



**Universidad del  
Rosario**

# **Small Business and the Informal Economy**

**Cristina Fernández**

**Faculty Advisors:**

Juan Miguel Gallego and Andrés García-Suaza

**Ph.D in Economics**

**Universidad del Rosario**

Facultad de economía

Ph.D in Economics

Bogotá-Colombia

2023

*To Mauricio, Amalia, Andrea, Isabela,  
and to my mother,  
who by example and direct advice,  
always taught me to follow my dreams,  
even if I thought that it was too late.*

## **Main idea**

Colombia is a country of small firms. The main idea of this dissertation is that there are disincentives for these firms to use capital and skilled workers, affecting their productivity and also keeping them from growing and becoming formal. These disincentives modify the impact of policies to reduce business and labor informality, and to increase capital access in small firms. They also impact social policy, since the poor and vulnerable workers are over-represented in small firms. The framework developed in this dissertation can be used in other developing countries with similar characteristics in tax schemes and credit access, even if information availability is limited at a firm level.

## **Acknowledges**

Most dissertations ideas are the result of a search to fulfill a Ph.D. requirement. In my case, the process was the opposite I decided to finish my Ph.D. in order to develop an idea that was already in my head.

After working several years of studying labor informality with Leonardo Villar, Guillermo Perry suggested me to perform a similar analysis including business informality and taking advantage of the new developments in this field. The task was not easy because neither information nor time were available. This is when when the idea of finishing my Ph.D. (leaved halfway at NYU a long time ago) took shape. Andrés Garcia and Juan Miguel Gallego, gently pushed me in this direction. The Universidad del Rosario was very welcoming and provided me with all the resources needed to develop this task, including the excellent advisory of Juan Miguel and Andrés; and periodic long and insightful meetings with Fernando Jaramillo and Alain Desdoigts. This four professors, and particularly Andrés, were always ready to share their wide knowledge and understanding of the problem of informality.

Outside the Universidad de los Andes, I also received a lot of support. Fedesarrollo allowed me to decrease my workload during several years. Marcela Eslava, Gabriel Ulyssea, Luis Fernando Mejía, Eduardo Lora, Luis Eduardo Arango, Luz Adriana Florez, María Angélica Arbeláez, Jesus Botero, among many others provided me with interesting comments. Cecilia Suescún was key in the editing process. Finally, I would also like to thank Fedesarrollo, DNP, Open Society and UNDP for financing me draft versions of this ideas.

# Contents

<b>Introduction</b>	<b>1</b>
<b>MSMEs and Informality: A new employer–employee database for Colombia</b>	<b>4</b>
<b>1 Introduction and executive summary</b>	<b>5</b>
<b>2 Literature review</b>	<b>8</b>
<b>3 Characteristics of the Business Universe</b>	<b>10</b>
<b>4 Extensive margin of informality or business informality</b>	<b>11</b>
4.1 General characteristics of business informality . . . . .	11
4.2 Rationality of the decreasing behavior of business informality . . . . .	13
<b>5 Intensive margin or labor informality</b>	<b>15</b>
5.1 General characteristics of labor informality (intensive margin) . . . . .	15
5.2 Rationality of the decreasing behavior of business informality . . . . .	17
<b>6 Dualism</b>	<b>18</b>
<b>7 Final thoughts and policy recommendations</b>	<b>21</b>
<b>A Methodology to build the database</b>	<b>26</b>

A.1	Definitions . . . . .	26
A.2	Construction of the business universe . . . . .	27
A.3	Representativeness analysis . . . . .	30
A.4	Imputed observations . . . . .	32
A.5	Value-added of observations obtained directly from the GEIH . . . . .	34
A.6	Firm profits . . . . .	36
 <b>The Impact of Tax Asymmetries on Labor and Business Informality: The Case of Colombia</b>		<b>40</b>
<b>A</b>	<b>Introduction</b>	<b>41</b>
<b>B</b>	<b>Literature review</b>	<b>44</b>
<b>C</b>	<b>A panorama of informality in Colombia</b>	<b>46</b>
<b>D</b>	<b>A model to understand informality with institutional frictions</b>	<b>51</b>
<b>E</b>	<b>Estimation of the model</b>	<b>55</b>
<b>F</b>	<b>Fitness of the model</b>	<b>56</b>
<b>G</b>	<b>Comparative statics</b>	<b>61</b>
G.1	Policies on the extensive margin or business informality . . . . .	62
G.2	Policies on the intensive margin or labor informality in formal firms . . . . .	65
<b>H</b>	<b>Final remarks</b>	<b>67</b>

<b>Appendices</b>	<b>71</b>
<b>Appendix A Sources of structural parameters and first guesses of estimated parameters</b>	<b>71</b>
<b>MSMEs and the Cost of Using Capital</b>	<b>74</b>
<b>A Introduction and literature review</b>	<b>75</b>
<b>B Data sources and facts. The case of Colombia</b>	<b>77</b>
<b>C The model structure</b>	<b>81</b>
<b>D Calibration of the model</b>	<b>85</b>
D.1 Structural and estimated parameters . . . . .	86
<b>E Policy simulations</b>	<b>89</b>
E.1 Eliminating the informal firm's friction in using capital . . . . .	89
E.2 Reducing size firm's friction in using capital . . . . .	92
<b>F Conclusions and policy implications</b>	<b>94</b>

## List of Figures

1	Distribution of firms by number of workers, value added and profits . . .	11
2	Extensive margin of informality (Business informality) . . . . .	12
3	Extensive margin of informality compared to Brazil and Mexico . . . . .	13
4	Net earnings, value added per worker and informality . . . . .	14
5	Labor and business informality . . . . .	16
6	Labor informality Comparisons . . . . .	17
7	Informality and economic subsectors . . . . .	20
8	EEG Information Sources (2019) . . . . .	27
9	Quantile-quantile plot: Labor income . . . . .	34
10	Estimation of the logarithm of the added value with two alternative methods . . . . .	36
11	Estimation of costs other than inputs from formal and informal firms as percentage of value added . . . . .	39
1	Sources of information of the EEG . . . . .	47
2	Extensive margin of informality according to 3 definitions of informality	49
3	Intensive margin of informality . . . . .	50
4	Intensive margin of informality . . . . .	51
5	Firm size distribution (Formal firms) . . . . .	58
6	Firm size distribution of extensive margin, intensive margin and share of skilled workers . . . . .	59
7	Taxonomy of informality . . . . .	60
8	$\theta$ , value added per worker and a taxonomy based on the latter . . . . .	61

9	Microeconomic impact of policies oriented to the extensive margin . . .	64
10	Microeconomic impact of policies oriented to the intensive margin . . .	66
1	Logarithm of capital to labor ratio . . . . .	79
2	Capital stock distributions . . . . .	80
3	Interest rates spreads SMEs/larger firms in the OECD (2020) . . . . .	81
4	MSMEs/SMEs spread in Colombia . . . . .	81
5	Cost of using capital . . . . .	87
6	Distribution of capital fit . . . . .	88
7	Micro effects of reducing the informal firms' distortion in accessing to capital (t=0) . . . . .	90
8	Micro effects of reducing the informal firms' distortion in accessing to capital (t=1) . . . . .	90
9	Micro effects of reducing the microbusiness distortion in accessing to capital (t=0) . . . . .	92
10	Micro effects of reducing the microbusiness distortion in accessing to capital (t=1) . . . . .	93

**List of Tables**

1	Net profits estimation . . . . .	15
2	Relative cost of hiring a worker formally and informally (2019) . . . . .	18
3	Firms below the tax threshold, registered as natural persons . . . . .	18
4	Determinants of salary differences, as an indicator of dualism in the labor market . . . . .	21
5	GEIH vs EEG coverage by sector and occupation . . . . .	28

6	Composition of the EEG database (2019) . . . . .	30
7	EEG vs GEIH. Number of workers . . . . .	31
8	Comparison of EEG(2019) with DANE’s Directory of Firms in terms of number of firms (thousands). . . . .	32
9	Availability of information to estimate productivity at the firm level . .	32
10	Imputation Errors . . . . .	33
11	Comparison EEG and GEIH. Intensive margin and labor income . . . .	34
1	Taxes and contributions faced by businesses (2019) . . . . .	42
2	Number of firms above and below the income thresholds and those reg- istered as “natural persons” . . . . .	42
3	Forces affecting the formality decisions across firm sizes . . . . .	43
4	Structural parameters used in the estimation . . . . .	55
5	Estimated parameters . . . . .	56
6	Estimated moments (observed & simulated data) . . . . .	57
7	Macroeconomic impact of policies oriented to the extensive margin (policy/baseline, constrained model) . . . . .	64
8	Macroeconomic impact of policies oriented to the intensive margin (pol- icy/baseline, constrained model) . . . . .	67
1	Representativeness of data source (thousands) . . . . .	78
2	Structural parameters . . . . .	86
3	Estimated parameters . . . . .	87
4	Model fit . . . . .	88
5	Baseline and simulation parameters for informal firms . . . . .	89

6	Macro effects of reducing the informal firm's distortion in accessing to capital (t=1) . . . . .	91
7	Macro effects of reducing the microbusiness distortion in accessing to capital (t=1) . . . . .	94

## Introduction

Traditional analysis understood business informality as a source of low productivity because of the misallocation of inputs. Therefore, formalizing firms would by itself increase productivity. Maloney (2004) and others controverted this approach by documenting an integrated market, and Perry et al. (2007) argued that there were several types of informality that coexist. Ulyseas (2018), formalized this idea through a model where small and large, and formal and informal firms interact, in accordance to their distributions observed in the data. Until now, a comprehensive analysis of this type was impossible to be implemented in Colombia due to the lack of a business census or periodic survey covering a comprehensive universe of firms.

The first paper contributes to fill this gap by developing a methodology to periodically describe the firm-employee relationship in countries that do not have an economic census or periodic surveys. This methodology was applied to the case of Colombia, and its analysis allowed, for the first time, to identify the main stylized facts of the business universe in the country. This exercise contributes to understand how the concepts of firm size, productivity, and inequality interact in an environment of high labor and business informality plus tax asymmetries between large and small companies.

The findings of this paper, allowed to confirm several facts as the immense prevalence of small firms with very low productivity. According to the EEG (2019), after excluding self-employment (43% of workers) and agriculture, micro business account for 93% of the firms, 31% of the workers and 8.9% of the value added. Also, small firms show a high incidence of poverty and vulnerability among workers and employers. Other stylized facts are high rates of business and labor informality, that are decreasing in the number of workers.

This diminishing behavior of business informality can be explained by larger firms finding more profitable to be formal while smaller firms not finding profitable, or not being able, to be formal. The diminishing behavior of the labor informality rate can be explained by higher enforcement in larger firms plus asymmetries in the tax scheme, that allows larger firms to obtain deductions. More specifically, the after tax relative cost of hiring formally is 10 times higher for small firms that do not pay taxes. In turn, the proliferation of small firms and their low productivity can be partially explained by

the lack of incentives to hire skilled workers and use capital, and therefore to grow.

The context of an asymmetric tax system was formalized in the second paper, by estimating a model based on Ulyssea (2018). The results of this model confirmed the impact of tax waivers and deductions in shaping informality, and found that enforcement plays a lesser role in Colombia than in the case of Brazil, which is plausible.

This paper also contributes to find the impact of policies orientated towards reducing labor and business informality in an environment of tax asymmetries. It was also showed that policies oriented to reduce labor informality have a limited impact and are costly, while policies oriented to reduce business informality are more effective, but do not necessarily reduce labor informality if they are directed towards low-productivity firms. These findings resembles the extensive history of policies trying to reduce informality in the country.

Finally, most of the literature detailed above does not include capital, or assumes that informal firms do not use capital (dualism) leaving unexplained an important side of the story. Introducing capital in the model is not easy, because evidence shows that small firms use very small amounts of capital and large firms use it intensively, and at the same time is important to have the same production function for formal and informal firms. In order to fill this gap, the third paper extends the model described in the second paper for the manufacturing sector by involving capital and a cost of capital use, that follows a logistic function.

This paper helps to explain the impact of policies to increase capital access in an environment of tax asymmetries plus high costs of using capital for small firms and low costs of using capital and interest rates tax deductions for larger firms. According to the results, policies aimed at reducing the cost of capital for informal firms increase productivity in the informal sector; but at the same time reduce the size of the formal sector, affecting overall productivity. Mean while, policies oriented to reduce the cost of using capital for small firms can increase formality, maintaining productivity relatively stable because both tales of the productivity distribution are equally affected.

The policy recommendations that follow from this paper are not easy to formulate because it is difficult to solve the asymmetries that affect small firms without creating larger distortions in the market. Behind all this reflection hinges the concept of

“economía popular”, used by the current administration. This approach understands that formalizing cannot be an objective by itself, and that is more important to formulate policies to increase firm’s productivity irrespective of them being informal. Through this is a sensitive road, it is important that it does not end working as an incentive to be informal, and segmenting the market even more. In order to be successful it also requires more sources of information at firm’s level.

According to the second paper reducing the tax waiver in small firms and adopting a flat profit tax, or simply enforcing firms to register as a business, can be an effective way to reduce labor informality, because it solves the asymmetry of tax deductions. However this policy is difficult to implement because of social, political and enforcing reasons. Other policy recommendations include a comprehensive single tax scheme (SIMPLES) that encompasses social security, or that allows formal labor tax deductions. Single tax schemes have proven to be effective in reducing informality, but also tend to cause dwarfism, and to be costly from the fiscal point of view. An alternative is a scheme of social security deductions vouchers that can be used when firms get larger. A more radical change in the social security institutions, might also be the optimal solution. Similarly, is difficult to formulate policies to increase capital access because it is a costly policy that inevitable would rely on taxpayers, higher intermediation rates and/or a risky environment for the financial sector.

# **MSMEs and Informality: A new employer–employee database for Colombia**

Cristina Fernández\*

## **Abstract**

In Colombia, as in many other countries, there is a lack of comprehensive and recurring information on business demographics. This paper creates a method to generate a periodic firm–employee database that allows for the first analysis of the entire universe of firms in the country. It also creates a methodology suitable for implementation in countries with limited business information. After excluding self-employment (43% of workers) and agriculture, this paper identifies five stylized facts on business demographics in Colombia: 1. Micro-businesses account for 93% of firms, 31% of workers and 9% of the value added. This granularity can be partially explained by a tax waiver on small businesses. 2. The informality of business (from 30% to 88% under the registration or taxes criteria, respectively) is decreasing in the number of employees. This behavior can be explained because small and unproductive firms cannot afford formality, whereas larger firms find profitable to operate formally. 3. Labor informality is high (71% among firms, and 27% among workers) and also decreasing in the number of employees. This diminishing trend, that in turn is related to small firms hiring few skilled workers, can be explained by a stronger enforcement on larger firms, plus asymmetric taxes. The after tax relative cost of hiring formally is ten times higher for small companies. 4. The incidence of poverty in micro-business is three times higher than in larger firms. 5. Overall, there are incentives for firms to remain small, but also incentives for small firms to do not use capital and skilled workers, affecting productivity and poverty.

**Keywords:** Informality, Firm informality, Informal labor market, Taxonomy of informality, Policy recommendations for informality, Firm size distribution, Informality and taxes, Informality and productivity.

**JEL codes:** J46, O17, L11, O47.

---

\*Fedesarrollo researcher. The great support of Andrés García-Suaza in the development of this work is appreciated. María Angélica Arbeláez, Eduardo Lora, Daniel Gómez, Juan Miguel Gallego, Fernando Jaramillo, Alain Desdoigts and Luz Adriana Florez were also fundamental in developing the work. Marlon Salazar and Camilo Ríos contributed to the initial steps of this document, Mónica Ortiz provided helpful comments, and Alexander Sarango and Cecilia Suescún assisted me in the final steps. This work was funded by the United Nations Human Development Report for Colombia. I am also grateful with the National Planning Department (DNP), which financed an earlier version of this document.

# 1 Introduction and executive summary

The research question of this article is what are the main stylized facts of the business universe in Colombia, with emphasis on size and informality, and its relationship with productivity and poverty. The main objective is to create a methodology to periodically describe the firm-employee relationship by generating a business database in countries that do not have an economic census or comprehensive periodic surveys.

In order to achieve this objective, this paper compiles information available from the micro business survey (EMICRON) and the structural surveys carried out in the manufacturing, retail, and service sectors (EAM, EAC and EAS). Missing information was filled by using the household survey (GEIH), which provides representative data at the quarterly level on employers and their businesses, by assuming one employer per firm and one firm per employer, and populations weights of employers equivalent to firms weights. This exercise was carried out for the year 2019, considering that the year 2020 was still affected by COVID-19. The resulting database does not include the agricultural and mining sector, due to low representativeness. Government and domestic services were also excluded because they do not really belong to the corporate universe. Self-employment was left out of the discussion, not because it is not important, but because it is too important (43% of workers) and divergent. In fact, self-employment in developing countries is closer to a substitute for unemployment than a wage work (Donovan, Lu, & Schoellman, 2023), gets heavily affected by factors such as necessity, flexibility, or the existence of a business idea, and tends to show fewer transitions to salaried jobs (Fernández, 2023b). Details on the compiled database (EEG later, by its component initials: EMICRON, Structural, GEIH) can be found in Annex A.

The effort to compile this database is important for three reasons. The first is that, in accordance with the recommendations of the CONPES (Departamento Nacional de Planeación, 2019), the new database allows to understand the behavior of the entire set of Colombian firms, including the smallest and the informal ones. As a example, it was key to estimate (Fernández, 2023a) and (Fernández, 2023c). Additionally, it contributes to setting the foundations for carrying out a new economic census in the country, by estimating expected results, identifying missing information, and providing an alternative to periodic follow ups. The second is that, after performing some imputations, the new database allows for a joint analysis of the economic and social aspects of the productive sector, since most of the observations are linked to the household survey, which is the

source for income distribution indicators. The third is that, the methodology used to generate this database can be applied periodically, not only to the Colombian case, but also to many developing countries with limited firm information available.

The EEG (2019) is not free of limitations, the most important of which is representativeness. Representativeness among firms with fewer than 10 workers is not an issue, since EMICRON is based on GEIH, and household surveys are an efficient way to capture small business. It also should not be an issue among the observations provided by structural surveys because they are supposed to incorporate all the business in selected sectors and size brackets. However, some consistency checks were carried out since the collection of data is not always optimal, and since the compilation of the EEG using the GEIH involved some assumptions. The EEG collects information from about 576 thousand firms, of which nearly 400 thousand are registered. This accounts for 61% of the 650 thousand firms identified as active in Colombia by the Statistical Directory of registered firms in 2019 in the relevant sectors (DEE Dane, 2023). Similarly, excluding self-employment, EEG firms employ 5.4 million workers, 62% of the total number of workers identified by the household survey in a comparable sample.

The second limitation is that, in order to make inferences in regard to labor informality, income distribution and worker vulnerability, it is necessary to make some imputations. These imputations were performed by using non-parametric techniques feed from different training bases, such as the same household database but with regard to questions asked to workers rather than employers (GEIH, 2019); the EMICRON (2019) and the 2021 household survey (GEIH, 2021, 2018 framework), which despite of providing post-pandemic information, asks employees whether the firm where they work is formal. Another variable that was missing was the value added among informal business with more than 10 workers (EEG observations obtained directly from the GEIH). This variable was calculated using the equivalence between value added and input remunerations.

The new database allows, for the first time in Colombia, to identify the main stylized facts of the entire business structure. The first of these facts is the immense prevalence of small firms with very low productivity. According to the EEG (2019), after excluding self-employment (43% of workers) and agriculture, micro firms account for 93% of the firms, 31% of the workers and 8.9% of the value added. The cause of the proliferation

of small firms is still a matter to explore, but common sense suggests that it might be related to the existence of tax-exempt thresholds, the limited availability of human capital, monitoring and control that increases with firm size, high costs of using physical capital (Fernández, 2023c), an inflexible labor market, and even a history of violence which erodes confidence and makes relevant to maintain a low profile to evade crime.

Other stylized facts are a high rate of business and labor informality, that are decreasing in the number of workers but not enough to allow the assumption that all larger firms are formal. This diminishing behavior can be explained by larger firms finding more profitable to be formal (because of access to credit, access to government facilities and demand, ability to export and, visibility among others) while smaller firms not finding profitable, or not being able, to be formal. The diminishing behavior of the labor informality rate can be explained by higher enforcement in larger firms plus asymmetries in the tax scheme, that allows larger firms to obtain deductions that can not be used by smaller firms. More specifically, the after tax relative cost of hiring formally is 10 times higher in firms that do not pay taxes.

This disincentive to hire formally derives into a disincentive to hire skilled workers. On top of this, and according to Fernández (2023c) small firms face higher costs to access capital than larger firms (the average interest rate charged in 2018 was 9% for larger business, 16% for SMEs and 29% for micro-business, according to the OECD (2020)). Having few access to capital and skilled labor, there is no surprise in finding small firms to be unproductive, and their workers and heads, very likely to be poor. In fact, estimations performed using the EEG(2019) find that the incidence of poverty in micro-firms is 32% among workers and 11% among employers, barely 3 times higher than the one observed in larger firms; in accordance to what is indicated in Eslava, Meléndez, and Urdaneta (2021). Likewise, small firms show higher incidence of low-skilled, migrants and other vulnerable workers.

In the analyses carried out along this paper, there seems to be a correspondence between highly productive formal firms that hire formal and more skilled workers, and unproductive informal firms that hire informal and low skilled workers; as well as productivity levels and sectors where only formal firms operate, and productivity levels and sectors where only informal firms operate. However, these trends do not imply that Colombia is a dual economy, according to some exercises performed in this pa-

per. This analysis is important because is consistent with firms transiting organically towards formality as they became more productive, and gives support to policies oriented to increase productivity in small firms as a way of decreasing poverty and increase formality.

The paper is structured as follows. Section 2 presents the literature review; Section 3, the main stylized facts derived from the EEG (2019); Sections 4 and 5, the relationship between firm size and business and labor informality, respectively; Section 6, an analysis of dualism in Colombia and Section 7, a reflection and some policy recommendations. These exercises are an example of the possibilities of analysis offered by the constructed database, but its potential goes beyond the scope of this paper.

