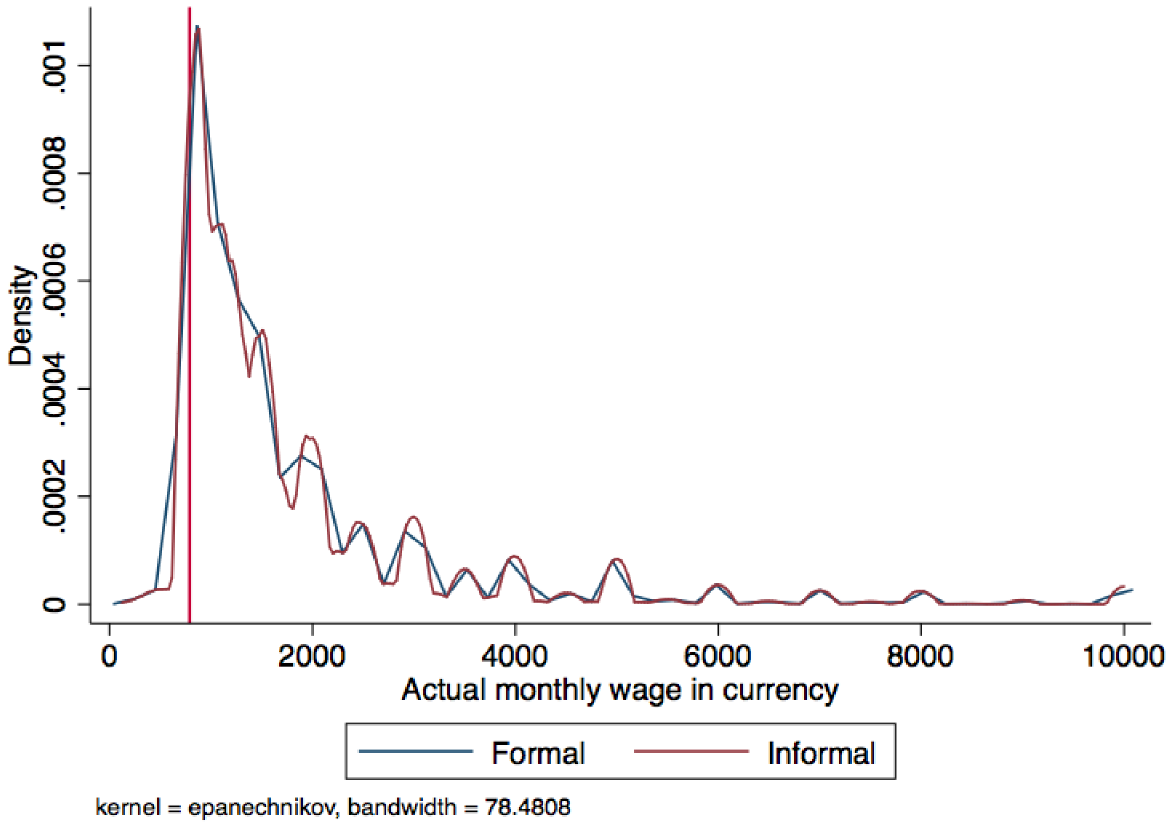


Figure 2.3: Correlation between urban area and GDP at the state level, 2015

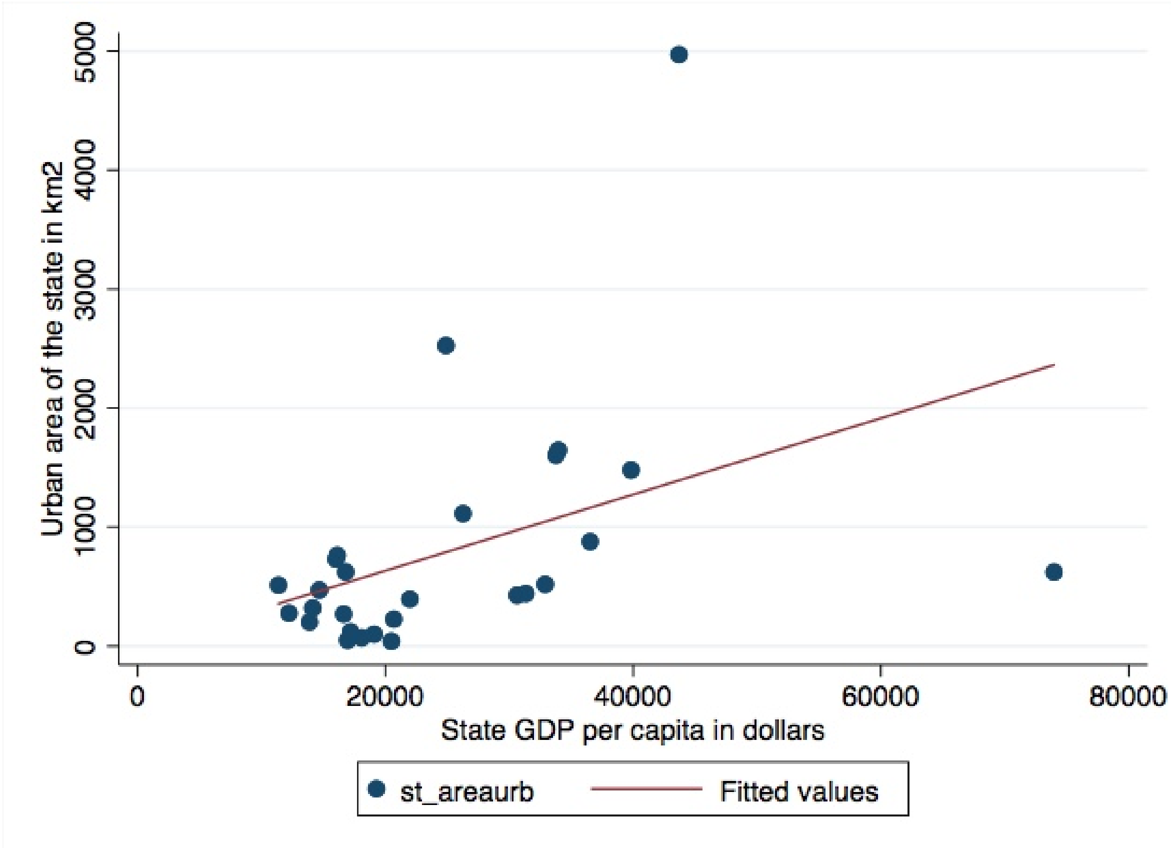
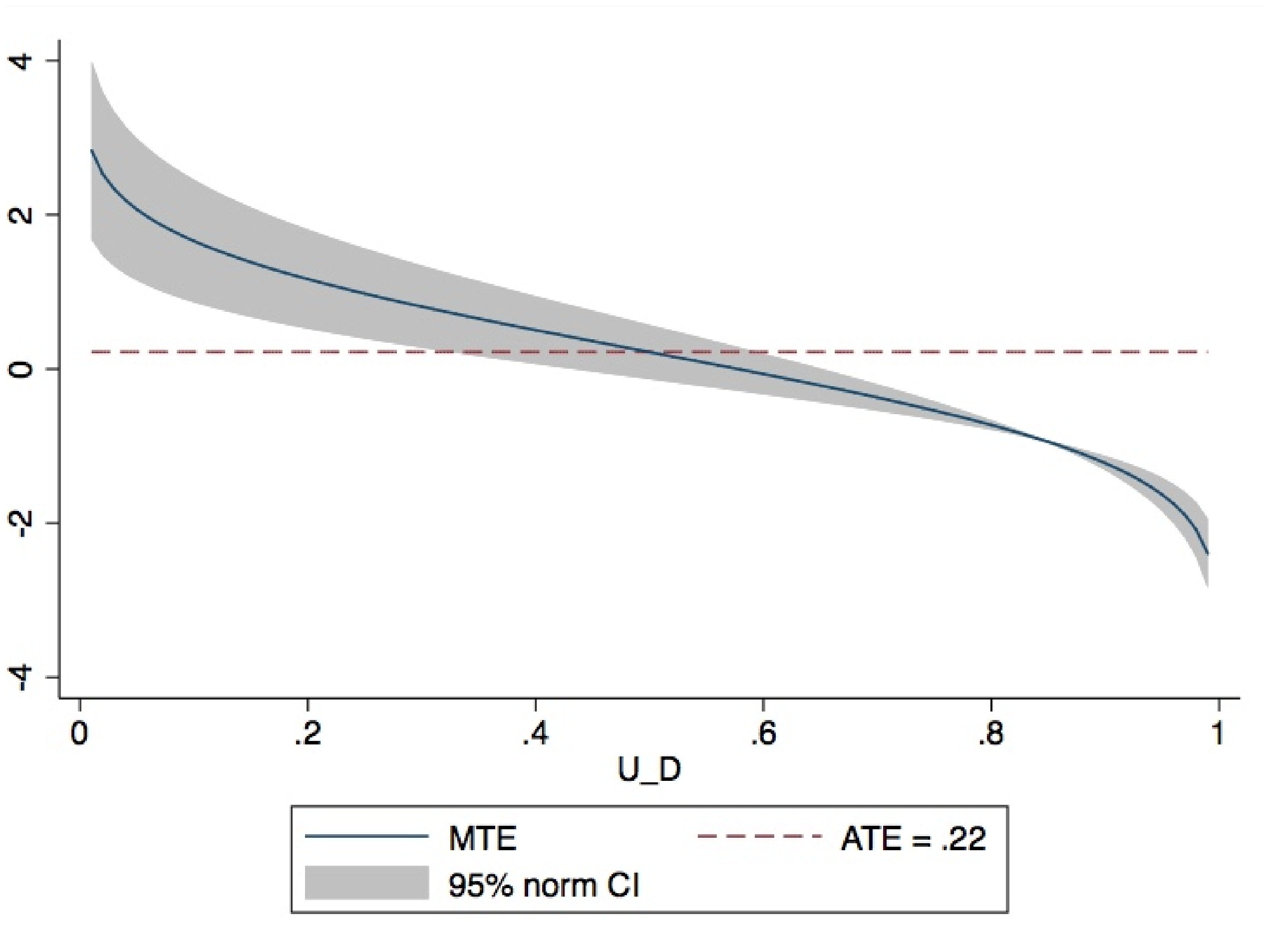


Figure 2.4: Marginal Treatment Effect



APPENDIX A

CHAPTER 1 APPENDIX

Notes:

1. Political and administrative division of the Russian Federation:

   - Federal districts (okrug): According to the classification established on May 2000, there are 7 federal districts (large regions) in Russia. This unit is equivalent to a region in the United States. Each federal district is composed for several federal subjects (ter).
   - Federal subjects (ter): According to the classification established on May 2000, there are 85 federal subjects in the Russian Federation. This unit is equivalent to a state in the United States.
   - Municipal district (Municipal Raion): They are within the limits of an administrative region. Each municipality is composed by one or more settlements.
   - Settlement: They can be urban or rural and they are composed of localities and communities. Settlements significantly vary in size. For example, Moscow City is a federal subject and also a settlement, but there are small villages that are also considered settlements. There are roughly 160 settlements in the RLMS.

2. Formal/informal status only considers individuals' main jobs. Main job is self-defined in the survey. It is the job the person spends most of her time.

Table A.1: **Description of Variables**

Variable	Notes
Hourly wage rate (log)	=ln(Labor earnings per month at main job with imputations / Hours of work per month at main job) Labor earnings per month are defined as follows: • Monthly average (over the last 12 months) after-tax real labor earnings from main job (2009-2016). • Real earnings are computed using CPI from December 2016 as the base year. • Labor earnings do not include payments in kind or other benefits. Hours of work per month are defined as follows: • Usual hours of work per month (2009-2016) Imputations: • Actual hours of work are only imputed for those who have missing value in "Usual hours worked" and worked in the last 30 days. *Source: Klara Peter*
Formal (binary)	=1 if the respondent is officially registered as an employee at a firm, =0 if individual is not officially registered as employee at a firm or if individual does not work at a firm. Note: In order to classify a worker as an"employee at a firm", the respondent has to answer "Yes, work at an enterprise or organization" to the question "Does this job belong to an enterprise or organization? I mean any organization or enterprise where more than one person works, no matter if it is private or state-owned. For example, any establishment, factory, firm, collective farm, state farm, farming industry, store, army, government service, or other organization". If the respondent is currently working but does not work at a firm, then she can work for a private person or being self-employed.

Variable	Notes
Female (binary)	=1 if the respondent is female; 0= if male
Age, Age squared	Year of survey minus year of birth; the mode of birth is computed in cases of inconsistencies across rounds.
Married (binary)	=1 if the respondent is married and living together, married and not living together, and married; =0 if has never been married, is living together unregistered, divorced, or widowed.
Education (categorical)	Respondent's level of education: =1 if lower than secondary; =2 if secondary; =3 if career or technical education; =4 if college or higher.
Education of the parents (categorical)	Highest level of education attained by either of the parents of the respondent: =0 if secondary or lower; =1 if technical education; =2 if college or higher.
Urban Status (categorical)	Respondent lives in: =1 Moscow at the time of the survey =2 Regional Center =3 Other City/Township =4 Village
Population of the settlement (log)	=ln(Population of the settlement)

Variable	Notes
Share of government employment in the community	= No. of individuals working in the government
Federal district (categorical)	Large regions of Russia: =1 if Central =2 if South =3 Northweast =4 Far East =5 Siberia =6 Ural =7 Volga
Year	Binary variable for the year of the survey (2009-2016)
Distance from settlement to labor office (log)	=ln(distance from settlement to closest federal labor inspectorate) Distance is measured as follows: • If there is no labor inspectorate in the city of residence: distance in kilometers from the center of the city of residence to the center of the city in which the federal labor inspectorate is located. • If there is a labor inspectorate office in the city of residence: distance from the closest border to the center of the city.
Ratio inspectors per economic entities.	=(No. inspectors / No. economic entities)x1000 Number of inspectors is defined at the federal subject level. Number of economic entities is defined at the federal subject level. *Source: ROSTRUD*