## 2 Literature review

Perry et al. (2007), in the leading World Bank report, makes one of the first attempts to understand the relationship between business dynamics and informality at the Latin American level. According to their findings, informal businesses encompass not only small subsistence businesses, but also larger businesses that do not comply with regulations. They also suggest that considerable efficiency gains can be obtained by shifting resources from low-productivity firms to high-productivity firms. In terms of policy recommendations, they argued that some firms may benefit from lowering the costs of informality and react to an increase in the costs of being informal, but they also understood that the best policy for small firms requires actions such as access to credit, formal education, training, and business development services.

On the other hand, one of the most exhaustive exercises to describe the taxonomy of companies in the context of informality was carried out by Levy (2018) for the case of Mexico. According to the author, the main cause of low productivity is the misallocation of resources, closely related to informality and nonsalaried work. To reach this result, he carried out an exhaustive analysis at the company level, which concludes the following about informal companies; 1. They absorb a significant amount of capital and labor. 2. They can be found in all sectors of the economy and throughout the territory. 3. They are not necessarily illegal because most of them hire workers through nonsalaried labor (that is legal in Mexico), and not necessarily completely informal. 4. Most of them are very small (less than 5 workers), but not all informal companies are small. The

proliferation of small businesses generates a proliferation of employers over workers, which is not always optimal. 5. It is a growing force. The composition of economic activity over time has shifted towards the informal sector. 6. They are unproductive. Firms that hire formal workers and/or are formal are more productive, but not all formal firms are more productive than all informal firms.

This analysis leads [Levy \(2018\)](#) to the premise that low-productivity firms absorb more capital and labor than they should, while the most productive do not receive enough resources; and this process is enhanced by the longevity of those unproductive small businesses. Furthermore, unproductive small businesses act in socially inefficient ways. For example, they choose to be informal when they can be formal. Being informal, they remain far from the control of the authorities and, therefore, do not accept cash, do not grow, do not use technology, and do not hire formal workers. In some cases, these firms have high entry rates and low survival rates, which generate short-term jobs. This analysis was then summarized by [Alvarez and Ruane \(2019\)](#) to estimate [Ulyssea \(2018\)](#) for the case of Mexico.

More condensed and equally rich is the description of business informality carried out in [Ulyssea \(2018\)](#), [Ulyssea \(2019\)](#) and [Ulyssea \(2020\)](#) for the case of Brazil. The main findings of these articles are the following: 1) The intensive and extensive margin, or labor and business informality, are decreasing on the size of the companies. 2) Informal companies are on average smaller, run by people with less education, pay lower wages, and are less productive than formal companies, but despite these differences, formal and informal companies in Brazil coexist in all sectors and levels of productivity, and neither there is evidence of the missing middle, another characteristic of dualism in the economy. 3) The wage gap between formal and informal workers, characteristic of segmented economies, since it means that formal and informal workers carry out different tasks in the economy, disappears when controlling for the characteristics of the companies (fixed effects). 4) The dynamic selection process takes place in both the formal and informal sectors, but it is weaker in the latter and, therefore, ex-ante rather than ex-post heterogeneity is the determinant of the dynamics of the company.

Based on all these stylized facts, [Ulyssea \(2018\)](#) developed a model which integrates two views of informality that use to antagonize: the dual perspective of the market ([Harris & Todaro, 1970](#); [Lewis, 1954](#)); and the De Soto margin, according to which

informality is due to high entry costs (De Soto, 1989, 2000). Another advantage of Ulyssea (2018) is that it is a general equilibrium model that considers both the intensive and extensive margin of informality, which do not always react in the same way. However, there are some characteristics of this analysis that make it difficult to apply to the Colombian case. The first of which is the availability and richness of the administrative records, the second is the assumption that all firms larger than ten workers are formal and hire formally, which is not the case of Colombia.

In the case of Colombia, Eslava, Haltiwanger, and Pinzón (2019) analyzed all non-micro formal manufacturing establishments and found that the size distribution exhibits a high concentration of old and small companies; pointing to a higher death rate for high-growth entrepreneurship, and a relatively high probability of long-term survival for small and likely unproductive firms. The authors consider this to be related with the low productivity growth rates observed in the country. More focused on informality, Fernández (2020) made a first attempt to understand the dynamics of business informality, but based on the Survey of Microestablishments (2013-2016), which was not representative. Although these articles made important contributions to understanding business dynamics and informality in Colombia, they were based on partial sets of information and therefore cannot address some of the main questions of this field for an aggregate level. This work seeks to fill this gap.

### 3 Characteristics of the Business Universe

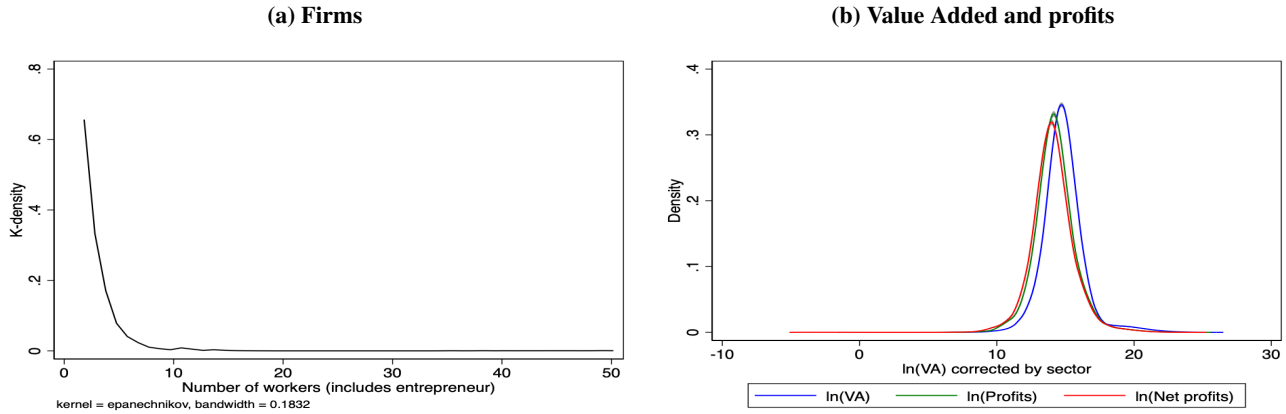
One of the most typical characteristics of emerging economies is the granularity of business demography. According to the EEG (2019), this is also the case in Colombia: excluding self-employment (44% of total employment) and agriculture, companies with 10 or fewer workers constitute 93% of the total firms, 31% of employment (9.4 employees per company) and 9% of value added. As claimed by (OECD, 2021), Colombia shows the highest percentage of micro businesses and self-employment among member countries (the average is 80%). Figure 1a illustrates the distribution of business by number of workers. Figure 1b shows the distribution of business in according to size, value added, profits, and profits net of the cost of keeping informality away from the authorities<sup>1</sup>. According to this figure, the value added exhibits a normal distribution, with

---

<sup>1</sup>All graphs include confidence intervals and are corrected by sector differences. Profits are calculated as the net value added of salary and tax costs; and net profits, as net profits of the costs of operating informally under the control of the authorities. The median monthly earnings of firms with paid employment are close to US\$1,500. To control for sector differences, the prediction errors of a regression between the logarithm of the variable of interest

a long tale to the right, which is reflected in the low contribution of micro-businesses to the GDP (9%). This figure also shows how profits replicate the behavior of value added, but with less extreme data, and lower averages, indicating the impact of taxes, contributions, and enforcement on earnings.

**Figure 1. Distribution of firms by number of workers, value added and profits**



Source: EEG (2019). Self-employment excluded. Value added and profits include confidence intervals

## 4 Extensive margin of informality or business informality

### 4.1 General characteristics of business informality

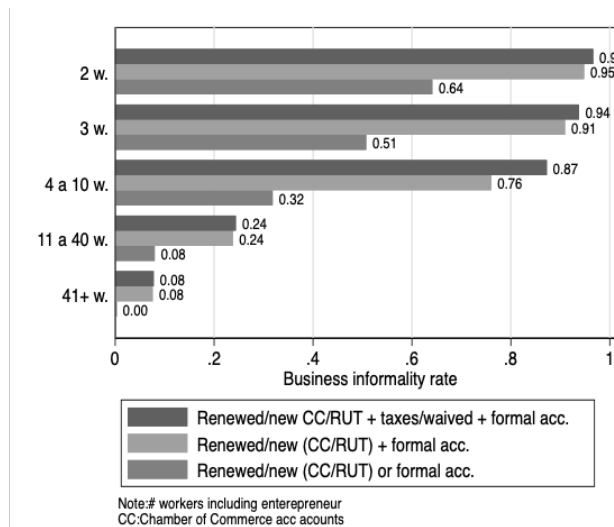
One of the problems when analyzing business informality is the existence of multiple criteria, at the point that one might account for a continuous variable, identifying compliance with a greater or lesser percentage of business regulations. This feature is very different in countries like Brazil where the single-tax (a scheme that considers compliance with a wide range of business regulations, including labor) prevails. Aware of these difficulties, DANE is creating a measurement methodology for business informality that considers four components: entrance, taxes, inputs, and product. Given that the methodology is still in progress, this paper uses the strictest measure of business informality (pays taxes or is exempt from paying them, has a renewed registration in the Chamber of Commerce, and maintains formal accounting) which also allows to assume that formal firms do not evade taxes. However, when the results are sensitive to the definition of informality, additional scenarios are included.

---

and the 3 economic sectors are obtained, and the regression constant is added. To avoid the impact of extreme values, 1% of the right tail of the distribution and negative values are removed from each database involved, since we work with logarithms

The first conclusion that can be drawn from the business informality analysis using the EEG (2019) is that it is widespread. Excluding self-employment, the strict business informality rate is 88%, and 30% of the firms do not have a chamber of commerce registration or are registered to pay taxes (RUT). As shown in Figure 2, business informality decreases monotonically with the size of the firms. However, between 8% and 24% of firms with 10 to 40 workers are informal. There are even some firms with more than 50 workers that are informal, according to the strict definition of informality (8%).

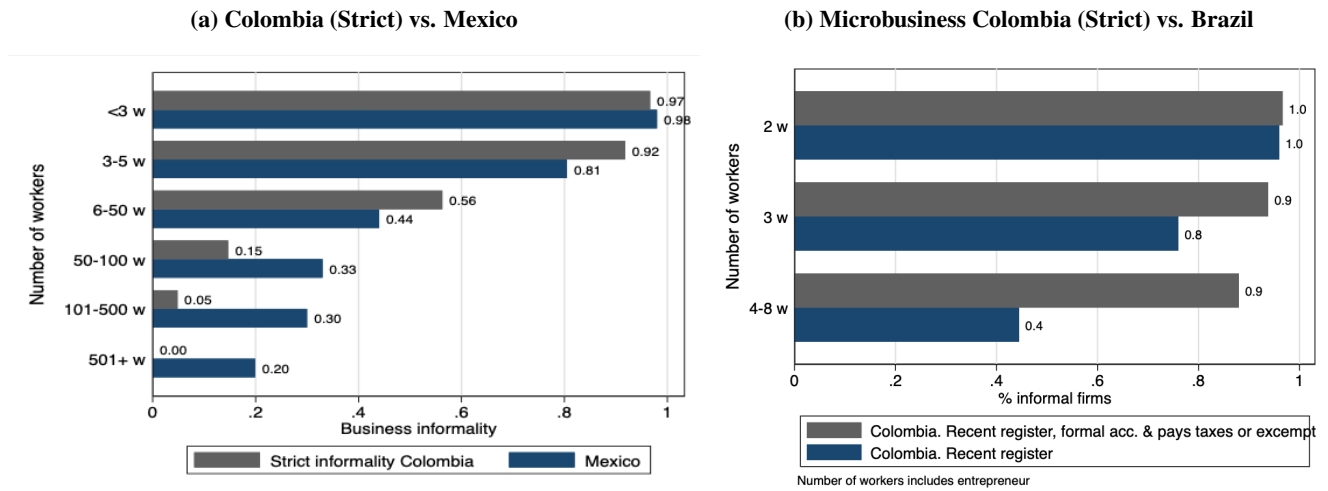
**Figure 2. Extensive margin of informality (Business informality)**



Source: EGG(2019). Self employment excluded.

Compared to Mexico (Figure 3a), Colombia’s rate is higher for businesses within 3 to 50 workers, but is lower afterwards. On the other hand, small Brazilian firms show levels of informality similar to the stricter informality criteria in Colombia, but the informality of Brazilian firms decreases faster with firm size. Therefore, assuming that firms with more than 10 workers are formal might be plausible in the case of Brazil, but not in the case of Colombia.

Figure 3. Extensive margin of informality compared to Brazil and Mexico



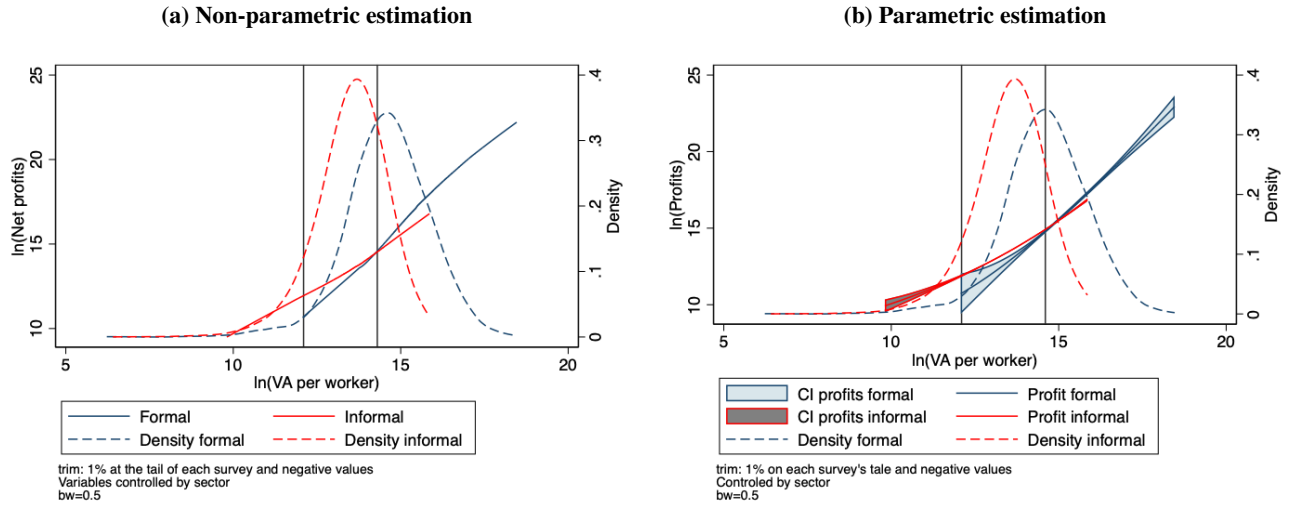
Source: EGG(2019), Alvarez and Ruane (2019) and Ulyssea (2018)

## 4.2 Rationality of the decreasing behavior of business informality

An explanation for the decreasing behavior of business informality on productivity is illustrated in Figure 4, that records the relationship between value added per worker, profits (solid lines) and densities (dotted lines) for formal and informal firms (blue/bold and light/red, respectively). In the first segment of the figure (less than COP\$150 thousand) the profits of informal firms are positive and no formal firms operate; in the second segment (less than COP\$2.2 million) the profits of informal firms are higher than those of the formal and accordingly, the informal firms outnumber the formal; and in the third segment, the opposite occurs.<sup>2</sup> Interestingly enough, firm distributions and profit curves cross at the same productivity level. According to Fernández (2023b), this taxonomy is not reproducible with self-employment.

<sup>2</sup>Ulyssea (2018) performs a similar methodology but divides the middle segment in two, those firms with lower productivity, which he identifies as parasitic, and those that are closer to the point at which it becomes more profitable to be formal, which he identifies as DeSoto firms

**Figure 4. Net earnings, value added per worker and informality**



Source: EEG (2019). Excludes self-employment. Net profits exclude enforcement costs.

The previous results are formalized in Table 1, which illustrates a regression between net profits, added value, and informality as explained in Equation 1. According to the first regression (1), profits grow with the added value per worker, and on average informality tends to be a profitable alternative. However, the impact of informality on profits tends to get smaller as the value added of the firm increases. Columns (2), (3) and (4) indicate that the results are robust to the inclusion of other control variables, and to the inclusion of clustered errors. It is important to note that this result, unlike Ulysea (2018) and Fernández (2023a) that use ex ante productivity, does not imply causality, and is impossible to disguise if formality is causing a higher value added per capita or the opposite.

$$l(Netprofits_i) = c + l(VA_{perworker}_i) + informal_i + l(VA_{perworker}_i) * informal_i + e_i \quad (1)$$

where  $Netprofits_i$  and  $VA_{perworker}_1$  are controlled by aggregate sectors (4 sectors) and  $informal_1$  is a dummy variable.

**Table 1. Net profits estimation**

	(1)	(2)	(3)	(4)
	Profits	Profits	Profits	Profits
Value-added per worker	1.900*** [28.22]	1.745*** [25.11]	1.745** [131.06]	1.745*** [11.24]
Informal firm	9.316*** [8.75]	6.952*** [6.35]	6.925* [28.00]	6.952* [3.25]
Value added per worker* informal firm	-0.634*** [-9.07]	-0.464*** [-6.46]	-0.464* [-27.09]	-0.464* [-2.99]
Robust errors	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Cluster errors informality	No	No	Yes	Yes
Cluster informality and survey	No	No	No	Yes
Observations	17808	17808	17808	17808
R-squared	0.733	0.751	0.751	0.751

Source: EEG (2019). Note: t statistics in parentheses, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Excludes self employment. Population weights are used. Value added and net benefits are in logarithms. The data is trimmed to 1% in each survey and negative VA values are not considered. Control variables: number of workers and four digit CIU

In sum, it is possible to conclude that the informality of business in Colombia is greater and more widespread, and less of a binary condition than in the Brazilian case. It has also been shown that business informality in Colombia decreases with the size of the firms, but this decrease is not as pronounced as in the case of Brazil. For this reason, it is not possible to assume that all companies with more than 10 workers are formal. This evidence emphasizes the importance of Dane extending the EMICRON survey or developing a periodic source of information for larger informal firms. The chapter also provided an explanation for the inverse relationship between firm size and business informality,

## 5 Intensive margin or labor informality

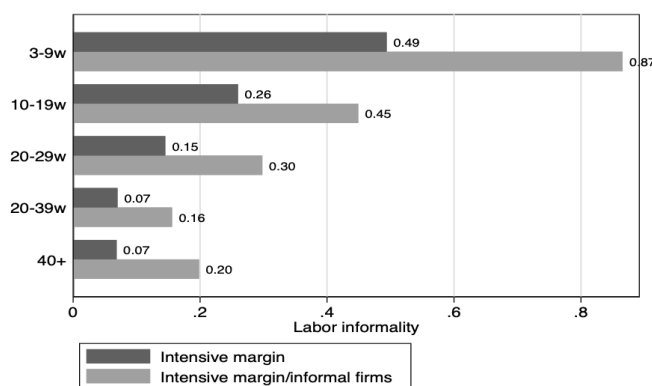
### 5.1 General characteristics of labor informality (intensive margin)

Although labor informality has a more straight-forward identification criteria than business informality, there is a large divergence of definitions between countries based on the internal laws (Fernández, Villar, & Gómez, 2017). The formality criterion used in this work is calculated as of health and pension contributions. The advantage of this criterion is that it allows international comparisons and does not get mixed with other topics such as the size of firms or business informality. The disadvantage is that it might identify exempted workers as informal (legal informality according to Levy (2018)).

Figure 5 shows the informality rate of formal and informal firms. The first observation that can be made from this Figure is that there are no formal firms hiring formal workers and informal firms hiring informal workers, rather the criteria of labor and busi-

ness informality are intertwined. Although the average informality rate in EEG firms is 72%, the average informality rate among formal firms is 33%, and among informal firms, 84%. Formal hiring by informal firms is explained by the fact that, in order to formally hire workers, a firm does not necessarily have to be fully formal due to lack of communication among entities. Figure 5 also corroborates the decreasing behavior of labor informality in the number of workers, but does not allow to assume that all workers in larger firms are formal, since the labor informality rate among firms with 10 to 40 workers is 19% and 6% in firms with more than 40 workers.

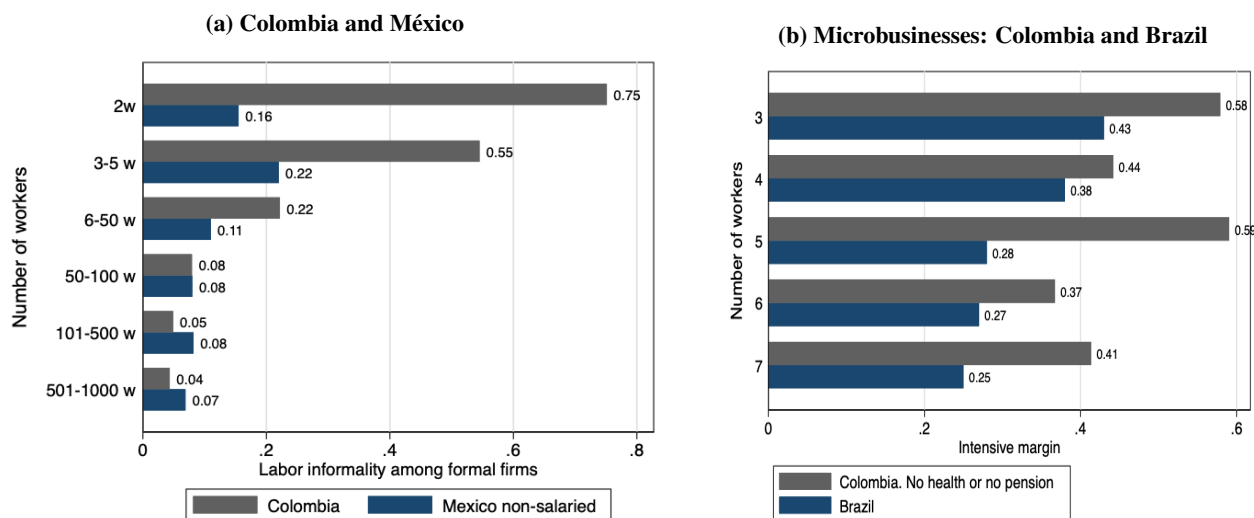
**Figure 5. Labor and business informality**



Source: EEG (2019).