Variable	Notes
Has supplementary insurance paid by firm ,	=1 if answers yes, =0 if not. The variable was constructed as follows: It takes the value of 1 if the respondent answers ”Enterprise or organization” to the question “Who pays for this supplementary medical insurance?”. It takes the value of 0 otherwise. Additionally, this question is only asked to those who declare having a supplementary voluntary medical insurance.
Receives unemployment benefits if loses job in t+1	=1 if answers yes, =0 if not The variable was constructed as follows: Conditional on being non-employed in t+1, It takes the value of 1 if the respondent answers “yes” to the questions “Do you receive unemployment benefits?” and “Are you registered with a state employment agency as unemployed?”. it takes the value of 0 otherwise.
Has had paid vacation in the last 12 months	=1 if answers yes, =0 if not The variable as constructed as follows: It takes the value of 1 if the respondent answers “Yes” to the question ”Are you given the following fringe benefits in this job: Regular paid vacation?”. It takes the value of 0 otherwise.
Is satisfied with job	=1 if answers yes, =0 if not. The variable was constructed as follows: It takes the value of 1 if the respondent answers “Somewhat satisfied” or “Fully satisfied” to the question ”How satisfied or unsatisfied are you with your job in general?”. It takes the value of 0 if the respondent answers “Neutral” “Somewhat unsatisfied”, or “Fully unsatisfied”.

Variable	Notes
Is satisfied with work contract	=1 if answers yes, =0 if not The variable was constructed as follows: It takes the value of 1 if the respondent answers “Somewhat satisfied” or “Fully satisfied” to the question ”How satisfied or unsatisfied are you with your work contract?”. It takes the value of 0 if the respondent answers “Neutral” “Somewhat unsatisfied”, or “Fully unsatisfied”.
Is satisfied with job pay	=1 if answers yes, =0 if not The variable was constructed as follows: It takes the value of 1 if the respondent answers “Somewhat satisfied” or “Fully satisfied” to the question ”How satisfied or unsatisfied are you with your work contract?”. It takes the value of 0 if the respondent answers “Neutral” “Somewhat unsatisfied”, or “Fully unsatisfied”.
Is concerned about the chance of job loss	=1 if answers yes, =0 if not The variable was constructed as follows: It takes the value of 1 if the respondent answers “Not concerned at all” or “Not very concerned” to the question ”How concerned are you that might lose your job?”. It takes the value of 0 if the respondent answers “Neutral” “A little concerned”, or “Very concerned”.
Is confident in finding a job if laid off	=1 if answers yes, =0 if not The variable was constructed as follows: It takes the value of 1 if the respondent answers “Fully certain” or “Fairly certain” to the question ”How certain are you that you will be able to find an equally good job as your present one if you are laid off?”. It takes the value of 0 if the respondent answers “Neutral”, “Fairly uncertain”, or “Fully uncertain”.

RA Results for Informality Definition 1 and RA results other outcomes under definition 2

Table A.2: Probit results for selection into employment participation - Weights for RA model

Variables	**Coef./SE**
Female	-0.305***
	(0.010)
Age	0.197***
	(0.003)
Age squared	-0.258***
	(0.004)
Married (=1)	0.161***
	(0.010)
Schooling categories	
High School	0.192***
	(0.016)
Technical/Vocational	0.533***
	(0.018)
College or more	0.763***
	(0.019)
Log population site	0.039***
	(0.005)
Log distance inspection	0.020***
	(0.007)
Unemployment rate at region	-0.043***
	(0.004)
Log real GDP per capita	0.392***
	(0.021)
Constant	-7.905***
	(0.285)

Dependent variable: Employed (=1).
Robust standard errors in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$

Table A.3: RA Results using Informality definition 1 and Log real hourly wage rate as outcome variable

Variables	**Untreated**	**Treated**	**Untreated W**	**Treated W**
Dep Var: Log wage rate	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	-0.265***	-0.304***	-0.253***	-0.298***
	-0.011	-0.004	(0.012)	(0.005)
Age	0.041***	0.041***	0.039***	0.039***
	-0.004	-0.002	(0.004)	(0.002)
Age squared	-0.058***	-0.054***	-0.056***	-0.052***
	-0.005	-0.002	(0.005)	(0.002)
Married (=1)	0.122***	0.044***	0.119***	0.043***
	-0.011	-0.005	(0.012)	(0.005)
Schooling categories (omited: primary or less)				
High School	0.082***	0.091***	0.080***	0.093***
	(0.016)	(0.008)	(0.017)	(0.009)
Technical/Vocational	0.176***	0.191***	0.173***	0.198***
	(0.019)	(0.009)	(0.019)	(0.009)
College or more	0.382***	0.469***	0.377***	0.474***
	(0.021)	(0.009)	(0.021)	(0.009)
Log population site	0.024***	0.031***	0.025***	0.032***
	(0.005)	(0.002)	(0.006)	(0.002)
Log distance inspection	-0.002	-0.022***	-0.003	-0.025***
	(0.008)	(0.003)	(0.008)	(0.003)
Unemployment rate at region	-0.056***	-0.030***	-0.058***	-0.030***
	(0.004)	(0.002)	(0.004)	(0.002)
Constant	3.920***	3.693***	3.966***	3.735***
	(0.103)	(0.042)	(0.105)	(0.043)
Formal	0.067***		0.066***	
	(0.007)		(0.007)	
Observations	62,929	62,929	62,929	62,929

Dependent variable: Log real hourly wage rate.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$

Table A.4: RA Results using Informality definition 1 and Log monthly earnings as outcome variable

Variables	**Untreated W**	**Treated W**	**Untreated**	**Treated**
Dep Var: Log monthly earnings	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	-0.388***	-0.403***	-0.379***	-0.399***
	(0.013)	(0.005)	(0.013)	(0.005)
Age	0.064***	0.048***	0.062***	0.046***
	(0.005)	(0.002)	(0.005)	(0.002)
Age squared	-0.087***	-0.064***	-0.086***	-0.062***
	(0.006)	(0.002)	(0.006)	(0.002)
Married (=1)	0.128***	0.043***	0.125***	0.043***
	(0.013)	(0.005)	(0.013)	(0.005)
Schooling categories				
High School	0.085***	0.081***	0.076***	0.084***
	(0.018)	(0.009)	(0.019)	(0.010)
Technical/Vocational	0.137***	0.162***	0.135***	0.171***
	(0.021)	(0.010)	(0.022)	(0.010)
College or more	0.319***	0.407***	0.314***	0.415***
	(0.023)	(0.010)	(0.024)	(0.010)
Log population site	0.017***	0.033***	0.023***	0.035***
	(0.006)	(0.002)	(0.006)	(0.003)
Log distance inspection	-0.021**	-0.025***	-0.020**	-0.028***
	(0.008)	(0.003)	(0.008)	(0.003)
Unemployment rate at region	-0.064***	-0.027***	-0.068***	-0.027***
	(0.005)	(0.002)	(0.005)	(0.002)
Constant	9.034***	8.806***	9.052***	8.832***
	(0.113)	(0.045)	(0.115)	(0.047)
ATE (Formal=1 vs Informal=0)	0.024***		0.025***	
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	60,310	60,310	60,310	60,310

Dependent variable: Log monthly earnings.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$

Table A.5: RA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (I)		Dep Var: (II)		Dep Var: (III)	
	Untreated	**Treated**	**Untreated**	**Treated**	**Untreated**	**Treated**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	-0.064	-0.224***	-0.222	0.073	0.217***	0.133***
	-0.063	-0.02	(0.155)	(0.095)	(0.029)	(0.015)
Age	0.006	0.035***	0.162***	0.142***	0.070***	0.097***
	-0.022	-0.008	(0.060)	(0.036)	(0.011)	(0.005)
Age squared	-0.007	-0.046***	-0.205***	-0.162***	-0.081***	-0.099***
	-0.029	-0.009	(0.077)	(0.044)	(0.014)	(0.007)
Married (=1)	0.249***	0.126***	0.018	0.041	0.239***	0.113***
	-0.065	-0.021	(0.152)	(0.092)	(0.029)	(0.016)
Schooling categories						
High School	-0.081	0.130***	0.399	0.234	0.088*	0.122***
	(0.112)	(0.046)	(0.250)	(0.166)	(0.046)	(0.028)
Technical/Vocational	0.193*	0.154***	0.573*	0.277	0.422***	0.285***
	(0.117)	(0.048)	(0.296)	(0.182)	(0.051)	(0.030)
College or more	0.308**	0.395***	1.151***	0.368*	0.682***	0.417***
	(0.120)	(0.047)	(0.308)	(0.192)	(0.052)	(0.030)
Log population site	0.105***	0.092***	-0.169***	-0.048	0.050***	0.039***
	(0.034)	(0.011)	(0.064)	(0.042)	(0.017)	(0.008)
Log distance inspection	0.096**	-0.025*	-0.040	-0.089	-0.004	-0.011
	(0.047)	(0.014)	(0.112)	(0.059)	(0.021)	(0.010)
Unemployment rate at region	0.076***	0.020***	0.212***	0.179***	0.013	0.019***
	(0.021)	(0.007)	(0.047)	(0.032)	(0.011)	(0.006)
Constant	-4.151***	-3.288***	-5.274***	-4.161***	-2.431***	-2.051***
	(0.626)	(0.194)	(1.428)	(0.895)	(0.276)	(0.137)
ATE (Formal=1 vs Informal=0)	0.053***		0.044***		0.309***	
	(0.002)		(0.015)		(0.006)	
Observations	51,989	51,989	2,114	2,114	48,033	48,033

Dependent variables: (I) Has Supplementary health insurance paid by the firm, (II) Receives unemployment benefits if loses job in t+1, (III) Paid vacation in the last 12 months.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table A.6: RA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (IV)		Dep Var: (V)		Dep Var: (VI)	
	Untreated	**Treated**	**Untreated**	**Treated**	**Untreated**	**Treated**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	0.085***	-0.031**	0.200***	0.052***	0.003	-0.148***
	(0.026)	(0.014)	(0.026)	(0.014)	-0.026	-0.014
Age	-0.036***	-0.035***	-0.040***	-0.042***	-0.020**	-0.035***
	(0.009)	(0.005)	(0.009)	(0.005)	-0.009	-0.005
Age squared	0.044***	0.043***	0.048***	0.052***	0.022*	0.039***
	(0.012)	(0.006)	(0.012)	(0.006)	-0.012	-0.006
Married (=1)	0.138***	0.090***	0.139***	0.049***	0.091***	0.057***
	(0.026)	(0.014)	(0.026)	(0.014)	-0.026	-0.014
Schooling categories						
High School	0.124***	0.043	0.143***	0.063**	0.080*	0.077***
	(0.039)	(0.027)	(0.039)	(0.026)	(0.041)	(0.028)
Technical/Vocational	0.190***	0.166***	0.230***	0.185***	0.111**	0.156***
	(0.045)	(0.028)	(0.045)	(0.028)	(0.047)	(0.029)
College or more	0.235***	0.349***	0.380***	0.461***	0.239***	0.384***
	(0.046)	(0.028)	(0.046)	(0.028)	(0.047)	(0.028)
Log population site	-0.006	0.023***	-0.028**	0.012*	-0.020	-0.005
	(0.014)	(0.007)	(0.014)	(0.007)	(0.014)	(0.007)
Log distance inspection	-0.035*	0.069***	-0.034*	0.046***	-0.050***	-0.027***
	(0.018)	(0.009)	(0.018)	(0.009)	(0.018)	(0.009)
Unemployment rate at region	0.056***	0.003	0.041***	0.012**	0.063***	-0.002
	(0.009)	(0.005)	(0.009)	(0.005)	(0.009)	(0.005)
Constant	0.313	0.354***	0.512**	0.379***	-0.298	0.187
	(0.241)	(0.127)	(0.241)	(0.126)	(0.247)	(0.126)
ATE (Formal=1 vs Informal=0)	0.089***		0.078***		-0.015***	
	(0.006)		(0.006)		(0.006)	
Observations	51,709	51,709	51,614	51,614	51,517	51,517

Dependent variables: (IV) Satisfied with job, (V) Satisfied with work contract, (VI) Satisfied with pay.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table A.7: RA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (VII)		Dep Var: (VIII)	
	Untreated	**Treated**	**Untreated**	**Treated**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	0.020	-0.010	-0.114***	-0.137***
	(0.027)	(0.014)	(0.029)	(0.014)
Age	-0.042***	-0.060***	-0.024**	-0.000
	(0.010)	(0.005)	(0.011)	(0.005)
Age squared	0.043***	0.069***	-0.002	-0.029***
	(0.012)	(0.006)	(0.014)	(0.006)
Married (=1)	0.096***	-0.021	0.031	-0.059***
	(0.027)	(0.014)	(0.029)	(0.014)
Schooling categories				
High School	0.033	0.078***	0.049	0.046
	(0.041)	(0.029)	(0.045)	(0.029)
Technical/Vocational	0.035	0.090***	0.058	0.021
	(0.047)	(0.030)	(0.050)	(0.030)
College or more	-0.007	0.192***	-0.009	0.105***
	(0.048)	(0.030)	(0.051)	(0.029)
Log population site	0.051***	0.000	0.038**	0.038***
	(0.015)	(0.007)	(0.017)	(0.007)
Log distance inspection	0.035*	-0.048***	-0.063***	-0.121***
	(0.019)	(0.009)	(0.021)	(0.009)
Unemployment rate at region	-0.011	-0.031***	0.015	-0.013**
	(0.010)	(0.005)	(0.011)	(0.005)
Constant	-0.195	0.842***	0.184	0.080
	(0.251)	(0.129)	(0.276)	(0.130)
ATE (Formal=1 vs Informal=0)	-0.024***		-0.044***	
	(0.005)		(0.006)	
Observations	51,816	51,816	47,227	47,227

Dependent variables: (VII) Not concerned about chance of job loss, (VIII) Confident in finding a job if laid off.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

IPWRA Results for Informality Definition 1 and RA results other outcomes under definition 2

Table A.8: IPWRA Results using Informality definition 1 and Log wage rate as outcome variable