Compared to Mexico (Figure 6a), the informality of labor in Colombia is higher, but this might be related to differences in the used criteria. Compared to Brazil (Figure 6b), informality looks higher and the diminishing pattern less clear. Both statistics indicate that in Colombia it is not possible to assume that labor informality is a problem confined to micro-business.

**Figure 6. Labor informality Comparisons**



Source: EEG (2019), GEIH (2019) and Ulyssea (2018). Thresholds: Income tax: COP\$46 million, VAT: COP\$100 million

## 5.2 Rationality of the decreasing behavior of business informality

The decreasing behavior of labor informality on size can also be explained by the rational decisions of the agents. Table 2 shows the after tax social security contribution for hiring a worker formally instead of informally. For relatively larger formal firms, it makes sense to hire workers formally, because formal hiring costs are tax-deductible, and the amount of tax savings is only 4.7% lower to the cost of social security charged to the employer. However, if the firm has income below the exempted tax threshold (COP\$46 million or US\$14,000 per year) and is registered as a natural person, it cannot make any deductions; and therefore, faces a relative cost of formally hiring a worker of approximately 47% of the salary, ten times higher than in the case of larger firms.<sup>3</sup> This mechanism is reinforced by greater oversight and control exercised by the authorities over larger firms. A more detailed explanation of this mechanism can be found in Fernández (2023a).

<sup>3</sup>For 2023 the main conclusions remaining valid, but the after tax relative cost of hiring formally is being reduced even more because the tax rate has increase to 35%. In this scenario, the obvious question is why larger firms to do not hire all their workers formally. Most likely, this is related with firing cost and a lack of flexibility in formal contracts. Another recent change is the introduction of a single tax scheme (SIMPLES), whose contributions are estimated based on gross income net of social security, a middle step in terms incentives for hiring formally between the firms that do not pay taxes and the larger firms.

**Table 2. Relative cost of hiring a worker formally and informally (2019)**

Income gross annual ( millions )	Income tax rate	Social security cost	Tax deduction	Difference between the cost of hiring a formal and informal worker
< 46	0%	47%	0	47%
> 46	29%	47%	$1.47 \times 29\% = 0.43\%$	$47\% - 0.43\% = -4.7\%$

Source: Own calculations, 2021 tax rates.

This tax-scheme-asymmetry impact is not marginal. According to the EEG (2019) and Table 3, 68% of the firms are below the income threshold, of which 39% are registered at the Chamber of Commerce, and of which 92% are registered under the figure of a natural person.

**Table 3. Firms below the tax threshold, registered as natural persons**

	Total (a)	Shares %	Registered (b)	% Registered in Chamber of Commerce (b/w)	Registered as a natural person (c)	% Registered as a natural person (c/b)
Total	576,316	100%	250,682	43%	217,639	87%
Below the threshold	391,075	68%	154,105	39%	141,075	92%
Over the threshold	185,240	32%	96,577	52%	76,564	79%

Source: EEG (2019) and own calculations. Excludes self-employment.

In sum the intensive margin is easier to analyze because of the availability of data and less diversity of criteria. The downtrend behavior observed in the data can be largely explained by enforcement and tax deductions that are only available to firms large enough to pay taxes. The high cost that face small firms in hiring formally translates into difficulties hiring skilled workers, affecting their productivity.

## 6 Dualism

From the analysis that has been carried out in the previous chapters, it can be deduced that formal firms are larger and more productive, and formal workers are more qualified and earn higher incomes than their informal peers. However, this does not necessarily imply that there are two inherently different economies with few communicating vessels (segmentation or dualism).

Identifying the degree of dualism in the economy is important because the policy recommendations diverge according to this indicator. For example, according to the

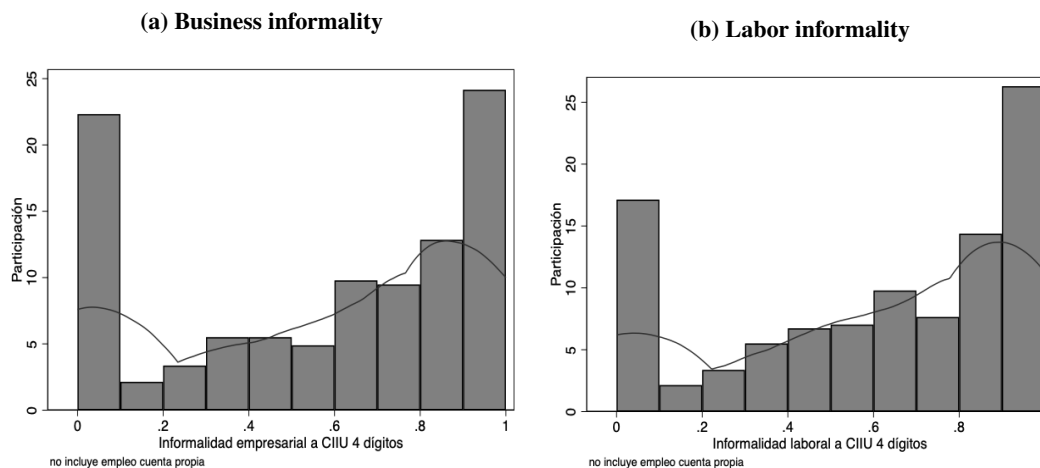
theory of the dual economy, the formalization of firms and workers would by itself improve the productivity, income, and welfare of workers, which is not necessarily valid if the reasons behind the formality decision continue to be valid. On the other hand a less segmented market allows workers and firms to transit organically to formality as they become more productive. In other words, the causality between informality and productivity is at the center of this discussion. The literature has designed several tools to determine the degree of segmentation of the economy. Perhaps the most widely used is the identification of transitions between formality and informality, widely used by [Maloney \(2004\)](#) to establish the little segmentation of the Brazilian and Mexican markets. Unfortunately, it is not possible to replicate the exercise because the household survey in Colombia does not contain panel information.

Another recurring exercise is the identification of sectors and geographical areas that are predominantly formal or informal, used by [Ulyssea \(2018\)](#) in the case of Brazil. Figure 7 replicates this exercise for the Colombian case, showing more segmentation than in the case of Brazil, but not complete segmentation. According to Figure 7a, the proportion of predominantly formal sectors (informality rate less than 10%) is only 22%, and the proportion of predominantly informal sectors (informality rate higher than 90%) is 23%; the remaining 55% are evenly distributed across middle rates of informality<sup>4</sup>. In the case of micro-business in Brazil the sectors not completely formal or informal is close to 70%. In the case of informality of the labor (Figure 7b), the segmentation is slightly less: 16% is predominantly formal and 26% predominantly informal.

---

<sup>4</sup>A flexible criterion of informality is used and a dis-aggregation to four-digit ISIC (376 sectors). When the stricter criterion is used, naturally, the percentage of informal sectors increases to 60%, but the percentage of formal sectors is lower at 10%.

Figure 7. Informality and economic subsectors



Source: EEG (2019). Own calculations. Panels (a) and (b) exclude self-employment.

Likewise, as has been repeatedly illustrated in this work, formal firms are on average more productive and have higher profits than informal ones; and there are segments where only formal or informal firms operate, but the distributions are far from being considered dichotomous because there is a wide range of productivity where the two types of firm coexist (see, for example, Figure 4).

Finally, an exercise used by Perry et al. (2007) and by Ulyseas (2018) seeks to identify whether the differences in worker productivity (using wages as a proxy) are due to the informality of their workers. In implementing this exercise, it is possible to identify not only whether productivity differences are associated with labor productivity, but also with business informality. The results are illustrated in 2 and Table 4. Column (1) of this table shows the exercise carried out with the household survey, where it can be controlled by characteristics of the workers, all some but not all firm's fixed effects. According to the table the informality of labor has an impact of 22% on wages differences. Column (2) estimates the equation without controlling for the worker's education. This omission adds 2 percentage points to the labor informality coefficient. Column (3) makes the same estimates with the EMICRON and obtains a coefficient 4 points higher. Columns (4), (5), and (6) include some firm fixed effects such as the informality of the firm (responsible for 6% of the total effect), the labor informality of the boss and his educational level, causing a significant reduction in the coefficient of labor informality (from 0.28 to 0.22). When full fixed effects are implemented (column 7), labor informality is only responsible for 14% of differences in productivity, not con-

trolling by skill. This indicate that there is not complete dualism in the market but a little more segmentation than in the Brazilian case [Ulyseia \(2018\)](#) where the coefficient of labor informality is not-significant.

$$l(wageinformal_{ij} - wageformal_{ij}) = workertraits_{ij} + firmtraits_j + e_{ij} \quad (2)$$

**Table 4. Determinants of salary differences, as an indicator of dualism in the labor market**

VARIABLES	(1) GEIH	(2) GEIH	(3) EMICRON	(4) EMICRON	(5) EMICRON	(6) EMICRON	(7) EMICRON
Labor informality	-0.2228***	-0.2339***	-0.2800***	-0.2682***	-0.2621***	-0.2165***	-0.1386***
Qualification	0.1427***						
Women	-0.1434***	-0.1362***	-0.1416***	-0.1428***	-0.1428***	-0.1466***	-0.0552***
Age	0.0261***	0.0261***	0.0274***	0.0272***	0.0276***	0.0266***	0.0119***
Age <sup>2</sup>	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0001**
Experience	0.0106***	0.0108***	0.0096***	0.0095***	0.0100***	0.0098***	0.0127***
Bussines informality skill owner +				-0.0867***	-0.0804***	-0.0564***	
Boss labor formality							-0.1408***
2to3 workers	-0.0852***	-0.0860***	-0.2078***	-0.2036***	-0.1994***	-0.1902***	
6to10 workers	0.0838***	0.0529***	0.0188	0.0132	0.0135	0.0061	
Observations	34,431	34,431	17,044	17,044	17,044	17,044	17,044
R <sup>2</sup>	0.262	0.257	0.311	0.312	0.315	0.321	0.939
R <sup>2</sup>	No	No	No	No	No	No	Yes
Firm fixed effects	all	all	all	all	all	all	all

Source: GEIH (2019) and EMICRON (2019).

The performed exercises make it possible to establish that the Colombian economy is more segmented than the Brazilian but it is not a dual economy. However, there is some degree of segmentation and for some firms and subsistence workers it is very difficult to transit into the formal market. For an analysis of segmentation between self-employed and other firms, see [Fernández \(2023b\)](#).

## 7 Final thoughts and policy recommendations

Before generating the recommendations to reduce informality, it is worth asking ourselves the inner reason of this task, because informality cannot be an objective in itself. The most important reasons to reduce business informality are fiscal, control of unfair competition, compliance with other standards including labor and productivity. Reducing informality for fiscal reasons makes little sense because the costs of monitoring small firms may outweigh the benefits in terms of revenue, and on top of that there is a welfare cost. If the reason is control of unfair competition, it must be considered that

although there are some benefits of operating in a small scale, the current regulation has also created inequities that tend to favor the largest and most productive firms. With respect to compliance with other standards, it is important to review the incentives that exist to comply with the different regulations. In the labor case, for example, it is important to consider that an increase in business formality does not necessarily imply an increase in labor formality if the firms to be formalized are small; because for these firms, it is very expensive to formally hire a worker.

Finally, reducing informality due to productivity requires rethinking. First, it must be considered that increases in productivity of small firms do not necessarily imply increases in the overall productivity of the country, because of their reduced participation in the added value; and the limited effectiveness of policies. However, it is possible to remove the constraints in accessing capital and skilled workers, amid other policies to increase productivity among small firms, and in turn, their feasibility of being formal and to provide means to the less favoured.

Some regulatory policy recommendations that imply structural changes are derived from this paper. As indicated by [Levy and Maldonado \(2021\)](#) the institutional apparatus should pass most of the benefit burden, which does not include aspects such as unemployment insurance or professional risks, to general taxes. In implementing this change, is important to consider that the government already finances some of the social security contributions through tax deductions. A first step towards this transition would be the establishment of a universal basic pension for low-income workers who cannot retire<sup>5</sup>, for which it is not necessary to change the private insurance regime. For workers who earn more than one minimum wage, a progressive contribution rate could be established ([Lora, Mejía, Benítez, Delgado, & Gutiérrez, 2021](#)) or at least a flat rate. In the same line, it should be considered the obligation of register any firm with mercantile activity as a firm and not a as a natural person, which is equivalent to the eliminate the profit tax waivers in small firms. This can eliminate the distortion that incentives informal hiring, but probably should be accompanied by another relief to small firms.

Some more operational recommendations include the generation of formal cost-deduction vouchers that firms too small to pay taxes can deduct when they grow. The advantage of this measure is that it encourages the growth of companies, which, as pre-

---

<sup>5</sup>The financing of this pension would not require the availability of RAIS income.

viously observed, is a particular problem in the Colombian case. There are also some recommendations regarding the single-tax scheme. The literature has associated this scheme with the reduction of informality in cases such as Brazil or Uruguay. However, it is important to consider the large fiscal costs and possible incentives for corporate dwarfism. Additionally, it should be kept in mind that little is achieved with the single tax in terms of well-being, if the conditions to comply with other regulations such as labor are not met. In this sense, the inclusion of social security in the single-tax, and/or the establishment of the single-tax rate on income net of salary and social security costs, can be effective solutions. Likewise, it is necessary to design additional mechanisms that encourage the growth of single-tax companies, such as the graduation of companies or decreasing tax benefits.

Another type of policy worth considering in this context is to increase productivity for equity purposes and that according to this paper can be a formality policy itself. Until recently, there was a perception that small companies did not grow and survived for long years with low levels of productivity. A recent study by [Angulo \(2023\)](#) indicates not only that there are some small companies that are growing; rather, it is possible to identify some common factors of these companies such as access to credit, technology, and formality (regardless of the mutual causality between this variable and productivity). However, more studies and databases are needed to be able to design what could be called a social productivity policy.

## References

- Acosta-Henao, M. (2020). *Law Enforcement and the Size of the Informal Sector*. Retrieved from <https://ssrn.com/abstract=3588010> (Draft Paper)
- Alaimo, V., Bosch, M., Kaplan, D. S., Pagés, C., & Ripani, L. (2015). *Empleos para Crecer*. Washington, D.C.: Inter-American Development Bank.
- Almeida, R., & Carneiro, P. (2012). Enforcement of Labor Regulation and Informality. *American Economic Journal: Applied Economics*, 4(3), 64–89.
- Alvarez, J., & Ruane, M. C. (2019). *Informality and Aggregate Productivity: The Case of Mexico*. International Monetary Fund.
- Amarante, V., & Perazzo, I. (2013). Trabajo por Cuenta Propia y Monotributo en el Uruguay. *Revista internacional del trabajo*, 132(3-4), 623–641.
- Angulo, R. (2023). *Índice Multidimensional de Robustez de Micronegocios (IMICRO). Diseño y Resultados para una Estrategia de Focalización de la Secretaría Distrital de Desarrollo Económico de Bogotá*. (Tech. Rep.). Bogotá, Colombia: Inclusión SAS.
- Antón, A., & Rasteletti, A. (2018). *Imposición al Trabajo en Contextos de Alta Informalidad Laboral: Un Marco Teórico para la Simulación de Reformas Tributarias y de Seguridad Social* (Documento para Discusión No. 646). Washington, D.C.: Inter-American Development Bank.

- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5–32.
- Chacaltana, J. (2016). Peru, 2002–2012: Growth, Structural Change and Formalization. *Revista CEPAL*.
- Chisari, O. O., Romero, C. A., & Martínez, J. P. V. (2016). Economy-Wide Effects of Improving Small Firms Access to Capital Markets: An Applied General Equilibrium Assessment. *Revista de Análisis Económico–Economic Analysis Review*, 31(1), 15–36.
- Chowdhury, M. J. A. (2009). *Microcredit, Micro-enterprises, and Self-Employment of Women: Experience from the Grameen Bank in Bangladesh* (Vol. 209; Working Paper). Dhaka, Bangladesh: Center for Microfinance and Development.
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157–175.
- DANE. (2013). *General Methodology: Large Integrated Household Survey – GEIH*.
- DANE. (2020a). *General Methodology: Annual Services Survey – EAS*.
- DANE. (2020b). *General Methodology: Annual Trade Survey – EAC*.
- DANE. (2020c). *General Methodology: Survey Micro-establishments – EMICRON*.
- DANE. (2021). *Economic Census of Colombia: Economic Census Methodological Document*.
- Departamento Nacional de Planeación. (2019). *CONPES 3956 - Formalización Empresarial*. Retrieved from <https://colaboracion.dnp.gov.co/CDT/Conpes/Econmicos/3956.pdf>
- De Paula, A., & Scheinkman, J. A. (2011). The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil. *Review of Income and Wealth*, 57, S8–S26.
- D’Erasmus, P. N., & Boedo, H. J. M. (2012). Financial Structure, Informality and Development. *Journal of Monetary Economics*, 59(3), 286–302.
- De Soto, H. (1989). *The Other Path: The Economic Answer to Terrorism*. New York, NY: Harper & Row.
- De Soto, H. (2000). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Basic Books.
- Dix-Carneiro, R., Goldberg, P. K., Meghir, C., & Ulyssea, G. (2021). *Trade and Informality in the Presence of Labor Market Frictions and Regulations* (Working Paper No. 28391). Cambridge, MA: National Bureau of Economic Research.
- Donovan, K., Lu, W. J., & Schoellman, T. (2023). Labor Market Dynamics and Development. *The Quarterly Journal of Economics*.
- Erosa, A., Fuster, L., & Martinez, T. R. (2023). Public Financing with Financial Frictions and Underground Economy. *Journal of Monetary Economics*, 135, 20–36.
- Eslava, M., Haltiwanger, J. C., & Pinzón, A. (2019). *Job Creation in Colombia vs the U.S.: “Up or Out Dynamics” Meets “The Life Cycle of Plants”* (Working Paper No. 25550). Cambridge, MA: National Bureau of Economic Research.
- Eslava, M., Meléndez, M., & Urdaneta, N. (2021). *Market Concentration, Market Fragmentation, and Inequality in Latin America* (Working Paper No. 11). NY, NY: United Nations Development Programme (UNDP).
- Estrada, D. A., & Hernández Rubio, A. (2019). *Situación Actual e Impacto del Microcrédito en Colombia*. Banco de la República de Colombia. Retrieved from <https://repositorio.banrep.gov.co/handle/20.500.12134/9723>
- Fairris, D., & Jonasson, E. (2016, February). *Determinants of Changing Informal Employment in Brazil, 2000–2010* (MPRA Paper No. 71475). Munich, Germany: University Library of Munich, Germany.
- Fajnzylber, P., Maloney, W. F., & Montes-Rojas, G. V. (2011). Does formality improve micro-firm performance? evidence from the brazilian simples program. *Journal of Development Economics*, 94(2), 262–276.
- Fernández, C. (2022). *Firms, Informality & Institutions. The Case of Colombia* (Documentos de Trabajo No. 020598). Bogotá, Colombia: Universidad del Rosario.
- Fernández, C., Villar, L., & Gómez, N. (2017). Taxonomía de la Informalidad en América Latina. *Coyuntura Económica*, 47(1 y 2), 137-167.
- Fernández, C. (2020). Informalidad Empresarial en Colombia. *Coyuntura Económica*, 50(1-2), 133-168.
- Fernández, C. (2023a). The Impact of Tax Asymmetries on Labor and Business informality. *Forthcomming*.
- Fernández, C. (2023b). *Informality, Productivity & Inequality: An Analysis from the Point of View of Firms and Workers*. (PNUD)
- Fernández, C. (2023c). Microbusiness and the Cost of Using Capital. *Forthcomming*.
- Fernández, C. (2023d). MSMEs and informality: A new employer-employee database for Colombia. *Forthcomming*.
- Fernández, C., & Villar, L. (2016). *The Impact of Lowering the Payroll Tax on Informality in Colombia* (Working Paper No. 015212). Bogotá, Colombia: Fedesarrollo.

- Granda, C., & Hamann, F. (2015). *Informality, Saving and Wealth Inequality in Colombia* (Working Paper No. 575). Washington, D.C.: Inter-American Development Bank.
- 'Haanwinckel, D., & Soares, R. R. (2021). 'workforce composition, productivity, and labour regulations in a compensating differentials theory of informality'. *The Review of Economic Studies*, 88(6), 2970-3010.
- Hamann-Salcedo, F. A., & Mejía, L. F. (2013). *Formalizando la Informalidad Empresarial en Colombia* (Vol. Capítulo 10. Pág.:399-427). Banco de la República de Colombia.
- Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, 60(1), 126.
- Jaramillo, F. (2020). *Informalidad – Margen Intensivo. Mimeo, note=Universidad del Rosario.*
- Jaramillo, M. (2013). *Employment Growth and Segmentation in Peru, 2001-2011: Country Case Study on Labour Market Segmentation* (Employment Working Paper No. 151). Geneva, Switzerland: International Labour Organization.
- Jenkins, S. P., & Kerm, P. V. (2007). *PARETOFIT: Stata module to fit a Type 1 Pareto distribution*. Statistical Software Components, Boston College Department of Economics.
- Kugler, A. D., Kugler, M. D., & Herrera-Prada, L. O. (2017). Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia's Reform. *Economía*, 18(1), 3–40.
- Levy, S. (2018). *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. Inter-American Development Bank.
- Levy, S., & Maldonado, D. (2021). *Misión del Empleo. Reporte Resumen.*
- Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(2), 139-191.
- Loayza, N. V. (2016). Informality in the Process of Development and Growth. *The World Economy*, 39(12), 1856–1916.
- Loayza, N. V., Ulyssea, G., & Utsumi, T. (2018). Informality and the Labor Market Effects of Mass Migration: Theory and Evidence from Syrian Refugees in Turkey. In *7th turkey labor market network meeting* (Vol. 9, p. 2020).
- Lora, E., Mejía, L. F., Benítez, M., Delgado, M. E., & Gutiérrez, D. (2021, April). *Reformas para una Colombia Post-COVID-19. Hacia un Nuevo Contrato Social* (Informes de Investigación No. 019238). Bogotá, Colombia: Fedesarrollo.
- Maloney, W. F. (2004). Informality Revisited. *World Development*, 32(7), 1159–1178.
- Medina, C., & Posso, C. (2010). *Technical Change and Polarization of the Labor Market: Evidence for Brazil, Colombia and Mexico* (Borradores de Economía No. 614). Bogotá, Colombia: Banco de la República.
- Mejía, D., & Posada, C. E. (2007). *Informalidad: Teoría e Implicaciones de Política* (Borradores de Economía No. 455). Bogotá, Colombia: Banco de la República.
- Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6), 1695–1725.
- OECD. (2007). *Eurostat-OECD Manual on Business Demography Statistics*.
- OECD. (2020). *Financing SMEs and Entrepreneurs 2020: An OECD Scoreboard*. OECD Publishing.
- OECD. (2021). *The Missing Entrepreneurs 2021*.
- OECD. (2022). *Financing SMEs and Entrepreneurs 2022: An OECD Scoreboard*. OECD Publishing.
- Paturot, D., Mellbye, K., & Brys, B. (2013). *Average Personal Income Tax Rate and Tax Wedge Progression in OECD Countries* (OECD Taxation Working Papers No. 15). Paris, France: OECD Publishing.
- Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., Mason, A. D., & Saavedra-Chanduvi, J. (2007). *Informality : Exit and Exclusion* (No. 6730). The World Bank Group.
- Rebolledo Abanto, P. A., & Soto Chavéz, R. (2004). Estructura del Mercado de Créditos y Tasas de Interés: Una Aproximación al Segmento de las Microfinanzas. *Revista Estudios Económicos*(11).
- Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8), 2015–2047.
- Ulyssea, G. (2019). *Formal and Informal Business Dynamics*.
- Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics*, 12, 525–546.
- Veyrune, M. R. M., Della Valle, G., & Guo, S. (2018). *Relationship Between Short-Term Interest Rates and Excess Reserves: A Logistic Approach*. International Monetary Fund.
- Vijlberger, W. P. (1992). Measuring income from family enterprises with household surveys. *Small Business Economics*, 4, 287-305.

## A Methodology to build the database

This appendix presents five methodological aspects that were considered for the creation of the database and its respective analysis. The first subsection presents the basic definitions used in the rest of the document, the second contains the guidelines that were adopted to build the database and standardize variables, the third explains the methodology to complete the database with imputations, the fourth presents the estimates of the productivity used in the segments of the new database that did not have this information, and the fifth and last subsection presents the methodology to measure the earnings of workers, when the information was not available.

### A.1 Definitions

Before explaining the methodology used in this paper, it is worth establishing some definitions of variables that will be used from now on.

- Worker: Person that has a job. It can be an employer, employee, or self employment.
- Employee: Person that has a job and a boss.
- Firm: Economic unit that develops a productive activity of goods or services, in order to obtain an income, acting as owner or lessee of the means of production<sup>6</sup>. Only firms with more than one employee are included. Firms with unpaid employees are excluded.
- Self employed: Business that do not have paid employees. This definition is in line with DANE and the OIT. Note that self-employment is not the same as one-worker unit, as a self employer can have several unpaid workers
- Firm size: Number of workers. Includes direct workers, direct contractors, unpaid workers, partners, and employers.
- Formal firm: The most used definition in this work is strict: recently registered firms, with formal accounting, that pay taxes or are exempt to do it. Other definitions used include: i) registered (recently or not) and with formal accounting, ii) registered (recently or not) and iii) registered (recently or not) or with formal accounting.

---

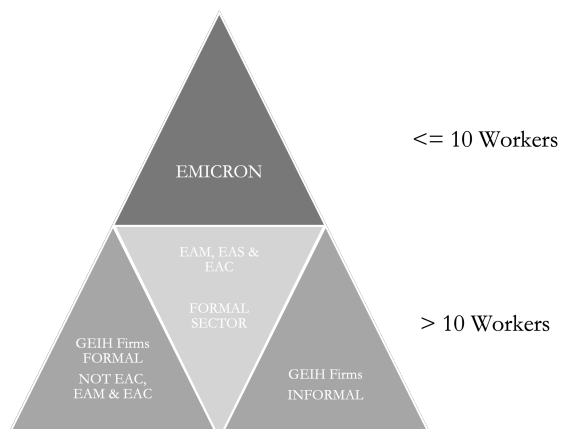
<sup>6</sup>DANE's definition of microbusinesses – EMICRON (2019).

- *Formal worker*: contributes to health and pensions. Informal workers do not contribute to health or to pension.
- *Skilled worker*: worker with tertiary education (some alternative exercises are performed defining skill as secondary or higher education).

## A.2 Construction of the business universe

The EEG (2019) business universe is constructed as follows: companies with fewer than 10 workers are represented by EMICRON; companies with more than 10 workers (11+) are represented by structural surveys (EAM, EAS, and EAC)<sup>7</sup>, and by the independent workers section of the GEIH if they are informal or belong to a sector or size not covered by structural surveys. Indeed, under the assumption that each employer represents a firm, and given that the employer sample is representative of the total number of employers in the country on a quarterly basis, it is possible to assume that the sample of employers is representative of the total number of businesses in Colombia<sup>8</sup>. Note that the questionnaire for each independent worker includes questions regarding the number of employees (paid, unpaid and partners), sector, and earnings, sector among other variables related to the firm. Figure 8 shows the sources of EEG information, and each of the individual sources is detailed below.

Figure 8. EEG Information Sources (2019)



Source: author's own estimates.

<sup>7</sup>Companies with 10 or fewer workers are discarded in these surveys. It is assumed that the companies referenced in these structural surveys are formal, since they filled out their registration number.

<sup>8</sup>These procedure can even capture some stock corporations, since the head of the corporation should declare himself as head of the company. The jobs of the second employers are not considered in the exercise, because the variables that describe the job are different and are generally not available.

- **EMICRON (Microbusiness Survey).** This cross-sectional employee-employer survey is a module of the Household Survey and, therefore, representative at the level of employers in a quarterly basis. Dane adjusted population weights to the universe of micro-entrepreneurs and to the loss of information in the data collection process. Its coverage excludes government business, air transport, the financial and insurance sector, and some geographical areas (DANE, 2020a)<sup>9</sup>
- **Household Survey (GEIH, Dec 2018 – Nov 2019).** This cross-sectional survey is representative on a monthly basis for large aggregates and on a quarterly basis, for less aggregated variables. To match the EMCIRON, the GEIH is adjusted in terms of coverage of sectors and occupations as shown in Table 5 and in terms of months (DANE, 2013)<sup>10</sup>

**Table 5. GEIH vs EEG coverage by sector and occupation**

	Private employee	Government employee	Domestic employer	Self employed	Employer	Other	Total
Agriculture etc	481,465	-	-	1,901,136	169,758	1,110,948	3,663,306
Mining	131,519	-	-	64,509	3,772	655	200,455
Manufacturing industry	1,402,949	2,507	-	894,878	120,292	84,112	2,504,739
Gas, water, electricity	133,911	18,489	-	34,651	1,332	1,475	189,858
Construction	793,616	1,074	-	591,080	85,662	3,797	1,475,229
Retail	1,690,327	-	17	2,075,631	193,279	257,703	4,216,957
Hospitality	525,751	3,852	-	958,729	22,854	8,192	1,519,378
ow. air transportation	22,065	26	-	494	74	9	22,668
Transportation	789,895	153	347	593,120	118,632	127,799	1,629,945
Communications	212,093	4,961	-	92,890	7,284	11,884	329,112
Financial activities	269,081	12,064	-	29,513	2,363	581	313,601
Real state	234,050	19	-	47,738	2,523	623	284,952
Professional activities	626,015	3,503	103	701,182	43,458	11,015	1,385,277
Public administration	1,088,918	813,354	167	640,373	21,848	14,892	2,579,552
ow. government	77,539	368,381	-	202,356	177	1,266	649,720
Artistic activities and other	358,199	3,055	695,974	965,008	36,621	14,549	2,073,406
ow household activities	5,581	34	694,529	32,535	-	258	732,938
<b>Total</b>	<b>8,737,787</b>	<b>863,032</b>	<b>696,608</b>	<b>9,590,437</b>	<b>829,677</b>	<b>1,648,225</b>	<b>22,365,766</b>
Included by sector	7,616,627				652,202	533,032	8,801,861
Not covered by EMICRON					100,827	0	100,827
Firms with one worker	29,900					7,458	37,358
<b>Total included</b>	<b>7,586,727</b>				<b>551,374</b>	<b>525,575</b>	<b>8,663,676</b>
<= 10 workers	2,820,664				502,533	526,445	3,849,642
> 10 workers	4,766,063				48,842	6,587	4,821,491
Not included	1,121,160	1,231,474	694,529	9,825,821	177,475	1,115,192	14,165,652

Source: GEIH (2019) and author calculations. Note 1: cells in gray are not included in EEG. Note 2: Employers not covered by EMICRON can be located in areas not included in the survey. Firms with one worker can be attributed to inconsistencies

- **Annual Manufacturing Survey (EAM, 2019).** This is a structural and exhaustive survey that considers all manufacturing establishments with 10 or more than 10 people employed nationwide

<sup>9</sup>EMICRON (and therefore the GEIH information used) includes micro business in 24 cities and their respective metropolitan areas, but does not include other territories on the periphery.

<sup>10</sup>EMICRON is carried out one month after the household survey, and therefore the GEIH used corresponds to December 2018 and January – November 2019

- **Annual Retail Survey (EAC, 2019).** This is a comprehensive structural survey that considers all retail economic units with 10 or more people employed nationwide, exempting used-goods retailers (auto resales included) (DANE, 2020b).
- **Annual Services Survey (EAS, 2019);** It is a comprehensive and structural survey that considers all formal economic units of service at the national level. Unlike EAC and EAM, the survey coverage cutoff level differs with the subsectors <sup>11</sup>
- **Household Survey (2021), 2018 framework** This database is not used directly, but to estimate variables such as the intensive margin of formality, or the probability that a formal firm or an informal hires vulnerable workers. In fact, despite the fact that this information was collected after the pandemic, it has the advantage of asking each employee if the firm where they work is formal or informal.

The observation unit used in this database is the “economic unit”, in line with what is suggested by OECD (2007). Company size in terms of workers includes direct workers, direct and indirect contractors, interns, unpaid workers, and partners, employers and owners<sup>12</sup>. In addition to sectors not included in EMICRON, such as finance, government, and household services, this analysis excludes the agricultural sector, for the sake of sector representativeness.<sup>13</sup>

The database was generated for 2019, which is the first year for which there is complete information on the required surveys. This information is also available for 2021, but it has limitations in the collection of information and is altered by the effect of COVID-19. In the event of a new Economic Census, the population weights used may be adjusted to ensure the representativeness of the survey. All the household surveys’ information uses the new expansion factors provided by DANE with the new survey framework that has population projections based on the last 2018 census.

Table 6 summarizes the data used in this document<sup>14</sup>: 29.2 thousand observations were used, of which 9.2 thousand come from EMICRON, 19,4 thousand from the structural surveys (EAM, EAC, and EAS) and 441 from the independent workers survey in

<sup>11</sup>A detailed list of cut levels by sectors can be found in DANE (2020c)

<sup>12</sup>This unit generally refers to the establishment, which is consistent with EMICRON, EAS and EAC. However, the information on the number of workers in the household survey and in the EAM refers to the unit that includes branches, etc.

<sup>13</sup>The number of workers in the agricultural sector in the EEG (2019) was only 30% of what is reported in the GEIH in firms with more than 10 workers.

<sup>14</sup>Including self employment with more than one worker, the amount of firms represented accounts for 1 million, and the number of workers, to 6.4 million. This wider data base is analyzed in Fernández (2023b)

the GEIH. These observations represent 576 thousand companies and close to 5.4 million workers.<sup>15</sup> Once the business universe was built, some variables were adjusted to standardized definitions<sup>16</sup>.

**Table 6. Composition of the EEG database (2019)**

		EMICRON >1 worker	EAC, EAS, EAM >10 workers	GEIH (independent workers survey, firms with >10 workers)	Total
Excluding self-employment	#Observations	9,264	19,451	441	29,156
	# of firms	533,602	19,451	23,262	576,316
	# of workers	1,650,729	3,168,968	582,758	5,402,456

Source: EEG (2019). Population weights framework 2018. Self-employment refers to firms with unpaid workers.

### A.3 Representativeness analysis

A first approximation to estimate the representativeness of the EEG (2019) is to divide the analysis in each of the sub-components of this database. The EMICRON sub component does not have a business universe record, because surveys and censuses at this level are costly and ineffective. However, this survey is estimated as a module of the household survey, which is considered a good practice for counting small businesses from the point of view of coverage [Vijberger \(1992\)](#). Likewise, the census nature of the structural surveys (EAC, EAM, and EAS) is supposed to clear existing doubts about their representativeness. Finally, regarding firms with more than 10 workers that are not represented in structural surveys, some doubts can arise from using the weights of independent workers as proxies of the firm’s weights, in spite of the fact that they were designed with a criterion related to availability of information, which also applies here.

To revise the representativeness of the in terms of the number of workers, the following exercise calculates the aggregate number or workers in EEG (2019) by multiplying the number of firms and the number of workers in each firm. Similarly, the total number of workers in the GEIH was restricted to the universe of the EEG as illustrated in [Table 5](#). According to [Table 7](#). there are 5.4 million workers reported in the EEG (2019) that compares to 8.6 million in the GEIH, which means a ratio of 62% By sector, this ratio is calculated at 73% in services, 53% in retail, 57% in manufacturing and 46% in construction. By firm size, a good representativeness of the number of workers is observed in firms with more than 50 workers, but in firms between 6 and 30 workers,

<sup>15</sup>It is important to note that whereas the information collected from the EMICRON and GEIH use population weights, the information in the structural surveys is collected using a census approach, and therefore the population weights are equivalent to 1.

<sup>16</sup>The homologation charts are available upon request.

**Table 7. EEG vs GEIH. Number of workers**

	EEG	GEIH adjusted	EEG/GEIH
Manufacturing	901	1,590	57
Construction	401	875	46
Retail	1,107	2,103	53
Services	2,993	4,096	73
Total	5,402	8,664	62
From 2to3	907	1,897	48
From 4to5	457	998	46
From 6to10	287	918	31
From 11to19	243	557	44
From 20to30	186	506	37
From 31to50	223	405	55
From 51to100	319	433	74
From 101+	2,781	2,950	94
Total	5,402	8,664	62%

Source: EEG (2019), GEIH (2019)

the representativeness is low (45%). Missing information might be due to recollection failures on either on the structural surveys (most likely on medium size firms) or the GEIH (most likely on larger firms).

Comparison of EEG (2019) with other sources of business information at the firm level is difficult due to different delimitation of coverage in terms of sectors and diversity of definitions of the firm (DANE, 2021)<sup>17</sup>. One exception is the Directory of Companies (DEST), collected by DANE, that restricts its coverage to companies with at least one formal employee who performed an economic activity during the year. The number of firms in DEST (648) can be compared with the EEG (2019) by restricting the later to those registered to the authorities (402 out of 576 thousand in EEG), which results in a coverage of nearly 62% as shown in Table 8. This similarity is achieved despite the fact that the data collection procedures and sources are totally different. At a sector level manufacturing shows the larger coverage (92%).

<sup>17</sup>The pre-count of the census that refers to the economic units is 2.5 million firms. However, the data is not comparable to the EEG due to differences in the definition of the firm.

**Table 8. Comparison of EEG(2019) with DANE’s Directory of Firms in terms of number of firms (thousands).**

	EEG	EEG Registered	DEST	Registered EEG/DEST
Retail	170	128	230	55%
Services	219	157	269	58%
Manufacturing	105	75	82	92%
Construction	83	42	67	62%
Total	576	402	648	62%

Source: DEST (2019), GEIH (2019) and EEG (2019, excluding self-employment).

## A.4 Imputed observations

Unlike other exercises that estimate the number of companies, such as the DEST or administrative records that only account for a few characteristics of the firm, the EEG has a good set of variables<sup>18</sup>. Table 9 presents the observed (O), assumed (A), imputed (I), and unavailable (ND) information in the EEG (2019) and in the databases used to estimate the imputed information: 1) the GEIH household survey (2019) in its segment aimed at workers, for firms with more than 4 workers; 2) the EMICRON database (2019) at the worker level, and 3) the GEIH household survey (2021, 2018 framework). Note that the advantage of these two last databases is that they contain simultaneous information on the characteristics of the workers an employers, and the formality and informality of the firms<sup>19</sup>.

**Table 9. Availability of information to estimate productivity at the firm level**

		EEG (2019)					Training databases		
		EMICRON Firms	EAM	EAC	EAS	GEIH Firms	GEIH Workers (2019)	GEIH Workers (2021)	EMICRON Workers (2019)
<b>Worker characteristics</b>	Labor informality	O	I	I	I	I	O	O	O
	Labor income (%)	O	I	I	I	I	O	O	O
	Value added per worker/wages	O	O	O	O	I	O	O	O
<b>Firm characteristics</b>	ISIC (2 digits)	O	O	O	O	O	O	O	O
	Area	O	NA	O	O	O	O	O	O
	# workers	O	O	O	O	O	NA	NA	O
	# worker (range or s)	O	O	O	O	O	O	O	O
	# Salaried, partners & non-remunerated	O	O	O	O	O	O	O	O
Business informality	O	O	O	O	O	NA	O	O	

Source: EEG (2019), GEIH (2019), EMICRON (2019) and GEIH (2021, Framework 2018). O: observed. I: imputed NA: not available

To carry out the imputations of the informality intensive margin, worker’s skills and labor income, artificial intelligence methods were used, and more specifically, the ”ran-

<sup>18</sup>As in Fernández (2023b) this set of variables can be increased to include skill level (%) poverty, income distribution, women (%), migrants (%), youth (%) and other vulnerable populations and using the same methodology presented in this section

<sup>19</sup>The EMICRON contains information about business informality because it is an employee-employer survey, but it is restricted to firms with less than 10 workers, which is an impediment when it comes to estimating what happens in larger firms.

dom forest” method developed by (Breiman (2001) and Cutler, Cutler, and Stevens (2012)). This method estimates the probabilities that a worker has a certain characteristic from decision trees whose nodes correspond to the characteristics of the firms, such as the total number of employees, the geographic area, and the operative sector, its formality status, and if the data is post-pandemic, and using different training databases.<sup>20</sup>. These estimates assume that the probabilities of occurrence of the dependent variables are the same at the level of workers and firms, conditional on certain characteristics of the firms. In other words, it is assumed that the probability of a worker being informal given some firm characteristics is equal to the probability of a firm having an informal worker given some firm characteristics. Table 10 shows robustness indicators of the random forest methodology: the error rate (ERR), the mean absolute error (MAE) and the out-of-bag error (OOB)<sup>21</sup>.

**Table 10. Imputation Errors**

	OOB	Error	MAE
Intensive margin	0.089	0.075	
Labor income	0.362	0.329	0.486

Source: EEG (2019) and GEIH (2019).

Another way, to analyze the results of the imputations is to compare the resulting variable multiplied by the number of workers with the GEIH observed variable. Table 11 shows this comparison using the two-tail unequal variance t-test. As it is shown, both the intensive margin and the labor income are statistically different, /footnotewhich is understandable given that the EEG uses other sources of information, and considers more covariables that the number of workers in the estimation but not necessarily different from the economic point of view.

<sup>20</sup>One advantage of the random forest is that it is possible to have data and variables omitted in some of the training bases, which is a clear difference from a traditional regression, and is key when you want to impute variables using different training bases.

<sup>21</sup>The OOB estimates the difference between the validation score of trees no including in the bootstrapping and the actual score

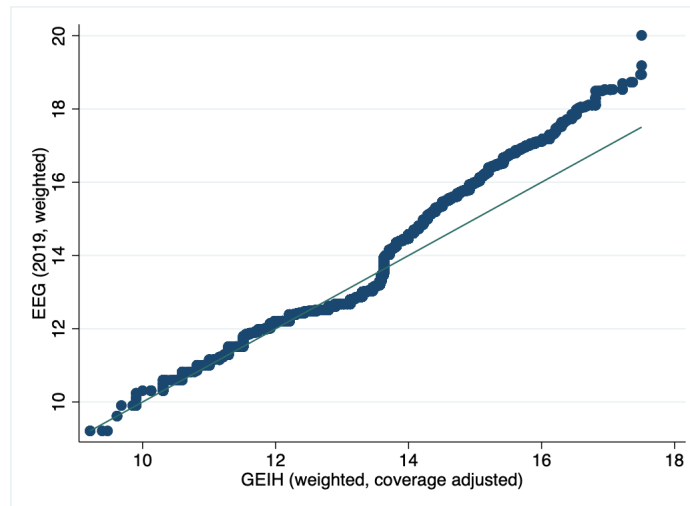
**Table 11. Comparison EEG and GEIH. Intensive margin and labor income**

	EEG	Std dev	GEIH	Std dev	Diff	95% CI	
Intensive margin							
All	0.29	0.403	0.37	0.482	-0.07	-0.08	-0.07
10 - workers	0.80	0.383	0.74	0.437	0.05	0.04	0.06
11 + workers	0.07	0.097	0.11	0.308	-0.03	-0.04	-0.03
Labor income (ln)							
All	14.24	0.967	13.76	0.673	0.48	0.47	0.50
10 - workers	13.35	0.722	13.43	0.690	-0.08	-0.10	-0.07
11 + workers	14.63	0.785	13.94	0.587	0.69	0.68	0.70

Source: EEG (2019) and GEIH (2019).