Variables	**Untreated [D=0]**	**Treated [D=1]**	**Selection Eq.**
Dep Var: Log wage rate	Coef./SE	Coef./SE	Coef./SE
Female	-0.258***	-0.305***	0.182***
	-0.014	-0.004	-0.013
Age	0.045***	0.040***	-0.011**
	-0.005	-0.002	-0.005
Age squared	-0.062***	-0.053***	0.021***
	-0.007	-0.002	-0.006
Married (=1)	0.104***	0.047***	0.200***
	-0.014	-0.005	-0.013
Schooling categories			
High School	0.070***	0.089***	0.080***
	(0.017)	(0.008)	(0.020)
Technical/Vocational	0.161***	0.188***	0.396***
	(0.020)	(0.009)	(0.023)
College or more	0.389***	0.462***	0.678***
	(0.022)	(0.009)	(0.024)
Log population site	0.017**	0.033***	0.010
	(0.007)	(0.002)	(0.007)
Log distance inspection	0.004	-0.022***	-0.009
	(0.009)	(0.003)	(0.011)
Unemployment rate at region	-0.049***	-0.029***	-0.015***
	(0.005)	(0.002)	(0.005)
Share public employment in community			0.754***
			(0.059)
Ratio inspectors per 1000 entities			0.270***
			(0.085)
Distance x Ratio Inspectors			-0.046**
			(0.020)
Constant	3.834***	3.709***	0.439***
	(0.129)	(0.042)	(0.129)
ATE (Formal=1 vs Informal=0)	0.066***		
	(0.007)		
Observations	62,929	62,929	62,929

Dependent variable: Log real hourly wage rate.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$

Table A.9: IPWRA Results using Informality definition 1 and Log monthly earnings as outcome variable

Variables	**Untreated [D=0]**	**Treated [D=1]**	**Selection Eq.**
Dep Var: Log earnings	Coef./SE	Coef./SE	Coef./SE
Female	-0.376***	-0.403***	0.183***
	(0.015)	(0.005)	(0.013)
Age	0.067***	0.047***	-0.008
	(0.006)	(0.002)	(0.005)
Age squared	-0.090***	-0.063***	0.018***
	(0.008)	(0.002)	(0.006)
Married (=1)	0.100***	0.048***	0.200***
	(0.016)	(0.005)	(0.014)
Schooling categories			
High School	0.069***	0.079***	0.082***
	(0.020)	(0.009)	(0.021)
Technical/Vocational	0.125***	0.159***	0.403***
	(0.023)	(0.010)	(0.024)
College or more	0.339***	0.401***	0.703***
	(0.025)	(0.010)	(0.025)
Log population site	0.002	0.034***	0.010
	(0.008)	(0.002)	(0.007)
Log distance inspection	-0.012	-0.025***	0.006
	(0.010)	(0.003)	(0.012)
Unemployment rate at region	-0.058***	-0.027***	-0.016***
	(0.006)	(0.002)	(0.005)
Share public employment in community			0.770***
			(0.061)
Ratio inspectors per 1000 entities			0.326***
			(0.087)
Distance x Ratio Inspectors			-0.058***
			(0.021)
Constant	9.008***	8.824***	0.350***
	(0.141)	(0.045)	(0.133)
ATE (Formal=1 vs Informal=0)	0.022***		
	(0.008)		
Observations	60,310	60,310	60,310

Dependent variable: Log rmonthly earnings.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01, \quad **p < 0.05, \quad *p < 0.1$

Table A.10: IPWRA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (I)			Dep Var: (II)		
	Untreated	**Treated**	**Selection Eq.**	**Untreated**	**Treated**	**Selection Eq.**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	-0.166**	-0.225***	0.146***	-0.306*	0.138	0.217***
	(0.077)	(0.020)	(0.013)	(0.161)	(0.095)	(0.059)
Age	0.006	0.033***	-0.021***	0.162**	0.133***	-0.015
	(0.031)	(0.008)	(0.005)	(0.064)	(0.036)	(0.021)
Age squared	-0.006	-0.044***	0.041***	-0.210***	-0.148***	0.031
	(0.040)	(0.009)	(0.006)	(0.081)	(0.044)	(0.026)
Married (=1)	0.268***	0.132***	0.204***	-0.048	0.077	0.180***
	(0.072)	(0.021)	(0.014)	(0.153)	(0.092)	(0.061)
Schooling categories						
High School	-0.089	0.115**	0.085***	0.617**	0.246	0.133
	(0.120)	(0.047)	(0.023)	(0.278)	(0.168)	(0.089)
Technical/Vocational	0.211*	0.143***	0.356***	0.648**	0.341*	0.433***
	(0.123)	(0.049)	(0.025)	(0.309)	(0.185)	(0.106)
College or more	0.392***	0.382***	0.611***	1.222***	0.411**	0.533***
	(0.137)	(0.047)	(0.025)	(0.322)	(0.195)	(0.111)
Log population site	0.096**	0.090***	0.010	-0.173**	-0.070	0.027
	(0.039)	(0.011)	(0.008)	(0.072)	(0.043)	(0.031)
Log distance inspection	0.105*	-0.033**	0.146***	-0.100	-0.063	0.286***
	(0.058)	(0.014)	(0.012)	(0.108)	(0.058)	(0.056)
Unemployment rate at region	0.090***	0.023***	-0.007	0.179***	0.177***	-0.026
	(0.024)	(0.007)	(0.005)	(0.048)	(0.033)	(0.020)
Share public employment in community			0.711***			0.711***
			(0.063)			(0.269)
Ratio inspectors per 1000 entities			0.175**			0.338
			(0.085)			(0.391)
Distance x Ratio Inspectors			-0.064***			-0.100
			(0.021)			(0.093)
Constant	-4.269***	-3.208***	-0.325**	-4.801***	-3.997***	-1.357**
	(0.839)	(0.197)	(0.138)	(1.454)	(0.908)	(0.601)
ATE (Formal=1 vs Informal=0)	0.052***			0.046***		
	(0.002)			(0.002)		
Observations	51,989	51,989	51,989	2,114	2,114	2,114

Dependent variables: (I) Has Supplementary health insurance paid by the firm, (II) Receives unemployment benefits if loses job in t+1.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. "Primary or less" is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table A.11: IPWRA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (III)			Dep Var: (IV)		
	Untreated	**Treated**	**Selection Eq.**	**Untreated**	**Treated**	**Selection Eq.**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	0.249***	0.131***	0.114***	0.125***	-0.035**	0.145***
	(0.033)	(0.015)	(0.014)	(0.029)	(0.014)	(0.013)
Age	0.086***	0.096***	-0.019***	-0.022*	-0.036***	-0.021***
	(0.013)	(0.006)	(0.005)	(0.011)	(0.005)	(0.005)
Age squared	-0.102***	-0.098***	0.038***	0.026*	0.044***	0.040***
	(0.016)	(0.007)	(0.007)	(0.014)	(0.006)	(0.006)
Married (=1)	0.175***	0.115***	0.181***	0.142***	0.088***	0.203***
	(0.033)	(0.016)	(0.015)	(0.029)	(0.014)	(0.014)
Schooling categories						
High School	0.068	0.119***	0.124***	0.118***	0.044*	0.085***
	(0.050)	(0.029)	(0.025)	(0.041)	(0.027)	(0.023)
Technical/Vocational	0.395***	0.281***	0.350***	0.188***	0.168***	0.354***
	(0.055)	(0.030)	(0.027)	(0.048)	(0.028)	(0.025)
College or more	0.635***	0.412***	0.589***	0.211***	0.349***	0.609***
	(0.058)	(0.030)	(0.027)	(0.050)	(0.028)	(0.025)
Log population site	0.059***	0.039***	-0.008	0.034**	0.018**	0.008
	(0.019)	(0.008)	(0.008)	(0.017)	(0.007)	(0.008)
Log distance inspection	0.026	-0.013	0.155***	-0.040*	0.064***	0.146***
	(0.024)	(0.010)	(0.013)	(0.021)	(0.009)	(0.012)
Unemployment rate at region	0.016	0.019***	0.024***	0.066***	-0.000	-0.005
	(0.013)	(0.006)	(0.006)	(0.010)	(0.005)	(0.005)
Share public employment in community			0.568***			0.709***
			(0.068)			(0.064)
Ratio inspectors per 1000 entities			0.239***			0.171**
			(0.090)			(0.085)
Distance x Ratio Inspectors			-0.094***			-0.063***
			(0.023)			(0.021)
Constant	-2.933***	-2.011***	-0.334**	-0.294	0.462***	-0.334**
	(0.323)	(0.138)	(0.149)	(0.287)	(0.128)	(0.138)
ATE (Formal=1 vs Informal=0)	0.308***			0.090***		
	(0.006)			(0.006)		
Observations	48,033	48,033	48,033	51,709	51,709	51,709