In terms of income dispersion, Figure 9 illustrates in a quantile plot, how the EEG consistently shows higher wages than the GEIH in the upper side of the distribution, this can be attributed to underestimation from the side of workers (GEIH) or overestimation from the side of employers (Structural Surveys).

**Figure 9. Quantile-quantile plot: Labor income**



Source: EEG, GEIH and author's own estimates.

## A.5 Value-added of observations obtained directly from the GEIH

DANE calculates the added value of the EMICRON firm and in structural surveys as gross production minus intermediate consumption (equation 3). However, this information is not requested from entrepreneurs in the GEIH. To calculate it, this article uses the added value estimate from the point of view of factor remuneration, according to which

gross production is equal to disposable income, which in turn is equal to workers' remuneration, gross surplus (GS), mixed income ( $\Pi_i$ ), and taxes on production (equation 4). Formally, the exercise can be explained as follows:

$$VA_i = GrossProduct_i - IntermediateConsumption_i + NetTaxes_i \quad (3)$$

$$VA_i = FactorRemuneration_i + TaxesOnProduction_i$$

$$VA_i = WorkersCompensation_i + GS_i + MixIncome_i + TaxesOnProduction_i \quad (4)$$

$$VA_i = (1 + \tau_w)w_{if}l_{if} + w_{ii}l_{ii} + GS_{ISIC}VA_i + \Pi_i + \tau_{vat}VA_i$$

$$VA_i(1 - \tau_{var} - GS_{ISIC}) = (1 + \tau_w)w_{if}l_{if} + w_{ii}l_{ii} + \Pi_i$$

$$VA_i = \frac{(1 + \tau_w)w_{if}l_{if} + w_{ii}l_{ii} + \Pi_i}{1 - \tau_{var} - GS_{ISIC}}$$

$$VA_i = \frac{(1 + \tau_w)w_iL^s(1 - P^{li}) + w_iL^s(P^{li}) + \Pi_i}{1 - \tau_{var} - GS_{ISIC}} \quad (5)$$

Where:

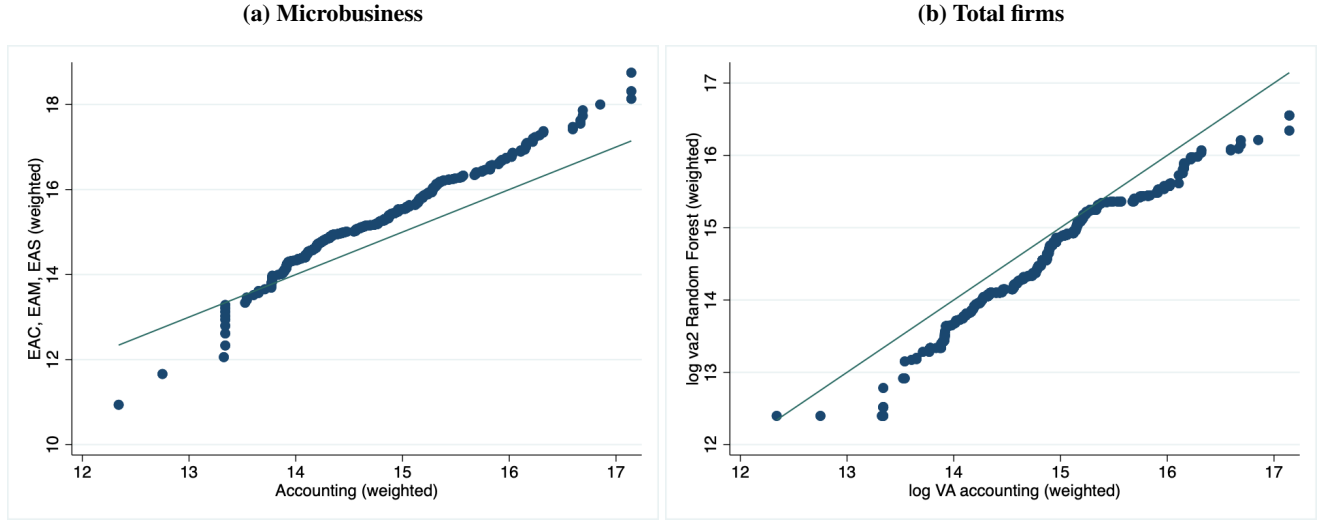
- $t_{vat}$ : value added taxes.  $t_{vat} = 0$  if *AnnualIncome* < 100M.
- $GS_{ISIC}$ : gross operating surplus at the ISIC 4-digit level (imputed as explained below).
- $\Pi_i$ : Mixed income of independent worker earnings.
- $L^s$ : total salaried workers.
- $P^{li}$ : probability of a worker being informal (estimated).

EEG (2019) provides all the variables needed to estimate equation 5, except gross operating surplus (GS), which represents capital income that is not included in mixed income (income earned by the head of the company), but is estimated according to the income-use matrix (DANE). This matrix has 61 groups of activities, which can be correlated with the four-digit ISIC categories<sup>22</sup>. The GS is assumed to be zero in companies

<sup>22</sup>It is assumed that all the ISIC activities that belong to a group contribute to the gross operating surplus in an equitable manner, which is why each ISIC will be charged with the average of the group of national accounts distributed among all the ISIC that belong to each group. one of the 61 categories of National Accounts activities

whose employers earn less than 100 million pesos per month. Figures 10a and 10b show the comparison of the value added and productivity estimated in logarithms between firms with more than 10 workers. Figures 10a compares the accounting method with the labor productivity data observed in the structural surveys (EAC, EAM, and EAS). Figures 10b compares the accounting method with the Random Forest estimation.

Figure 10. Estimation of the logarithm of the added value with two alternative methods



Source: EEG (2019) and own calculations based on EMICRON, EAC, EAM and EAS.

## A.6 Firm profits

Firms' gross profits can be estimated by using the variables of: sum of value added, VAT, income tax net of deduction of formal payroll (as is the case in Colombia) and formal and informal payroll, as illustrated in equation 6<sup>23</sup>.

$$\begin{aligned}\pi^{f,i} &= (1 - \tau_y) \left\{ (1 - \tau_{vat}) VA_i - (1 - t_w) w_i l_i^f \right\} - w_i l_i^i \\ \pi^{f,i} &= (1 - \tau_y) \left\{ (1 - \tau_{vat}) VA_i - (1 - t_w) w_i L^s (1 - P^{li}) \right\} - w_i L^s (P^{li})\end{aligned}\quad (6)$$

where

- $\pi^{f,i}$ : profits for formal or informal firms.

<sup>23</sup>Partner earnings are assumed to be part of profit and unpaid workers receive no income.

- $\tau_y$ : income tax (28%), 0% if the firm earns less than 46M per year or the firm is informal.
- $\tau_{vat}$ : VAT (19%), 0% if the firm earns less than 100M per year or the firm is informal.
- $t_w$ : social security contributions (47%)
- $w_i$ : average salary of the firm for salaried workers.
- $l_i^f$ : formal workers.
- $l_i^i$ : informal workers.
- $L^s$ : total salaried workers.
- $P^{li}$ : intensive margin, probability that a worker is informal in the firm (estimated in EMICRON and GEIH and assumed in structural surveys).

Following [Ulyssea \(2018\)](#), it is also possible to estimate profits after deducting the implicit cost of enforcement face by informal firms. Equation 7 illustrates the estimation of firm profits net of the costs implied by the monitoring and control activities by the authorities. Like in [Ulyssea \(2018\)](#), this is a cost function, quadratic on informal labor, that reflects the increasing cost of remaining out of sight of the authority, mediated by a parameter specific to the type of business and informality of the labor<sup>24</sup>. Note that, unlike [Ulyssea \(2018\)](#), informal firms are allowed to hire formal workers. This is particularly important if the informality criterion is strict or includes the payment of taxes. In other words, the firm can hire formally even if it is not registered in the Chamber of Commerce, or does not pay taxes or is exempt from them.

---

<sup>24</sup>For this calculation, the used parameters were obtained from [Fernández \(2022\)](#).

$$\pi^f = VA_i - (1 - \tau_w)w_iL^s(1 - P^{li}) - \frac{1 - L^s(P^{li}P^c)}{b_c}w_iL^s(P^{li}P^c) - \quad (7)$$

$$L^s \frac{P^{li}(1 - P^c)}{b_{nc}}w_iL^sP^{li}(1 - P^c)$$

$$\pi^i = (1 - \tau_y) \left\{ (1 - \tau_{vat})VA_i - (1 - \tau_w)w_iL - s(1 - P^{li}) \right\} - \frac{1 - L^s(P^{li}P^c)}{b_i}w_iL^s(P^{li}P^c) -$$

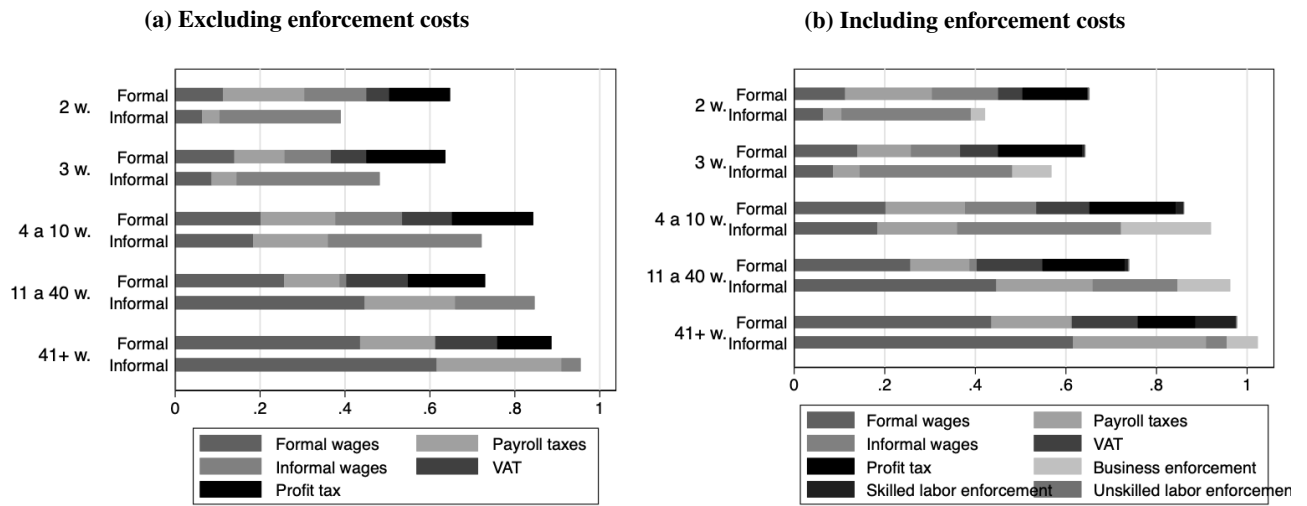
$$L^s \frac{P^{li}(1 - P^c)}{b_i}w_iL^sP^{li}(1 - P^c)$$

where

- $\pi_i^f$ : earnings of formal firms.
- $\pi_i^i$ : earnings of informal firms.
- $VA_i$ : value-added.
- $b_c$ : enforcement cost parameter, skilled workers, higher = lower cost.
- $b_{nc}$ : enforcement cost parameter, unskilled workers, higher=lower cost.
- $b_i$ : enforcement cost parameter, informal firms.
- $P^c$ : probability of having a skilled worker.

Figures 11a and 11b show the cost structure of firms: wages, regulatory and tax costs of formal and informal firms without including and including the costs that the monitoring and control of the authorities entail for the firm, as a percentage of value added. As is possible to observe in the results, the costs of operating formally for smaller firms are higher, while the costs of being informal for larger firms are higher; but enforcement moves the switching point to a smaller firm size.

**Figure 11. Estimation of costs other than inputs from formal and informal firms as percentage of value added**



Source: EEG (2019). Self-employment excluded. w refers to number of workers

# **The Impact of Tax Asymmetries on Labor and Business Informality: The Case of Colombia**

Cristina Fernández\*

## **Abstract**

This paper illustrates how two well-intended policies to reduce informality, such as the income tax waiver for small firms and the income tax deduction of labor cost, end up generating a large number of small firms hiring workers, most likely unskilled, without a formal contract. This paper also shows the difficulties in reducing informality in the complex regulatory environment of Colombia. Policies oriented to reduce labor informality have a limited impact and are costly from the fiscal point of view; policies oriented to reduce business informality are more effective, but do not necessarily reduce labor informality if they are directed towards low-productivity firms. The methodology used to illustrate these facts was the estimation of [Ulyssea \(2018\)](#) for the case of Colombia, in an environment of asymmetric tax and social security regulations.

**Keywords:** Informality, Firm informality, Business informality, Informal labor market, Taxonomy of informality, Policy recommendations for informality.

**JEL codes:** D22, D58, E24, J21, J46, O17.

---

\*Economics Ph.D. student at Universidad del Rosario and researcher of Fedesarrollo. I would like to the invaluable help of Andres García-Suaza, Fernando Jaramillo, and Juan Miguel Gallego in developing this paper. Comments by Gabriel Ulyssea, Luis Eduardo Arango, and Luis Fernando Mejía were also very helpful. This working paper was funded by the Colombia Científica-Alianza EFI Research Program, with code 60185 and contract number FP44842-220-2018, funded by The World Bank through the call Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology, and Innovation. I also would also like to thank the Departamento Nacional de Planeación (DNP) for financing a previous version of this paper.

## A Introduction

One of the first questions asked when approaching the analysis of the Colombian productive sector and the labor market is why there are so many small, low-productivity firms hiring informal workers. One usual explanation is the stronger enforcement on larger firms, but this explanation falls short when considering the weak enforcement institutions in Colombia. An alternative hypothesis is the existence of regulations that incentive this behavior. The research question of this paper is how the complex set of tax, social security, and labor policies in Colombia shapes business and worker informality throughout the size distribution of firms.

The main hypothesis of this article is that two main features of the institutional framework are causing this firm behavior: First, there is an income threshold under which firms are exempted from paying taxes. In Colombia, businesses are not required to pay income taxes if they earn a yearly gross income lower than COP\$46 million (nearly US\$14,000 of 2019)<sup>1</sup>. There is another threshold for exempted VAT at COP\$100 million (nearly US\$26,000). Second, there is the possibility of deducting formal labor costs, including wages and payroll taxes, from the income tax base. Due to this deduction, income tax payers are almost indifferent between hiring workers formally and informally (the effective rate after this deduction is nearly 4.7%), while exempt firms face higher incentives to hire informally because they do not get the tax exemption (the effective rate is nearly 47%, ten times higher).<sup>2</sup> This incentive, which has been relatively ignored in the literature, proves to be an efficient policy to promote formal hiring among taxpayers.

Table 1 summarizes the labor contributions after receiving the tax deduction faced by businesses paying or not income tax and assuming a 28.8% income tax rate ( $\tau_y$ ), and a 19% VAT tax rate ( $\tau_{vat}$ ). According to this table, taxpayers face VAT and income taxes, but lower employer contributions; whereas smaller firms do not pay VAT and income taxes but face a higher relative cost of hiring formally vs. informally<sup>3</sup>.

---

<sup>1</sup>The threshold to pay income taxes is not fixed and depends upon the percentage of expenditures reported by the entrepreneur. This threshold exists only for businesses registered as natural persons, but most businesses do not have incentives to register as legal entities. Also, for simplicity, other income deductions are ignored. This assumption is supported by the fact that a firm must meet a series of requirements to validate these deductions

<sup>2</sup>In 2023 the same arguments continue to be valid but differences are larger because the tax rate increased to 35%. Also, a single tax scheme was implemented that allow deductions of social security payments

<sup>3</sup>It might be easier to understand the problem with an example: A formal firm with an income of COP\$1000 and a wage bill of COP\$500. This firm should pay a payroll bill of COP\$237 ( $\tau_w = 47\%$ ) and having a tax base of \$235 (income wage bill payroll) and should pay an income tax of COP\$76 ( $\tau_y = 28.8\%$ ). Without formal deduction for labor costs, the tax base will be (COP\$1000) and the tax, is COP\$288. The value the firm saves in taxes for hiring workers formally is (COP\$288-COP\$76=COP\$212), and the value of the payroll bill net of tax deductions is equal to (COP\$237-COP\$212=COP\$23),

Although the COP\$46 million tax threshold seems low, it covers an important percentage of micro-businesses in Colombia, which in turn accounts for around 93% of all firms in the country, excluding agriculture and self-employment. As shown in Table 2, and according to the Colombian Micro-Business Survey (EMICRON), 77% of the micro businesses are below the income and VAT thresholds; 10% above the income tax threshold but below the VAT threshold, and 13% above both thresholds. Table 2 also shows that 86% of small formal micro-businesses are registered as 'natural persons' and, therefore, suitable to be exempted from income tax.

**Table 1. Taxes and contributions faced by businesses (2019)**

Gross income Yearly \$COP millions	VAT $\tau_{vat}$	Income tax $\tau_y$	Social security $\tau_w$	Tax deduction $(1 + \tau_w)\tau_y$	Relative cost of hiring informally: social security and tax deductions $\tau_w - (1 + \tau_w)\tau_y$
< 46	0%	0%	47%	0%	47%
> 46 & < 100	0%	29%	47%	$147\% \times 29\% = 42\%$	$47\% - 42\% = 4.7\%$
> 100	14.5%	29%	47%	$147\% \times 29\% = 42\%$	$47\% - 42\% = 4.7\%$

Source: Own calculations. Note: VAT includes income tax deductions.

**Table 2. Number of firms above and below the income thresholds and those registered as “natural persons”**

	Total	%	Informal	%	Formal	%	Formal firms registered as “natural person” (%)
Microbusiness	1,411,670	100%	1,370,133	100%	41,536	100%	73%
<46	1,087,547	77%	1,077,821	79%	9,725	23%	86%
>46 & <100	145,088	10%	137,608	10%	7,480	18%	87%
>100	179,033	13%	154,704	11%	24,329	59%	63%

Source: EMICRON and own calculations. Excludes firms with one worker. Informal: not registered recently or not formal accounts.

Table 3 summarizes the hypothesis on how these two institutional characteristics (tax waivers and tax exemptions) interact with the minimum wage and enforcement policies to encourage or constrain extensive and intensive informality margins. Larger firms tend to show a lower intensive margin since they can deduct formal labor costs and face high enforcement policies, and a lower extensive margin, resulting from both high enforcement policies and taxes. On the other hand, smaller firms face high incentives to both be informal and hire informally, because of the high formal labor cost. They are

that is equivalent to a tax rate of 4.8%

also more affected by the minimum wage because small firms don't get the incentive to deduct costs, and because they also tend to hire more unskilled workers. These mechanisms shape not only the distribution of informality but also the distribution of firms, productivity, and labor through different firm sizes.

**Table 3. Forces affecting the formality decisions across firm sizes**

Gross Income	Intensive margin Labor informality	Extensive margin Business informality
Below the threshold	<ul style="list-style-type: none"> <li>• Low control and enforcement</li> <li>• High labor costs</li> </ul> <i>High incentives to informal hiring</i>	<ul style="list-style-type: none"> <li>• Low control and enforcement</li> <li>• No taxes</li> </ul>
Above the threshold	<ul style="list-style-type: none"> <li>• High control and enforcement</li> <li>• Low labor cost</li> </ul> <i>High incentives to formal hiring</i>	<ul style="list-style-type: none"> <li>• High control and enforcement</li> <li>• Income tax and VAT</li> </ul>

Source: Own hypothesis.

The methodology used to test this hypothesis is to estimate [Ulyssea \(2018\)](#), a model that has been shown to accurately replicate the Brazilian informality case, with and without these three institutional constraints. According to the results, both models well replicate the distribution of the number of firms, the intensive and extensive margin, and the share of skilled workers in different firms for the case of Colombia. However, the unconstrained model accommodates a higher informality in the country by assuming a lower enforcement on business formality, and the constrained model by assuming a lower enforcement on both business and labor informality. The lower role of the enforcement mechanism in the new model suits well the case of developing countries, where enforcement institutions tend to be rather weak.

This contribution has important policy implications because it moves the focus of recommendations from payroll taxes and enforcement to a holistic analysis of the tax, labor, and social security institutions. This shift is welcomed, since payroll taxes policies have shown positive but limited effects on informality; and enforcement policies applied to small businesses can be ineffective, and even harmful for employment and welfare. Comparative statics also allows us to observe the impact of fiscal and labor policies on informality. Policies oriented to reduce labor informality have a limited impact and are costly from the fiscal point of view, and policies oriented to reduce business informality are more effective, but do not necessarily reduce labor informality if they are directed towards small firms.

Although the model is estimated for the Colombian case, the main conclusions of this paper can be applied to a variety of developing countries, which typically face high

informality and have institutional constraints in the form of minimum wages, formal labor cost tax deductions, and tax brackets for low-income firms (Paturot, Mellbye, & Brys, 2013).

Self-employment continues to be an unfilled gap in explaining informality and a pending item on the agenda. In fact, I have realized that this topic, although tremendously important for the Colombian case, should be analyzed with an occupational choice model as explained by Fernández (2023b)), rather than with a firm's perspective model as Ulyssea (2018), and therefore is left for future research.

This paper is structured as follows. Section B briefly reviews the existing literature, Section C shows the main facts of informality in Colombia, Section D presents the theoretical model that results from introducing institutional arrangements to the model of Ulyssea (2018), sections E and F estimates the model and shows the fitting for the Colombian case, Section G illustrates some comparative static exercises and Section H concludes.

## **B Literature review**

Historically, the analysis of informality has debated between the segmented and integrated view. The segmented view, led by Lewis (1954) and Harris and Todaro (1970), advocates for two different markets with few transitions between them: one including low-educated and low-productivity firms and the other high-educated and high-productivity firms. The integrated view led by Maloney (2004) stand up for a sole market where firms and workers freely transit between formality and informality. However, some others, such as De Soto (1989) and De Soto (2000) support a third view, where firms and workers transitions are limited by excessive regulation. Recently, informality has been understood as a phenomenon that encompasses the three different types of informality mentioned above, depending on the size and distribution of productivity of firms (Fernández et al., 2017; Loayza, 2016; Perry et al., 2007; Ulyssea, 2018).

According to Perry et al. (2007), informal firms include small subsistence firms, regulations-restricted firms, and firms of a larger size that do not comply with regulations. They argue that some firms can benefit from lowering the costs of informality and react to an increase in the costs of being informal, but they also understand that small

firms require policies such as access to formal credit, training, and business development services. [Loayza \(2016\)](#) reinforced this framework and policy recommendations with a model calibrated for Latin American countries, where the type and amount of informality is closely related to the cost of formality, which, in turn, harms productivity.

Although these models and analysis stress the importance of lowering the cost of formality, most empirical analysis has found a limited and sometimes non-existent impact of policies such as payroll taxes and registration cost cuts. A positive example is the reduction in payroll taxes from 29,5% to 16% that occurred in Colombia, 2012. This measure reduced the informality rate of dependent workers by 4.8 pp., which is consistent with a 2.4 pp. reduction in the informality rate of the economy ([Fernández & Villar, 2016](#)). Similarly, [Kugler, Kugler, and Herrera-Prada \(2017\)](#) found that this reduction increased formal employment in more than 3% and reduced informal employment in 2.9%. However, the reduction in Peru's labor cost of micro-businesses from 54% to 17% of mean wages did not generate a significant reduction in informality when controlling for growth ([Chacaltana, 2016](#)). According to [M. Jaramillo \(2013\)](#), this could be related to lack of enforcement.

[Ulyssea \(2018\)](#) shifted the focus of the analysis from the cost of formality to the cost of informality. He developed a model that departs from [Melitz \(2003\)](#) but includes an internal and external margin of informality. When applying this model to the Brazilian case, the author found that while reducing payroll taxes can be an effective policy recommendation, enforcement is the main driver of the shape of informality in the size distribution of firms and the most effective variable to reduce informality. This emphasis on enforcement could have been inspired by the Brazilian case. According to ['Haanwinckel and Soares \(2021\)](#), Brazil increased the percentage of workers supervised by labor inspectors by 34% between 2003 and 2012, along with a new incentive scheme. Impact analyses of this increase in enforcement are mixed: [Fairris and Jonasson \(2016\)](#) found a positive impact and [Almeida and Carneiro \(2012\)](#) did not find an impact on formality.

Another possible explanation for the limited impact of reducing informality costs is the interaction of these policies with tax and labor regulations. [Antón and Rasteletti \(2018\)](#) included a more detailed structure of the tax institutions in their model for the case of Mexico, but their analysis was more focused on the fiscal impact of formalization policies. [Dix-Carneiro, Goldberg, Meghir, and Ulyssea \(2021\)](#) formulated a model with tax and minimum wage frictions, but in this case the model was oriented to analyze the impact of trade. [Acosta-Henao \(2020\)](#) analyzed the relationship between tax rates and an enforcement rate, whose effectiveness depends on the size of the informal sector (measured by its assets) and government expenditure. According to his findings, this enforcement perspective creates a Laffer curve, where an increase in taxes can reduce informality under specific circumstances.

The single-tax schemes literature tends to consider a richer institutional framework, particularly in the cases of Brazil and Uruguay, where the single-tax includes social security contributions. [Alaimo, Bosch, Kaplan, Pagés, and Ripani \(2015\)](#) estimated that these schemes formalized about half a million Brazilian micro-businesses and two million jobs between 2000 and 2005; and together with [Fajnzylber, Maloney, and Montes-Rojas \(2011\)](#), found that the firms that opted for the single-tax scheme have shown higher levels of income, profits, hire more workers, and are more capital intensive. According to [Amarante and Perazzo \(2013\)](#), the single tax policies in Uruguay increased the number of formal self-employed workers from 6% to 23% between 2006 and 2010.

In the case of Colombia, [Hamann-Salcedo and Mejía \(2013\)](#) and [Mejía and Posada \(2007\)](#), through a general equilibrium model, found that regulation institutions act as a legal barrier to the entrance of formal firms into the market. Similarly, the recent Employment Mission by [Levy and Maldonado \(2021\)](#) suggested that the behavior of employment and informality in Colombia was mainly due to the complicated labor and social security systems, but did not elaborate much on their interaction with the tax system. This paper aims to fill this gap.

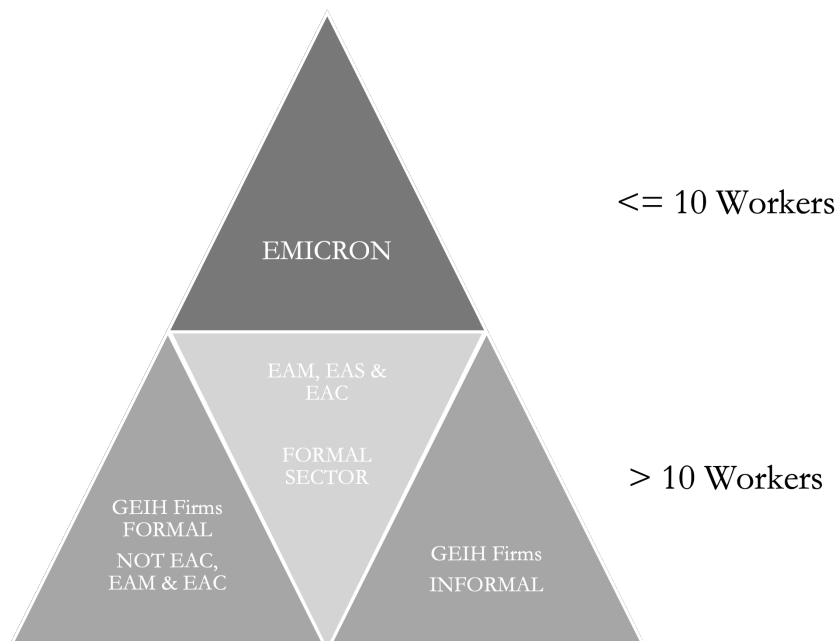
## **C A panorama of informality in Colombia**

One of the limitations in analyzing informality in Colombia, and particularly business informality, is the lack of a business census or a representative survey to simultaneously analyze the entire universe of firm, including informal and informal firms, and micro-business and larger firms. The last economic census was conducted in 1990. [Fernández](#)

(2023d) filled this gap by compiling information on nearly half million firms from EMICRON (employer-employee microbusiness survey), the structural surveys of manufacturing, services, and trade (EAM, EAS, and EAC), and the section of the household survey (GEIH) that asks questions to entrepreneurs<sup>4</sup>.

Figure 1 shows the procedure used to build this new database (EEG: EMICRON, structural and GEIH surveys): information of firms with ten workers or fewer is taken from EMICRON; firms with more than 10 workers are represented by the structural surveys (EAM, EAS, and EAC) and by the GEIH if they are informal or belong or are not covered by the structural surveys. The set of variables collected through this procedure is rather wide and allows generating a good characterization of informality in Colombia. A representativeness analysis of this survey, as well as other features, can be found in the original article.

Figure 1. Sources of information of the EEG



Source: Fernández (2021).

The EEG allows the identification of the following stylized facts that should be considered when modelling informality for the case of Colombia.

**1. Extensive margin of informality.** Business informality in Colombia is high and more continuous than binary, in the sense that there are firms that represent a larger or

<sup>4</sup>This paper uses the version of EEG (2019) that includes self-employment, because otherwise the size distribution of firms would be biased.

smaller portion of the regulations. This is unlike the case of Brazil, where the single tax scheme embodies a large part of the legislation, generating a more discrete business informality variable.

Although it is difficult to estimate and model a continuous variable of business informality, it makes sense to use different informality scenarios in the analysis. Figure 2 shows the extensive margin, according to three different definitions of business informality. As shown in the figure, informality is widely spread among small firms and across all informality definitions, even after excluding one-worker firms. The informality of business is also of considerable size for companies with more than 10 workers (42% among firms with 10 to 40 workers). Therefore, assuming that these firms are formal, as in Ulyssea (2018), is not realistic for the Colombian case. Under this criteria the business informality rate in Colombia is 88% but it can be as low as 30% if we consider a lighter criteria.

It is very important to have the perspective of business informality for several definitions of the variable, but for the specific purpose of this paper, formal firms are restricted to those that pay taxes or are exempted and, therefore, it is more accurate to have a strict definition of informality considering three criteria: Renewed or new registration with the Chamber of Commerce<sup>5</sup>, formal accounts<sup>6</sup>, and being tax contributors or exempted<sup>7</sup> (lighter bar in the graph).

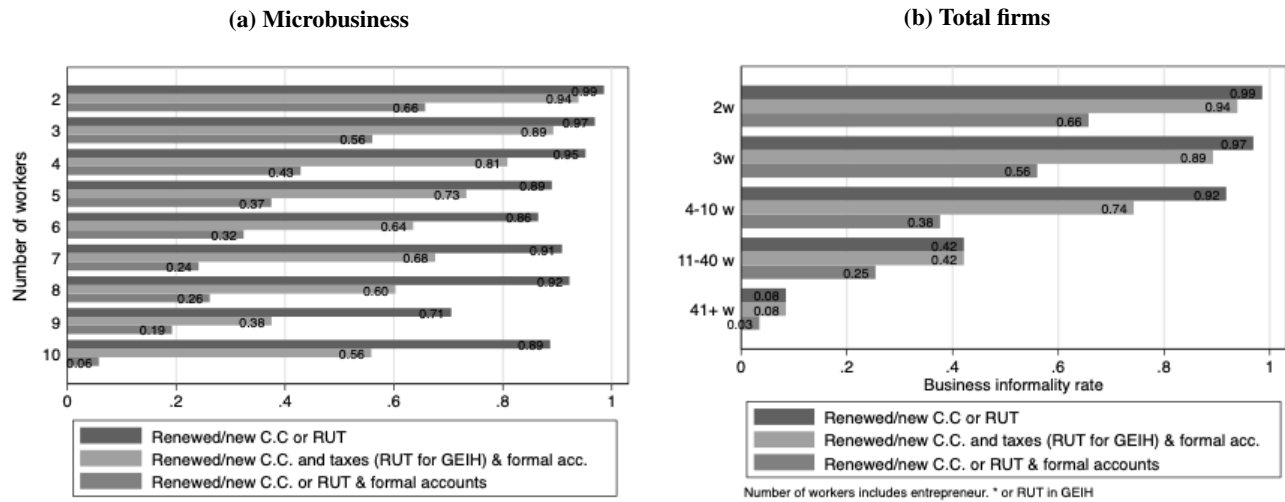
---

<sup>5</sup>The question in the GEIH also includes RUT registration (income tax registration).

<sup>6</sup>General Balance or Profit and Losses Statement or with a daily register book

<sup>7</sup>In the case of micro-business, it also includes as formal firms that not supposed to pay taxes. Firms with more than 10 workers (GEIH) are assumed to pay taxes.

Figure 2. Extensive margin of informality according to 3 definitions of informality

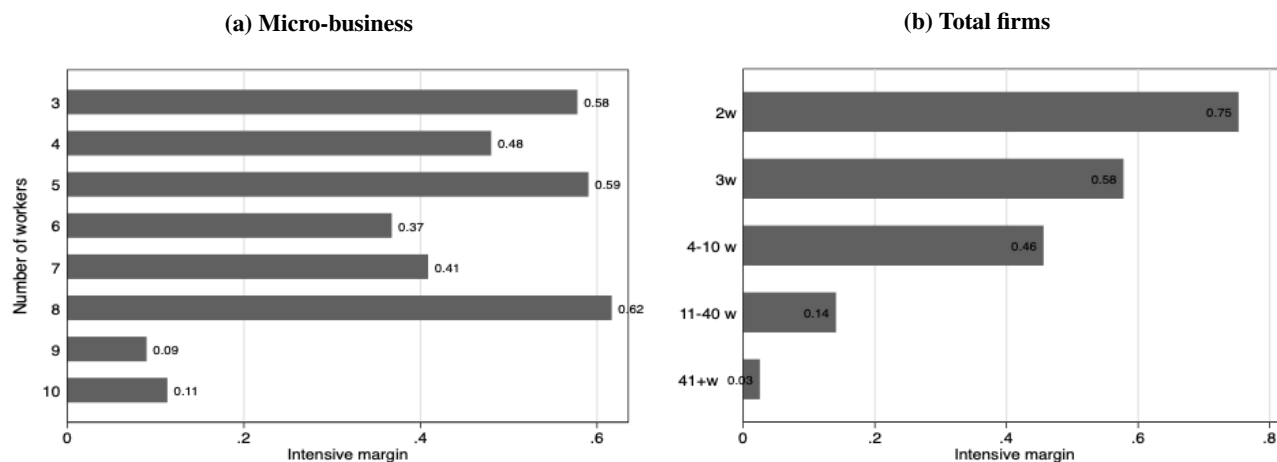


Source: EMICRON (2019) and Fernández (2021).

**2. Intensive margin of informality** The intensive margin of informality is defined as the percentage of workers who do not contribute to health or pensions and are hired by formal companies with at least 2 dependents (3 workers). Figure 3 shows this statistic for microbusiness (EMICRON, 2019) and for all firms (EEG, 2019)<sup>8</sup>. The intensive margin for micro-business does not show a smoothly decreasing behavior, probably due to the number of observations available, but arranging the information by ranges generates a decreasing intensive margin. However, it is important that the intensive margin remains significant for relatively large firms (14% among formal firms with 10 to 40 workers). Therefore, we do not assume a null intensive margin for firms with more than 10 workers, as literature often do. The labor informality rate among EEG(2019) firms is 71%.

<sup>8</sup>The informality rate is observed for firms with less than 10 workers (EMICRON). Informality for larger firms (bigger than 10 workers) outside of these structural surveys is estimated through a random forest procedure feed with the same GEIH survey but regarding questions asked to workers and the GEIH (2021) that asks workers about the registration status of their place of work, as shown in Fernández (2023d). Non-remunerated workers, partners, and entrepreneurs are also classified according to their informality status and therefore are included in the denominator.

**Figure 3. Intensive margin of informality**



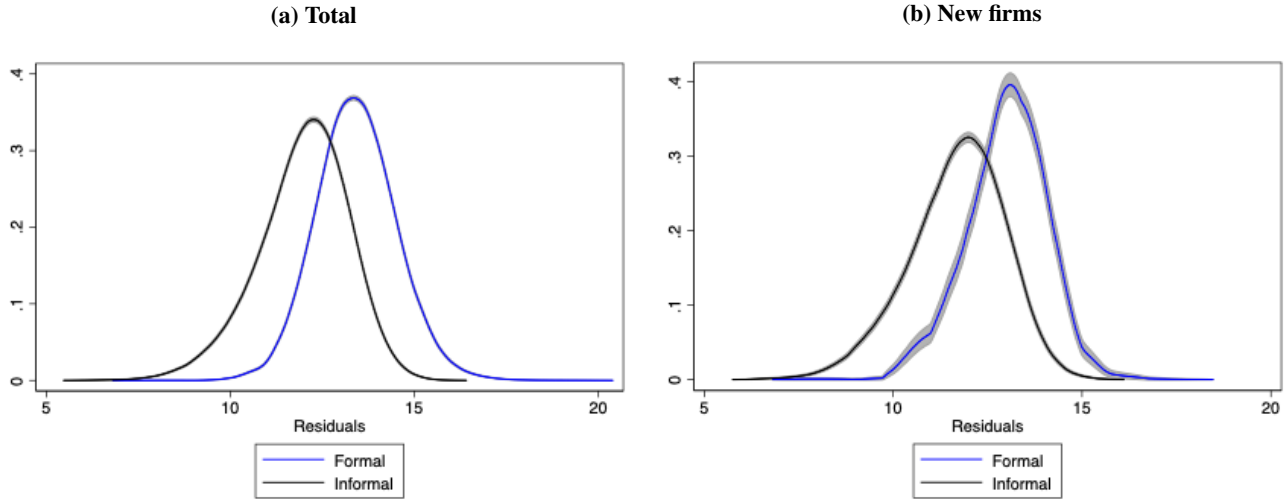
Source: Fernández (2023d). The intensive margin for firms with more than 10 workers is estimated.

**3. Value-added distributions.** EEG (2019) also allows for the comparison of per-worker value-added distributions between formal and informal firms<sup>9</sup>. According to Ulyssea (2018), the overlap of the formal and informal per-worker value added distributions, with formal firms having a higher mean, is consistent with heterogeneous firms that take the formality/informality decision with limited information. As shown in Figure 4, the Colombian case shows a high overlap between formal and informal distributions, suggesting a high uncertainty. The perfect conditions foreseen in his model would predict totally independent distributions.

According to this panorama, Ulyssea (2018) is a good model to understand informality in Colombia for four reasons: includes micro-business and larger companies simultaneously in the analysis, contemplates the two margins of informality, considers some firms and worker heterogeneity, and allows modifications to obtain a detailed institutional framework, as shown in the next section.

<sup>9</sup>The EMICRON and the structural surveys provide this information for micro business and firms in the services, trade, and manufacturing sector. The value added of other firms is estimated using factor remunerations. The complete methodology can be found in Fernández (2023d)

Figure 4. Intensive margin of informality



Source: Fernández (2023d).

## D A model to understand informality with institutional frictions

To show the importance of institutions in shaping informality, this article estimates [Ulyssea \(2018\)](#) for the case of Colombia. The idea of this model is that, based on an economic assessment, firms decide to comply or not with labor, tax regulations subject to waivers, and tax-free brackets.

This model assumes the same production function and prices (normalized to 1) for formal and informal firms, one factor of production (labor), and a single good. The concave production function for each sector is  $y_s(\theta) = \theta l_s^\alpha$  with  $\alpha < 1$  and  $s = I, F$ , where the subscript  $S$  refers to whether the workers are hired by a formal or informal firm. The model considers two levels of qualification (skilled (1) and unskilled (2)) aggregated through a CES function with shares of skilled workers  $\eta_i$  and  $\eta_f$  and an elasticity of substitution  $\frac{1}{1-\rho}$ . Therefore,  $l_F = (\eta_i l_{F1}^\rho + (1 - \eta_f) l_{F2}^\rho)^{\frac{1}{\rho}}$ .

In addition to regulation, informal firms face fixed and variable production costs, and an enforcement cost applied to the extensive margin that increases in size, and of the following form  $T(l_I) = (1 + \frac{l_I}{b_I})$ , where  $b_I$  is an exogenous parameter. Since informal firms are assumed to only hire informal workers, the earnings of informal firms (I) can be expressed as follows:

$$\begin{aligned} \max_{l_{i1}, l_{i2}} \pi &= \theta l_I^\alpha - \left(1 + \frac{l_{I1}}{b_I}\right) w_1 l_{I1} - FC_I \\ s.t. \quad l_I &= (\eta_I l_{I1}^\rho + (1 - \eta_I) l_{I2}^\rho)^{\frac{1}{\rho}}, \quad l_I > 0, \quad \alpha < 1, \quad w_1 > w_2 \end{aligned}$$

On the other hand, formal firms  $F$  must pay income taxes ( $\tau_y$ ) charged over the value added net of formal costs and vat taxes charged on revenues ( $\tau_{vat}$ ).  $\tau_y$  is equal to zero if income is lower than the income tax threshold ( $Y_y$ ) and  $\tau_{vat}$  is equal to zero if income is lower than the VAT threshold ( $Y_{vat}$ ). Formal firms can hire formal workers, paying social security contributions ( $\tau_w$ ) and a minimum wage, if binding, on unskilled wages<sup>10</sup>; and / or informal workers, who are affected by enforcement costs, like those faced by informal firms  $F(l_{Fs}) = (1 + \frac{l_{Fis}}{b_s})$ ,  $s = 1, 2$ .

Workers are assumed to perform the same activities within the formal firm, subject to their level of qualification. They are hired according to their relative costs, but there is a critical threshold  $\tilde{l}$  above which only informal workers are hired.  $\tilde{l}$  can be seen as an endowment of informal workers for all firms, even large formal ones<sup>11</sup>. Skilled workers are assumed to face a higher enforcement on skilled workers; therefore, the threshold for this type of labor is lower. This feature generates three types of formal firms: i) Firms that hire only informal workers ( $l_{F1} < \tilde{l}_1$  &  $l_{F2} < \tilde{l}_2$ ), ii) Firms that hire some formal skilled workers ( $l_{F1} > \tilde{l}_1$  &  $l_{F2} < \tilde{l}_2$ ) and iii) Firms that hire some formal skilled and unskilled workers ( $l_{F1} > \tilde{l}_1$  &  $l_{F2} > \tilde{l}_2$ ). The profits of formal firms ( $F$ ) can be expressed as follows:

<sup>10</sup>According to Loayza, Ulyseas, and Utsumi (2018). Whether the minimum wage is binding or not can be written as  $\max(wage_{unskilled}, w_{min})$  then  $t_{wmin} = \frac{\max(wage_{unskilled}, w_{min})}{w_{min}}$ .