Dependent variables: (III) Paid vacation in the last 12 months, (IV) Satisfied with job.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table A.12: IPWRA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (V)			Dep Var: (VI)		
	Untreated	**Treated**	**Selection Eq.**	**Untreated**	**Treated**	**Selection Eq.**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	0.240***	0.051***	0.145***	0.026	-0.146***	0.145***
	(0.029)	(0.014)	(0.013)	(0.030)	(0.014)	(0.013)
Age	-0.028**	-0.042***	-0.021***	-0.013	-0.036***	-0.021***
	(0.011)	(0.005)	(0.005)	(0.011)	(0.005)	(0.005)
Age squared	0.031**	0.052***	0.040***	0.014	0.040***	0.040***
	(0.014)	(0.006)	(0.006)	(0.015)	(0.006)	(0.006)
Married (=1)	0.163***	0.045***	0.203***	0.085***	0.056***	0.202***
	(0.029)	(0.014)	(0.014)	(0.030)	(0.014)	(0.014)
Schooling categories						
High School	0.139***	0.064**	0.087***	0.075*	0.075***	0.086***
	(0.042)	(0.027)	(0.023)	(0.043)	(0.028)	(0.023)
Technical/Vocational	0.236***	0.187***	0.355***	0.115**	0.152***	0.354***
	(0.048)	(0.028)	(0.025)	(0.050)	(0.029)	(0.025)
College or more	0.371***	0.463***	0.612***	0.255***	0.375***	0.612***
	(0.050)	(0.028)	(0.025)	(0.051)	(0.029)	(0.025)
Log population site	0.001	0.009	0.009	0.001	-0.007	0.009
	(0.017)	(0.007)	(0.008)	(0.017)	(0.007)	(0.008)
Log distance inspection	-0.027	0.042***	0.146***	-0.062***	-0.027***	0.145***
	(0.021)	(0.009)	(0.012)	(0.022)	(0.009)	(0.012)
Unemployment rate at region	0.049***	0.010*	-0.005	0.062***	-0.004	-0.004
	(0.011)	(0.005)	(0.005)	(0.011)	(0.005)	(0.005)
Share public employment in community			0.712***			0.709***
			(0.064)			(0.064)
Ratio inspectors per 1000 entities			0.175**			0.165*
			(0.085)			(0.085)
Distance x Ratio Inspectors			-0.065***			-0.063***
			(0.021)			(0.021)
Constant	-0.033	0.454***	-0.345**	-0.466	0.235*	-0.336**
	(0.284)	(0.127)	(0.139)	(0.294)	(0.127)	(0.139)
ATE (Formal=1 vs Informal=0)	0.078***			-0.016***		
	(0.006)			(0.006)		
Observations	51,614	51,614	51,614	51,517	51,517	51,517

Dependent variables: V) Satisfied with work contract, (VI) Satisfied with pay.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table A.13: IPWRA Results using Informality definition 2 and different non-wage labor market outcomes as dependent variable

Variables	Dep Var: (VII)			Dep Var: (VIII)		
	Untreated	**Treated**	**Selection Eq.**	**Untreated**	**Treated**	**Selection Eq.**
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Female	0.050*	-0.010	0.147***	-0.157***	-0.133***	0.115***
	(0.030)	(0.014)	(0.013)	(0.033)	(0.014)	(0.015)
Age	-0.045***	-0.060***	-0.022***	-0.018	-0.002	-0.019***
	(0.011)	(0.005)	(0.005)	(0.013)	(0.005)	(0.005)
Age squared	0.046***	0.068***	0.042***	-0.010	-0.027***	0.039***
	(0.015)	(0.006)	(0.006)	(0.016)	(0.006)	(0.007)
Married (=1)	0.121***	-0.022	0.204***	0.032	-0.055***	0.186***
	(0.031)	(0.015)	(0.014)	(0.033)	(0.014)	(0.015)
Schooling categories						
High School	0.010	0.079***	0.083***	0.057	0.048*	0.120***
	(0.044)	(0.029)	(0.023)	(0.048)	(0.029)	(0.025)
Technical/Vocational	-0.004	0.090***	0.353***	0.038	0.022	0.348***
	(0.050)	(0.030)	(0.025)	(0.054)	(0.030)	(0.027)
College or more	-0.021	0.190***	0.608***	-0.027	0.101***	0.590***
	(0.052)	(0.030)	(0.025)	(0.056)	(0.030)	(0.027)
Log population site	0.067***	-0.001	0.011	0.048**	0.036***	-0.006
	(0.017)	(0.007)	(0.008)	(0.020)	(0.007)	(0.008)
Log distance inspection	0.017	-0.048***	0.148***	-0.078***	-0.123***	0.153***
	(0.021)	(0.009)	(0.012)	(0.024)	(0.009)	(0.013)
Unemployment rate at region	-0.003	-0.029***	-0.006	0.026**	-0.014**	0.025***
	(0.011)	(0.005)	(0.005)	(0.013)	(0.005)	(0.006)
Share public employment in community			0.707***			0.599***
			(0.063)			(0.069)
Ratio inspectors per 1000 entities			0.181**			0.249***
			(0.085)			(0.091)
Distance x Ratio Inspectors			-0.065***			-0.097***
			(0.021)			(0.023)
Constant	-0.245	0.833***	-0.329**	0.051	0.141	-0.347**
	(0.290)	(0.130)	(0.138)	(0.329)	(0.131)	(0.150)
ATE (Formal=1 vs Informal=0)	-0.024***			-0.043***		
	(0.005)			(0.006)		
Observations	51,816	51,816	51,816	47,227	47,227	47,227

Dependent variables: (VII) Not concerned about chance of job loss, (VIII) Confident in finding a job if laid off.
Clustered standard errors at the individual level in parentheses.
Model includes education of the parents, urban status, indicator for if a SOE closed in the community in the past 12 months, region and year dummies. “Primary or less” is excluded from the schooling categories.