<sup>11</sup>A way to understand this, is the existence of some activities within the firm subject to be performed by independent workers, that pay social security on their own.

$$\begin{aligned}
\pi_f &= \max_l \left\{ (1 - \tau_y - \tau_{vat}) \theta l_F^\alpha - C(l) \right\} \\
C(l) &= \left(1 + \frac{l_{Fi1}}{b_s}\right) l_{Fi1} w_1 + \left(1 + \frac{l_{Fi2}}{b_s}\right) l_{Fi2} w_2 + \lambda_1 (1 + \tau_w) (1 - \tau_y) l_{Ff1} \\
&\quad + \lambda_2 (1 + \tau_w) (1 - \tau_y) (1 + \tau_{wmin}) l_{Ff2} \\
s.t. \quad l_F &= (\eta_i l_{F1}^\rho + (1 - \eta_f) l_{F2}^\rho)^{\frac{1}{\rho}}, \quad l_F > 0, \quad \alpha < 1, \quad w_1 > w_2 \ \& \ b_2 > b_1 > 0, \\
l_{Ffs} &= l_{Fs} - \tilde{l}_s \quad s = 1, 2, \quad \lambda_s = 1 \quad \text{if } l_{Ffs} > \tilde{l}_s \quad \text{and } 0 \text{ otherwise,} \\
\tau_y &= 0 \quad \text{if } \theta l_F^\alpha < Y_y, \quad \tau_{vat} = 0 \quad \text{if } \theta l_F^\alpha < Y_{vat}, \quad \tau_{wmin} = \frac{\max(w_2, w_{min})}{w_{min}}.
\end{aligned}$$

More clearly, the maximization problem according to size of the firm is the following:

1. Firms that hire only informal ( $l_{F1} < \tilde{l}_1$  &  $l_{F2} < \tilde{l}_2$ )

$$\begin{aligned}
\max_{l_{i1}, l_{i2}} \pi &= (1 - \tau_y - \tau_{vat}) \theta l_F^\alpha - \left(1 + \frac{l_{Ii1}}{b_1}\right) w_1 l_{Ii1} + \left(1 + \frac{l_{Ii2}}{b_2}\right) w_2 l_{Ii2} - FC_I \\
s.t. \quad l_I &= (\eta_i l_{FI}^\rho + (1 - \eta_f) l_{I2}^\rho)^{\frac{1}{\rho}}, \quad l_I > 0, \quad \alpha < 1, \quad w_1 > w_2 \ \& \ b_2 > b_1 > 0
\end{aligned}$$

2. Firms that hire some formal skilled workers ( $l_{F1} > \tilde{l}_1$  &  $l_{F2} < \tilde{l}_2$ )

$$\begin{aligned}
\max_{l_{i1}, l_{i2}} \pi &= (1 - \tau_y - \tau_{vat}) \theta l_F^\alpha - \left(1 + \frac{\tilde{l}_1}{b_1}\right) w_1 \tilde{l}_1 + \left(1 + \frac{l_{Fi2}}{b_2}\right) w_2 l_{Fi2} - FC_F \\
&\quad - (1 + \tau_w) (1 + \tau_y) w_1 l_{Ff1} \\
s.t. \quad l_F &= (\eta_i l_{F1}^\rho + (1 - \eta_f) l_{F2}^\rho)^{\frac{1}{\rho}}, \quad l_F > 0, \quad \alpha < 1, \quad w_1 > w_2 \ \& \ b_2 > b_1 > 0, \\
l_{Ffs} &= l_{Fs} - \tilde{l}_s, \quad s = 1, 2, \quad \tau_{wmin} = 0 \quad \text{if } \theta l_F^\alpha < Y_y, \quad \tau_{vat} = 0 \quad \text{if } \theta l_F^\alpha < Y_{vat}, \\
\tau_{wmin} &= \frac{\max(w_2, w_{min})}{w_{min}}.
\end{aligned}$$

3. Firms that hire some formal skilled and unskilled workers ( $l_{F1} > \tilde{l}_1$  &  $l_{F2} > \tilde{l}_2$ )

$$\begin{aligned}
\max_{l_{i1}, l_{i2}} \pi &= (1 - \tau_y - \tau_{vat})\theta l_F^\alpha - \left(1 + \frac{\tilde{l}_1}{b_1}\right)w_1\tilde{l}_1 + \left(1 + \frac{\tilde{l}_2}{b_2}\right)w_2\tilde{l}_2 - FC_F \\
&\quad - (1 + \tau_w)(1 + \tau_y)(w_1 l_{Ff1} + \max(w_2, w_{min})l_{Ff2}) \\
s.t. \quad l_F &= (\eta_i l_{F1}^\rho + (1 - \eta_f)l_{F2}^\rho)^{\frac{1}{\rho}}, \quad l_F > 0, \quad \alpha < 1, \quad w_1 > w_2 \ \& \ b_2 > b_1 > 0, \\
l_{Ffs} &= l_{Fs} - \tilde{l}_s, \quad s = 1, 2, \quad \tau_{vat} = 0 \text{ if } \theta l_F^\alpha < Y_y, \quad \tau_{vat} = 0 \text{ if } \theta l_F^\alpha < Y_{vat}, \\
\tau_{wmin} &= \frac{\max(w_2, w_{min})}{w_{min}}.
\end{aligned}$$

The informal worker thresholds ( $\tilde{l}_s$ ,  $s = 1, 2$ ) can be derived by equalizing the marginal cost of hiring informally ( $w_2 + \frac{2w_2 l_{F2}}{b_2}$ ) and formally  $\{(1 + \tau_w)(1 + \tau_y) \max(w_2, w_{min})\}$ .

The threshold for unskilled workers can be derived as follows:  $\tilde{l}_2 = \frac{(1 + \tau_w)(1 + \tau_y)b_2 w_2^*}{2}$ ,  $w_2^* = \max(w_2, w_{min})$ . Similarly, the threshold for skilled workers that do not face the minimum wage restriction can be written as:  $\tilde{l}_1 = \frac{(1 + \tau_w)(1 - \tau_y)b_1}{2}$ .

Each period, a potential number of  $M$  firms enters the market. These firms only observe a vague production function  $v$ , which is a parameter that distorts true productivity.  $v$  is identically distributed among all participants. This uncertainty is key, since it allows us to have formal and informal firms at the same level of productivity. To enter the formal sector, firms must pay a fixed cost, denominated in units of production,  $E_f > E_i$ . Once companies enter the market, they obtain their true productivity from the function  $F(\theta/v)$ , which remains constant once companies enter the market. There is a positive relationship between  $v$  and  $\theta$ . This ex ante uncertainty mechanism explains why formal and informal firms coexist in the same productivity ranges. If firms face unexpectedly low productivity, they exit the market before they start producing. Firms face the possibility of exiting the market  $\kappa_i$  or  $\kappa_f$ , depending on whether they are formal or informal. Since prices and productivity remain constant, the firm value function can be expressed as  $V_s(\theta, w) = \max\left\{0, \frac{\pi_s(\theta, w)}{\kappa_s}\right\}$ ,  $s = I, F$  and the respective expected value as:  $V_s^e(\theta, w) = \int V_s(\theta, w) dF\left(\frac{\theta}{v}\right)$ ,  $s = I, F$ .

## E Estimation of the model

To observe the impact of tax, labor, and social security institutions on the behavior of informality, the model detailed above was estimated for the case of Colombia with and without institutional constraints. Following [Ulyssea \(2018\)](#), the model was estimated using a two-stage minimum distance (SMD) estimator. This estimator uses the value of some structural parameters and the guess values of estimated parameters, wages for skilled and unskilled workers, productivity shocks, and the distribution functions of some key variables of the economy. Some moments of these distributions are compared with the same moments obtained from the distributions of the real data, and the estimated parameters are adjusted until the model converges<sup>12</sup>. Therefore, the initial and final values for the estimated parameters might diverge.

The estimation of the model involves gathering information of the structural parameters in the economy, as well as giving an accurate guess for the initial values of the estimated parameters, which is important given that the SMD is a procedure to find local rather than global minimums. The sources of the structural parameters and the initial values of the estimated parameters for the unconstrained and institutional-constrained models are detailed in Annex 1.

**Table 4. Structural parameters used in the estimation**

		Brazil	Mexico	Colombia unconstrained	Colombia constrained
$w_1$	Equilibrium wage of skilled informal workers	844	2.35	985	973
$w_2$	Equilibrium wage of unskilled informal workers	591		688	
$w_2$	Equilibrium wage of unskilled workers			820	
$\tau_w$	Payroll/Payroll Contributions	0.37	0.35	0.34	0.47
$\tau_y$	Income taxes	0.29	0.00	0.14	0.29
$V_0$	Pareto parameter, ex ante productivity	7.00	7.08	7.30	7.30
$\kappa_f$	Exit probability of formal firms	0.13	0.08	0.16	0.16
$\gamma_f$	Cost of operating in the formal sector / unskilled wage	0.45	0.45		
$\gamma_i$	Cost of operating in the informal sector / unskilled wage			0.05	0.05
$\tau_{vat}$	VAT tax (net of income tax deduction)				0.145
$y_{ty}$	Monthly threshold for paying income tax (40% deductions)				5597
$y_{tvat}$	Monthly threshold for paying VAT				9995

Sources: Ulyssea (2018), Alvarez and Ruane (2019) and those referred above in the text.

Table 4 summarizes the parameters used by [Ulyssea \(2018\)](#) and [Alvarez and Ruane \(2019\)](#) to analyze the cases of Brazil and Mexico; and the parameters resulting from

<sup>12</sup>To give robustness to the results, again following [Ulyssea \(2018\)](#) I simulated 300,000 observations and 20 databases

the calibration in the case of Colombia. Table 5 shows the initial and final values of the estimated parameters in the Colombian case, and the final parameters of the Brazilian and Mexican parameters. There are five important differences in the final version of the Colombia case with respect to the values registered in other countries. First, the cost of enforcement is lower in the case of Colombia. As explained in the introduction, the weakness of Colombia’s enforcement authorities is consistent with this results.

Another important difference from the case of Brazil is the higher ex post variance in productivity, which is consistent with the productivity of formal and informal firms having a greater overlap section in the case of Colombia (Fernández, 2023d); and the lower entry cost to the formal sector is probably related to Colombia which has periodic rather than upfront registration costs.

**Table 5. Estimated parameters**

		Brazil	Mexico	Colombia		
				Initial	Without constraints	With institutional constraints
$b_f$	Intensive margin cost, skilled workers	2.61	2.35	2.61	5.71	24.2
$b_f$	Intensive margin cost, unskilled workers	4.94		4.94	4.72	41.3
$b_i$	Cost of extensive margin	5.01	4.58	5.01	29.1	8.92
$\delta_i$	Exit probability of informal firms	0.38	0.27	0.24	0.25	0.27
$\gamma_i$	Cost of operating in the informal sector / w2	0.25	0.19			
$\gamma_f$	Cost of operating in the formal sector / w2			0.33	0.36	0.36
$x_i$	Pareto shape parameter	3.08	1.57	2.99	3.01	2.44
$E_f$	Cost of entering the formal sector	4282	93193	3214	3077	2946
$E_i$	Cost of entering the informal sector	2023	8	457	799	765
$\alpha$	Cobb Douglas function parameter	0.6	0.32	0.50	0.65	0.57
$\sigma$	Productivity variance ex post	0.24	0.27	0.24	0.33	0.54
$\rho$	CES elasticity parameter	0.29		0.29	0.37	0.30
$\eta_I$	Skilled worker CES share. Informal firms	0.48		0.59	0.58	0.63
$\eta_F$	Skilled worker CES share. Formal firms	0.59		0.88	0.77	0.78

Sources: Ulyssea (2018), Alvarez and Ruane (2019) and those referred above in the text.

## F Fitness of the model

Table 6 shows the moments of the distributions of the main variables estimated with the model and with the data, which serves as support for the calibration process. These moments were specifically selected for the Colombian case, and for this reason they are not compared with the Brazilian ones. As can be seen in the results, it replicates

remarkably well all parameters, except for the case of business informality among large firms.

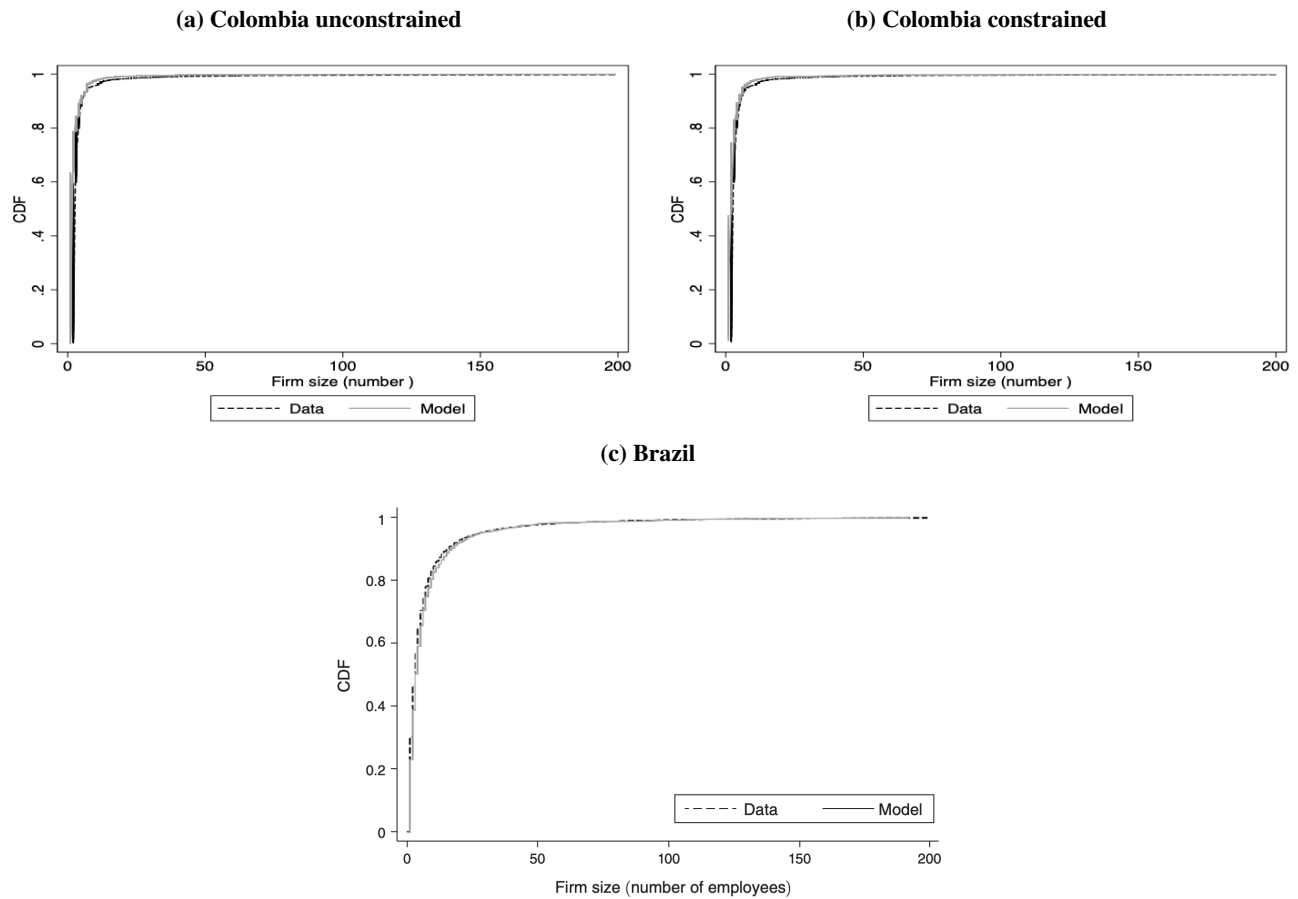
In addition to estimating the similarity of the models in terms of means, the distributions by size of some of the key variables were compared. Figure 5 shows the formal distribution of companies for Brazil and Colombia (constrained and unconstrained).

**Table 6. Estimated moments (observed & simulated data)**

	<b>EEG and GEIH</b>	<b>Model with institutional constrains</b>
Moment 1. Formal workers among dependents (GEIH, workers)	37%	37%
Moment 2. Informal workers among the unskilled (GEIH, workers)	61%	61%
Moment 3. Informal workers among the qualified (GEIH, workers)	28%	28%
Moment 4. Business informality rate	93%	96%
Moment 5. Business informality rate in firms with 1/3 dependents (2/4 workers)	97%	96%
Moment 6. Business informality rate in firms with 4/9 dependents (5/10 workers)	83%	86%
Moment 7. Business informality rate in firms with 10+ dependents (11+ workers)	22%	24%
Moment 8: Intensive margin in firms with 2/4 dependents (3/5 workers)	55%	55%
Moment 9: Intensive margin in firms with 5/19 dependents (6/20 workers)	24%	21%
Moment 10: Intensive margin in firms with 20+ dependents (21+ workers)	3%	4%
Moment 11: % of informal firms with 1/4 dependents (2/5 workers) in total informal firms	95%	95%
Moment 12: % of informal firms with 1/9 dependents (1/10 workers) in total informal firms	99%	99%
Moment 13: % of formal firms with 1/9 dependents (2/10 workers) in total formal firms	54%	51%
Moment 14: % of formal firms with 10/19 dependents (11/20 workers) in total formal firms	18%	18%
Moment 15: % of formal firms with 20/39 dependents (21/40 workers) in total formal firms	13%	14%
Moment 16: % of formal firms with 40+ dependents (41+ workers) in total formal firms	15%	18%

Sources: EEG and model estimates under the strict informality scenario.

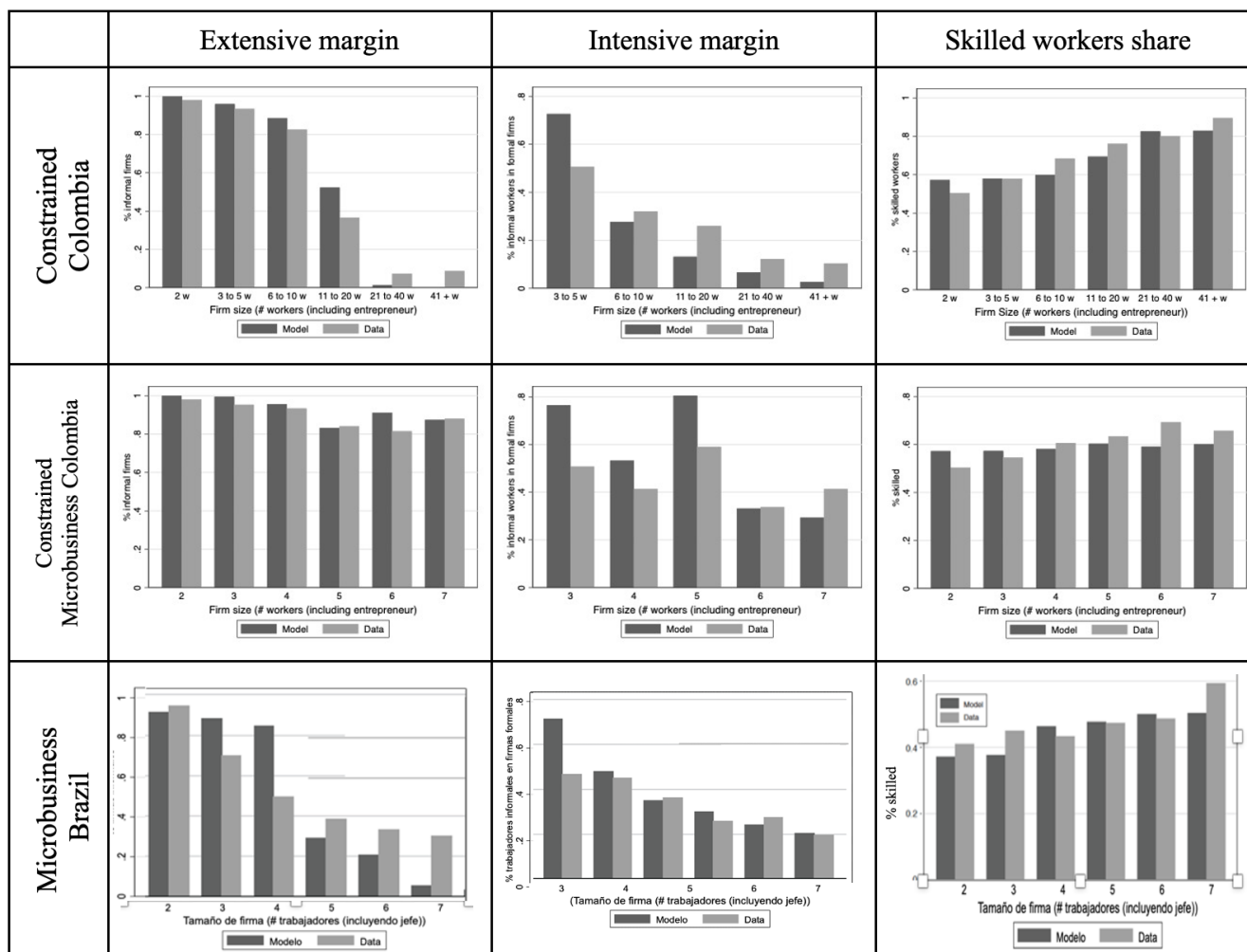
Figure 5. Firm size distribution (Formal firms)



Source: Author's calculations EEG (2019), GEIH (2019) and Ulyssea (2018).

Figure 6 shows the extensive and intensive margin of informality and the share of skilled workers in formal and informal firms, for unconstrained Colombia, for constrained Colombia, constrained microbusiness (which is more comparable to Ulyssea (2018)) and Brazil. As shown in the figure, the prediction of the model fits the data very well, considering the erratic behavior of the indicator for firms between 5 and 10 workers.

**Figure 6. Firm size distribution of extensive margin, intensive margin and share of skilled workers**

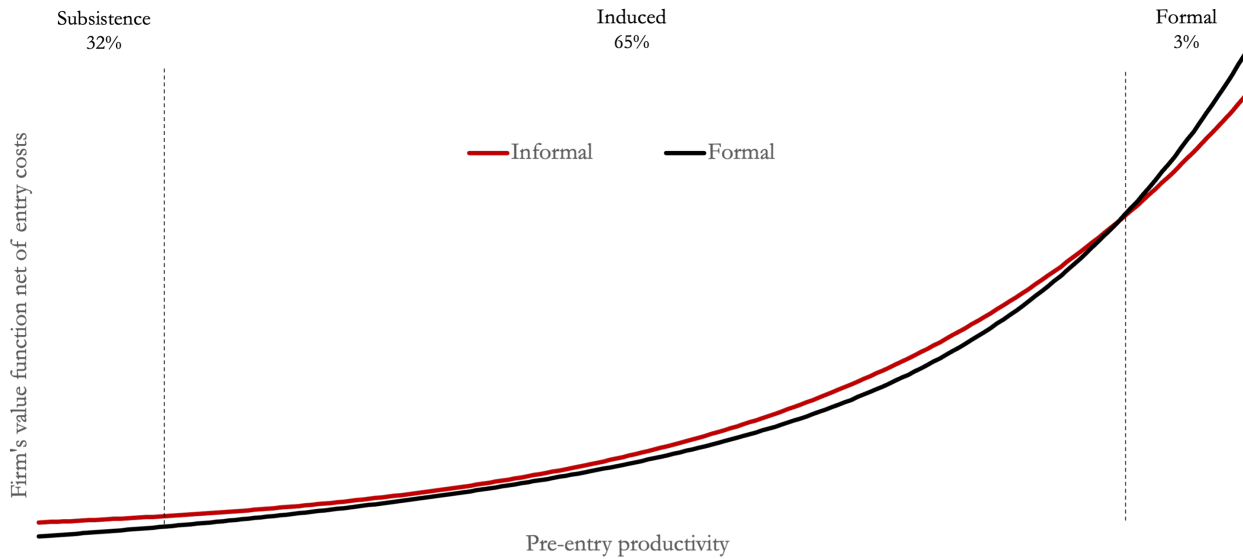


Source: Author's calculations EEG (2019), GEIH (2019) and Ulyseia (2018).