Significance: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Fines and penalties imposed by labor inspectors

Table A.14: Amount of fines in 1,000 rubles and number of penalties by year

Year	Fines	Penalties
2009	4.92	2,427.98
2010	5.08	2,057.83
2011	5.86	2,199.43
2012	6.10	1,977.42
2013	10.11	2,273.40
2014	4.07	3,054.46
2015	3.22	3,330.44
2016	2.29	3,856.12

Table A.15: Amount of fines in 1,000 rubles and number of penalties by okrug

Okrug	Fines	Penalties
Central	6.49	2,671.85
North West	5.84	1,988.13
South	5.75	4,077.17
Volga	4.29	2,518.04
Urals	4.22	2,304.26
Siberia	4.32	2,168,00
Far East	4.50	1,825.46

APPENDIX B

CHAPTER 2 APPENDIX

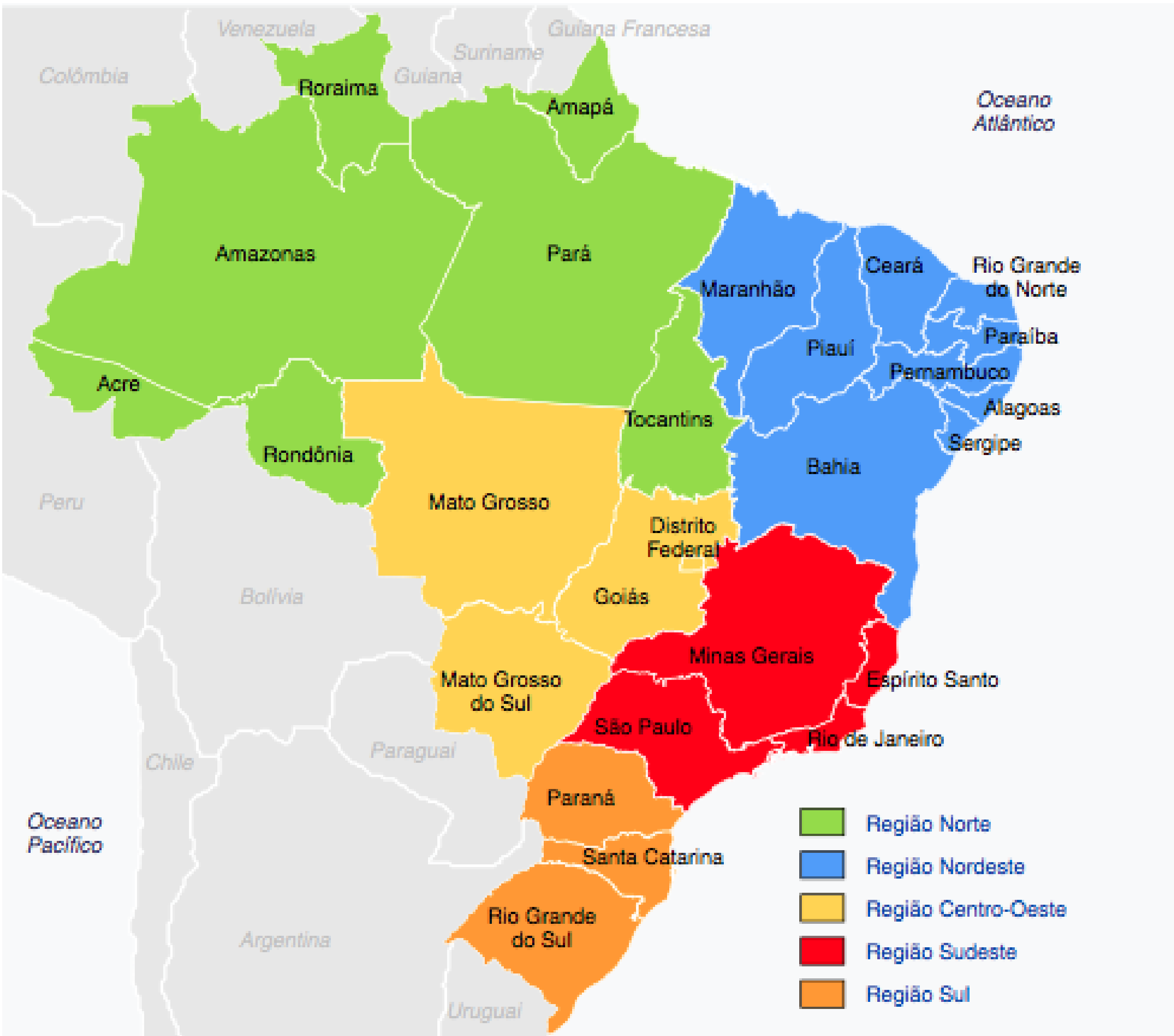
Notes:

1. Political and administrative division of Brazil:

   - Regions: There are 5 administrative regions in Brazil created by the Brazilian Institute of Geography and Statistics. The regions are: North, Northeast, Central-West, Southeast and South region. States in each region share economic and geographic characteristics. See Figure B.1 for details.
   - Federative Units: There are 27 federative units in Brazil: One Federal District, where the administrative capital of the country is located, and 26 states. This unit is equivalent to a state in the United States. For the purpose of this work, all the federative units are going to be called "states".
   - Municipalities: Each state is divided in municipalities. There are over 5,500 municipalities in the country. In this paper, municipalities are not used.

2. Formal/informal status only considers individuals' main jobs. Main job is defined in the survey as the the job the person spent most of her time during the previous 365 days.

Figure B.1: Regions of Brazil



Source: Felipe Menegaz, distributed under CC Attribution-Share Alike 3.0 Unported license.

Table B.1: **Description of Variables**

Variable	Notes
Hourly wage rate (log)	=ln(Labor earnings per month at main job / Weekly hours of work * 4.33) Labor earnings per month are defined as follows: • Monthly average (over the last 12 months) after-tax real labor earnings from main job in 2015 • Real earnings are computed using average CPI from 2015 as the base year. • Labor earnings do not include payments in kind or other benefits. Hours of work per week are defined as follows: • Usual hours worked per week for 2015 Factor of conversion from weekly hours to monthly hours: 4.33
Formal (binary)	=1 if the respondent has a signed workers card or if the respondent has an official registration (CNPJ) of her entrepreneurial activity or free-lance activity. =0 if individual does not have a signed workers card or does not have a CNPJ Note: In order to classify the job of a worker, the respondent has to answer the following question "In your job you work as: [5 options are displayed]" The options are: employee, maid, self-employed, entrepreneur with at least one employee, non-salaried worker, construction worker working for myself state farm, farming industry, store, army, government service, or other.

Variable	Notes
Female (binary)	=1 if the respondent is female; 0= if male
Age, Age squared Age at the time of the survey.	
Married (binary)	=1 if the respondent is married or living together, =0 if has never been married, divorced, or widowed.
Race (categorical)	Respondent's self-declared race: =1 if white (excluded category) =2 if afro-brazilian =3 if asian-brazilian =4 if indigenous =5 if mixed
Education (categorical)	Respondent's highest level of education: =1 if primary school or less (excluded category) =2 if middle and high school =3 if college =4 if grad school.
Urban Status (binary)	Respondent lives in: =1 if urban =0 if rural

Regions (categorical)	Large regions of Brazil: =1 if North =2 if Northeast =3 Center-West =4 Southeast (excluded category) =5 South
GDP per capita	GDP per capita is measured as follows: • State GDP is deflated by using the CPI then each state GDP is divided by the size of the state population at that year *GDP source: Brazilian Institute of Geography and Statistics* *CPI source: Fundacao Getulio Vargas* *Population Source: Brazilian Institute of Geography and Statistics*
Inspectors per state	=ln(inspectors per state) *Source: Ministry of Labor and Employment*
Ratio inspectors per inspection office	=ln (No. inspectors / No. offices) Number of inspectors is defined at the state level. Number of inspection offices is defined at the state level *Source: Ministry of Labor and Employment*
Urban area of the state	=ln(Urban area of the state in sq. km) *Source: Instituto Brasileiro de Geografia e Estatistica*

BIBLIOGRAPHY

Almeida, R. and P. Carneiro. (2012). Enforcement of Labor Regulation and Informality. *American Economic Journal: Applied Economics*, 4(3), 64-89

Arias, O. and M. Khamis. (2008). Comparative Advantage, Segmentation and Informal Earnings: A Marginal Treatment Effects Approach. IZA discussion paper 3916, Institute for the Study of Labor

Belskaya, O., K. S. Peter, and C. Posso. (2014). College Expansion and the Marginal Returns to Education: Evidence from Russia. IZA Discussion Papers 8735, Institute for the Study of Labor

Bobba, M., L. Flabbi, and S. Levy. (2018). Labor Market Search, Informality, and Schooling Investments, IDB Working Paper Series No. IDB-WP-863, Inter-American Development Bank

Bosch, M., M. B. Cobacho, and C. Pages. (2014). "Effects of Non-Contributory Systems on Informality: Taking Stock of Eight Years of Implementation of Mexico's Seguro Popular" In Social Insurance, Informality, and Labor Markets, edited by M. Frolich, D Kaplan, C. Pages, J. Rigolini, and D. Robalino, 80-109. Oxford: Oxford University Press, 2014

Botelho, F. and V. Ponczek. (2011). Segmentation in the Brazilian Labor Market. *Economic Development and Cultural Change*, 59(2), 437-463

Cameron S. V. and J. Heckman. (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*, 106(2), 262-333

Cameron S. V. and J. Heckman. (2001).The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*, 109(3), 455-499

Carneiro, P., J. Heckman, and E. J. Vytlacil. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(4), 2754-2781

Cardoso, A. and T. Lage. (2005). A Inspeção do Trabalho no Brasil, *Revista de Ciencias Sociais Dados*, 48(3), 451-490

Gimpelson, V. and R. Kapeliushnikov. (2014). Between Light and Shadow: Informality in the Russian Labour Market. IZA discussion paper 8279, Institute for the Study of Labor

Haanwinckel, D., and R. Soares. (2016). Workforce Composition, Productivity, and Labor Regulations in a Compensating Differentials Theory of Informality. IZA Discussion papers 995, Institute of Labor Economics

Heckman, J., D. Schmierer and S. Urzua. (2010). Testing the Correlated Random Coefficient Model. *Journal of Econometrics*, 158(2), 177-203

Heckman, J., S. Urzua, and E. Vytlacil. (2006). Estimation of Treatment Effects under Essential Heterogeneity, working paper

Heckman, J. and E. Vytlacil. (2005). Structural Equations, Treatment Effects, and Econometric Policy

Evaluation, Econometrica , 73(3), 669-738

Hirschman, A.O., Exit, Voice, and Loyalty: Responses to Decline in FIrms, Organizations, and States. (1970). Harvard University Press

IPEA. (2016). Boletim Mercado de Trabalho, Conjuntura e Analise No. 60, April 2016

Koettl, J. and M. Weber. (2014). "Does Formal Work Pay? The Role of Labor Taxation and Social Benefit Design in the New Member States"' In Social Insurance, Informality, and Labor Markets, edited by M. Frolich, D Kaplan, C. Pages, J. Rigolini, and D. Robalino, 147-188. Oxford: Oxford University Press, 2014

La Porta, R. and A. Shleifer. (2014). Informality and Development. *Journal of Economic Perspectives*, 28(3), 109-126

Lehman, H. (2014). " Informal Employment in Transition Countries: Empirical Evidence and Research Challenges", *Comparative Economic Studies*, 57 (1), 1-30

Lehman, H. and A, Muravyev. (2014). "Labor Market Institutions and Informality in Transition and Latin American Countries" in Social Insurance, Informality, and Labor Markets, edited by M. Frolich, D Kaplan, C. Pages, J. Rigolini, and D. Robalino, 9-29, Oxford: Oxford University Press, 2014

Lehmann, H. and A. Zaiceva. (2015). Re-defining Informal Employment and Measuring its Determinants: Evidence from Russia. *Journal of International Development*, 27 (4), 464 ? 488

Levy, S., Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico. (2008). Brookings Institution Press

Lewis, W.A. (1954). Economic Development with Unlimited Supplies of Labour, The Manchester School, 22(2), 139-191

Lokshin, M. and Z. Sajaia. (2004). Maximum likelihood estimation of endogenous switching regression models. *The Stata Journal*, 4(3), 284-289

Maddala, G. S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press

Magnac, T. (1991). Segmented or competitive labor markets.*Econometrica*, 165-187

Maloney, W. (2004). Informality Revisited. *World Development*, 32 (7), 1159-1178

Meghir, C., R. Narita, and J.M. Robin. (2015). Wages and Informality in Developing Countries. *American Economic Review*, 105 (4), 1509-1546

Pagés, C., J. Rigolini, and D. Robalino. (2014). "Social Insurance, Informality, and Labor Markets: How to Protect Workers While Creating Good Jobs" in Social Insurance, Informality, and Labor Markets, edited by M. Frolich, D Kaplan, C. Pages, J. Rigolini, and D. Robalino, 9-29. Oxford: Oxford University Press, 2014

Perry, G., W. Maloney, O. Arias, P. Fajnzylber, A. Mason, and J. Saavedra-Chanduvi. (2007). Informality: Exit and Exclusion, The World Bank

Pratap, S. and E. Quintin. (2006). Are labor markets segmented in developing countries? A semiparametric approach. *European Economic Review*, 50, 1817-1841

Slominckyk, F. and V. Gimpelson. (2015). Informality and Mobility: Evidence from Russian Panel Data, *Economics of Transition*, 23 (2), 299-341

Tannuri-Pianto, M.E., D. Pianto, and O. Arias. (2004). Informal Employment in Bolivia: A Lost Proposition?. Working Paper Departamento de Economia Universidade de Brasilia 317

Ulyssea, Gabriel. (2018). "Firms, informality, and development: Theory and evidence from Brazil." *American Economic Review*, 108.8, 2015-47

Roy, A. (1951). Some Thoughts on the Distribution on Earnings, *Oxford Economic Paper*, 3(2), 135-146

Viollaz, M. (2016). Are Labor Inspections Protecting Workers' Rights? Adding the Evidence from Size?based Labor Regulations and Fines in Peru. Forthcoming *International Labour Review*

Willis, R. and S. Rosen. (1979). Education and Self-Selection. *Journal of Political Economy*, 87(5), S7-S36