One of the most important features of this model is that it allows one to identify three types of firms: the Subsistence firms that are the ones that cannot operate formally; the Induced, which are those that could operate as formal but prefer to be informal because it is more profitable, and the Formal Firms. According to Figure 7, the Subsistence Informality is relatively small in Colombia, which makes sense since the fixed entrance cost of the formality is relatively small. Also, it is important to consider that unipersonal companies are excluded from this exercise. On the other hand, Induced firms represent quite a large portion of informality, and the informal and formal firm values are similar,

implying that policies that change the relative variable cost of operating formally might be effective<sup>13</sup>.

**Figure 7. Taxonomy of informality**

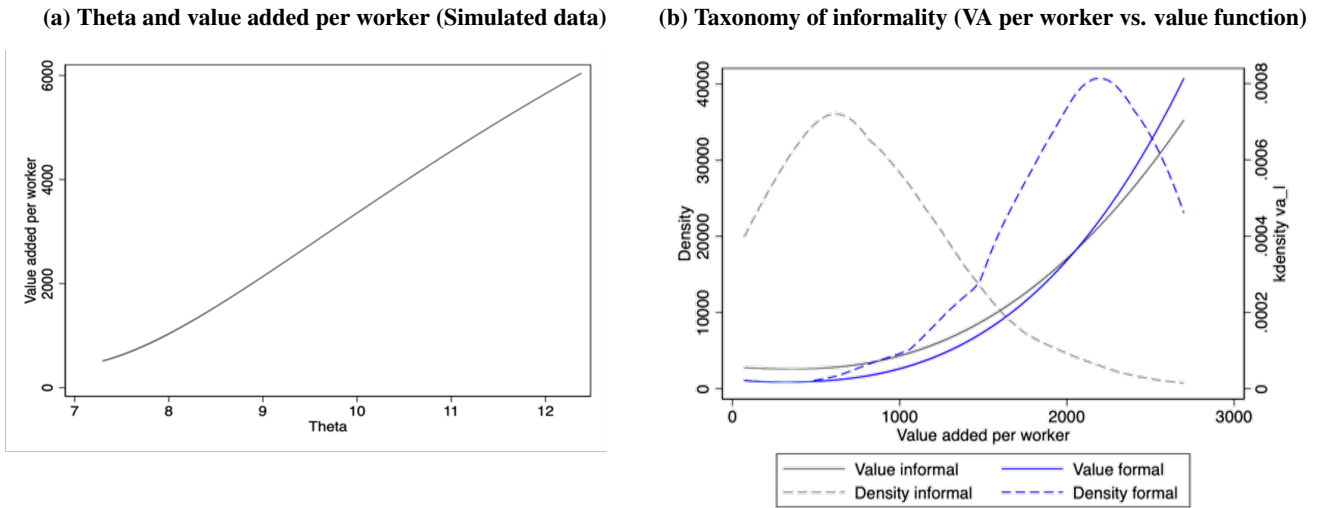


Source: Model simulations.

This taxonomy is interesting because it considers the heterogeneity of informality in Latin America and is very useful to understand the impact of different policies. However, one inconvenience is that  $\theta$  is the parameter that allows one to identify the different types of informality. An alternative is to build a taxonomy based on the value added per worker. As shown in Figure 8a, there is a lineal relationship between this variable and  $\theta$ , and it can be corroborated in Figure 8b that the taxonomy is similar to that based on  $\theta$ . Figure 8b also shows that the density of formal and informal firms, along their size, responds to incentives to be formal and informal. This finding allows policy targeting to be more efficient. However, it is important to note that the intersection between the formal- and informal-sector curves occurs later in the value added than in the pre-productivity taxonomy.

<sup>13</sup>Ulyssea (2018) proposed a much complex taxonomy which depends on a scenario in which entry barriers were eliminated. I prefer the one used here, not only because it is easier to implement but also because in the Colombian case the scenario without entry cost is affected by non-linearities and becomes difficult to use to formulate a taxonomy.

**Figure 8.**  $\theta$ , value added per worker and a taxonomy based on the latter



Source: Simulated data.

## G Comparative statics

To analyze the impact of policies on a set-up with a complex institutional arrangement as in the Colombian case, on top of the traditional policies to reduce informality, some tax scenarios were added. The objective of introducing these scenarios is not to suggest that tax policies should be implemented to reduce informality, but rather as informality being a side effect of these policies.

The exercises were divided into two groups: policies to reduce business informality and policies to reduce labor informality in formal firms, understanding that these two objectives can be strongly related. To observe changes clearly, the level of changes in parameters tried to resemble extreme cases, but in some exercises, this was not possible without generating a corner solution, and softer versions of the change were analyzed instead. The best example of this limitation is the reduction in payroll taxes that, as [F. Jaramillo \(2020\)](#) pointed out, should follow the following rule  $\tau_w > \tau_y + \tau_w * \tau_y$  to be able to find an internal solution of the model.

As in [Ulyssea \(2018\)](#), the impact of these policies was analyzed on the lifetime value of the firm net of entry costs, for each percentile of the productivity distributions (which is positively related to size) and for each of the following groups: i) “always formal”: firms that are formal in the base case and in the counterfactual; ii) “always informal”:

firms that are informal in the base case and in the counterfactual; iii) “switchers”: informal firms in the base case becoming formal in the counterfactual. The macroeconomic implications of the policies were also included in the analysis.

## G.1 Policies on the extensive margin or business informality

Four policies were considered to reduce business informality: a 50% reduction in formal entry costs to a level that is almost equal to informal entry costs, 45% more enforcement of informality ( $b_i = 1$ ) and the elimination of income and vat taxes ( $\tau_y = 0$  and  $\tau_{vat} = 0$ ). As shown in Figure 9, in the enforcement scenario, the value of “always informal” firms decreases because they reduce their size inefficiently to avoid being caught; and because it becomes more costly to operate informally, at the point that some firms decide to formalize (“switchers”). Although the most affected informal firms are the large ones, considering enforcement is increasing in size, most of the firms that decide to formalize are the smaller ones, because they are exempted from income tax, so it is relatively less expensive to become formal. Similarly, in the scenario of a reduction in formal entry costs, the ‘switchers’ tend to be small firms since their fixed costs are a larger share of total costs. The increase in the number of formal firms in both scenarios increases the skilled wage hurting “always formal” firms.

Regarding tax policies, the decrease in VAT and income tax is positive for formal firms and negative for informal firms because of competition in the labor market. It also led firms to formalize, particularly the larger ones, which are subject to pay taxes. The set of “switchers” that opt to formalize is more restricted in the VAT scenario, since the VAT threshold is higher.

This is consistent with the macro impacts of the policies illustrated in Table 7. In fact, these policies induce the entry of new companies, the formalization of others, and a consequent reduction in the informality of the business. As formal firms hire a higher proportion of skilled workers, the skilled wage increases. Meanwhile, the unskilled wage substantially decreases in the enforcement scenario, where informal firms get severely hurt; and marginally decreases in the entry cost scenario, because of general equilibrium means.

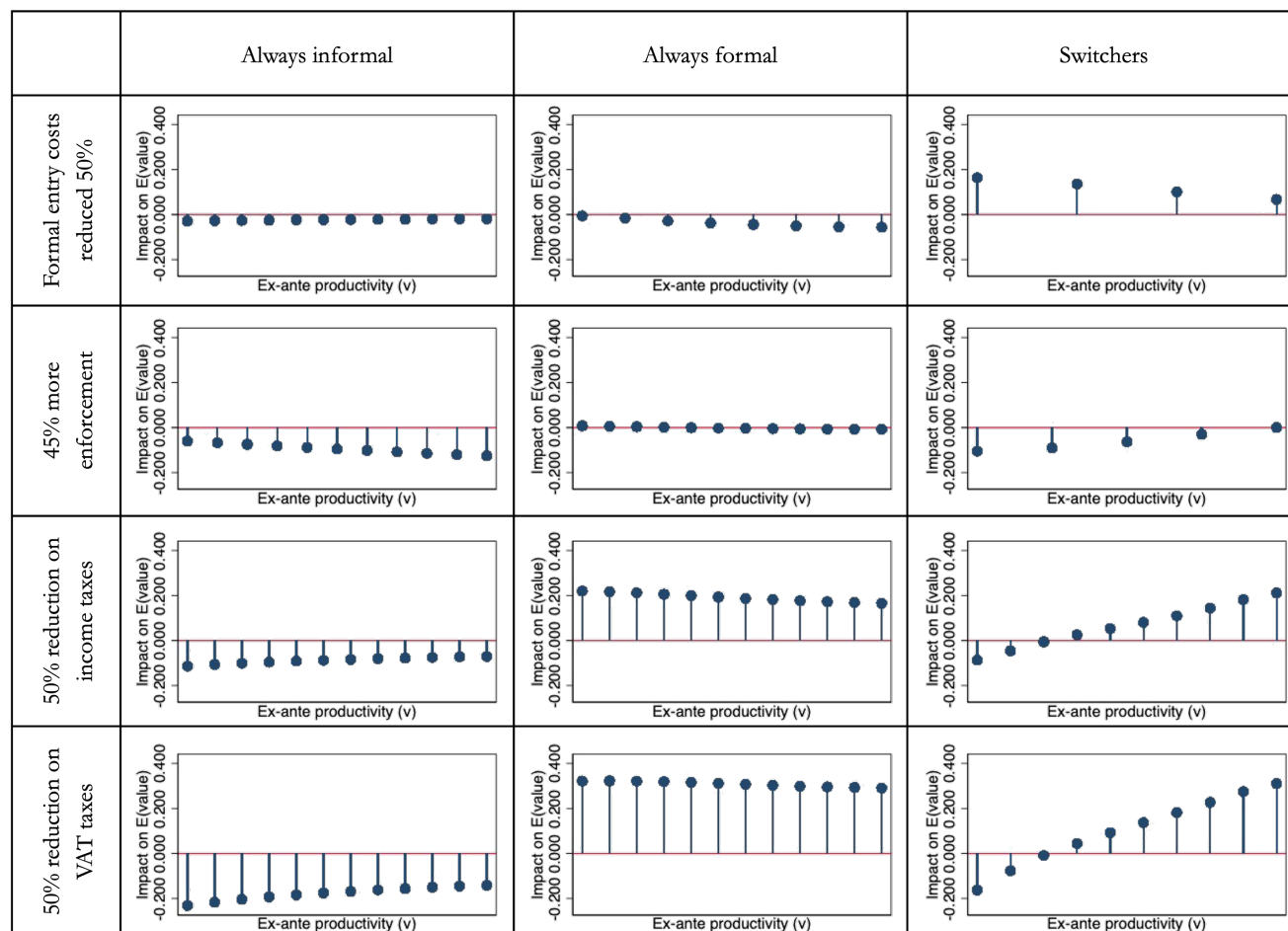
The reduction in entry costs scenario, where the firms that formalize are mostly small, causes the percentage of firms that hire informally to increase, augmenting the informal-

ity rate of skill workers. Meanwhile, in the enforcement and VAT scenario, a reduction on the informality rate of skilled workers is observed, because larger firms are the ones that hire formally. Interestingly, labor informality increases when the income tax is eliminated because the labor cost deduction disappears.

Taxes revenues are mostly driven by social security contributions, and consequently decrease in the scenarios with larger labor informality. Meanwhile, as the reduction of entrance costs is biased towards small firms, this scenario is characterized by a smaller reduction of the informal GDP and a lower output. Finally, whereas in the entry cost and enforcement scenarios welfare decreases along with the unskilled wage; the increase in welfare is impressive in the tax scenarios, probably due to an unrealistic assumption of a perfectly efficient zero-cost government.

In sum, extensive margin policies are successful in reducing business informality. This would naturally reduce the informality of labor, as in the VAT scenario. However, if the firms that formalize are the smaller ones, as in the enforcement or the entry-cost scenario, labor informality remains unchanged and can even increase due to firms entering the market. In the income tax scenario, the informality of labor increases because the tax deduction disappears. One key question is the importance of formalizing firms if they do not pay taxes and do not hire workers formally.

**Figure 9. Microeconomic impact of policies oriented to the extensive margin**



Source: Simulated data.

**Table 7. Macroeconomic impact of policies oriented to the extensive margin (policy/baseline, constrained model)**

	Lower entry cost	Higher enforcement	Lower income tax	No VAT tax
Informal firms	0.62	0.94	0.76	0.74
Informal GDP	0.62	0.77	0.55	0.48
Informal workers	1.08	0.87	1.17	1.07
Skilled informal workers	1.12	0.79	1.24	1.14
Unskilled informal workers	1.03	0.98	1.08	0.97
Skilled premium	1.11	1.15	1.08	1.07
Skilled wage	1.06	1.01	1.12	1.23
Unskilled wage	0.96	0.88	1.04	1.14
Mass of firms	1.13	1.03	1.09	1.08
Output	1.01	1.06	1.03	1.04
Tax revenues	0.96	1.08	0.90	0.96

Sources: Simulated data.

## G.2 Policies on the intensive margin or labor informality in formal firms

Figure 10 shows the impact of decreasing payroll taxes in 4.5 pp, double the enforcement on informal hiring by formal firms ( $b_1 = 2$  and  $b_2 = 0.5$ ), and an elimination of the income tax waiver on small firms (threshold for income taxes = 0), which, as shown before, generates a tax deduction on payroll taxes for small firms.

As shown in the Introduction, payroll costs are higher for small firms that do not take advantage of the tax deduction; but these firms do not hire many workers formally. Therefore, the reduction of payroll taxes is mostly beneficial to productive “always formal” firms, which are large formal workers employers, increasing labor demand, wages, and hurting informal firms, by general equilibrium means.

In contrast, the increase in enforcement hurts small “always formal” firms that hire more workers without a formal contract. This lowers unskilled wages and benefits large “always formal” and “always informal” firms. The impact of eliminating the tax waiver on small firms is very similar to the impact of reducing payroll taxes, but large “always formal” firms get relatively more benefits, and small “always formal” firms end up being hurt.

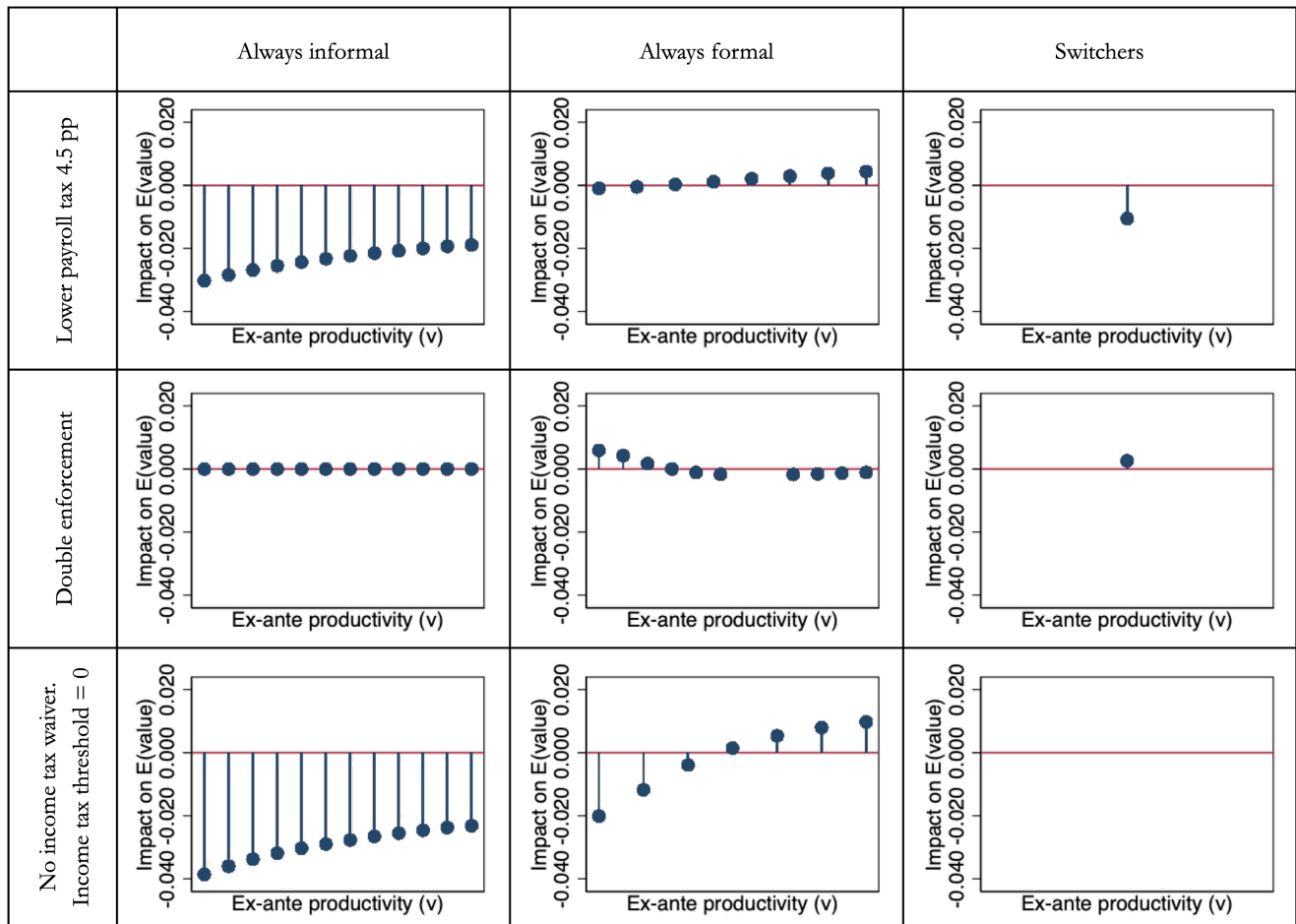
Consistently, Table 8 illustrates that intensive margin policies reduce labor informality to some extent. However, the impact varies according to the policies. In the reduction of payroll taxes, labor informality gets lower among skilled workers, since larger firms are the ones that get more benefits, whereas in the elimination of income tax deduction scenario, labor informality gets lower among unskilled workers. Consequently, skilled wages increase relatively more in the payroll scenario, whereas unskilled wages increase relatively more in the income tax scenario. Enforcement in the intensive margin scenario has a lower impact on labor informality (-1pp), since the formalization effect is controlled by the fact that it becomes more profitable for new firms to operate fully informal than to operate formally and face a high enforcement on the hiring process.

Similarly, as the payroll policy is more favorable for larger firms, production increases more than in other scenarios. From a fiscal point of view, as expected, the reduction on payroll taxes reduces tax revenues and the elimination of the tax waiver increases them. Welfare moves accordingly to changes in taxes.

Reducing the tax waiver in small firms and adopting a flat profit tax, or simply enforcing firms to register as a business, can be an effective way to reduce labor informality, because it solves the asymmetry of tax deductions. However this policy is difficult to implement because of social, political and enforcing reasons.

In sum, labor informality policies have a consistent but limited impact on labor informality and few impacts on business informality. However, it is important to consider that strong enforcement on the intensive margin can make it more profitable for small firms to become informal 100%. Another drawback of policies oriented to reduce labor informality is the fiscal cost. This result is in line with the evaluations of the 2012 reduction of payroll taxes in Colombia.

Figure 10. Microeconomic impact of policies oriented to the intensive margin



Source: Simulated data.

**Table 8. Macroeconomic impact of policies oriented to the intensive margin (policy/baseline, constrained model)**

	Lower payroll tax	Higher labor enforcement	No income tax waiver
Informal firms	1.00	1.00	1.00
Informal GDP	0.95	0.99	0.97
Informal workers	0.93	0.98	0.94
Skilled informal workers	0.93	0.98	0.99
Unskilled informal workers	0.93	0.97	0.88
Skilled premium	1.01	1.00	0.93
Skilled wage	1.03	1.00	1.00
Unskilled wage	1.02	1.00	1.08
Mass of firms	1.00	1.00	1.00
Output	1.02	1.01	1.02
Tax revenues	1.02	1.01	1.00

Sources: Simulated data.

## H Final remarks

The recent Colombian Employment Mission ([Levy & Maldonado, 2021](#)) recommended an integral reform of the regulatory system to reduce unemployment; and warned against partial equilibrium solutions to the employment problem, which can generate wrong incentives and unwanted results. This paper presents a clear example of this situation. The waiver of income tax for small firms is an adequate policy to reduce informality in small firms, and the formal cost deductions from the income tax is an effective, though barely analyzed policy, to reduce business informality. However, the combination of both policies with nonlinear parameters ends up generating a large share of small firms hiring workers informally, and therefore with low possibilities of hiring skilled workers, which is essential to enhance productivity.

A possible solution to this problem is to generate vouchers of social security payments, accountable for firms below the income tax threshold that hire workers through a formal contract, that can be used once the firm gets bigger. This recommendation, in addition to generating incentives to formalize, generates incentives to grow. Other policies to address this problem include a single-tax scheme that includes social security or at least one in which taxes are estimated after deducting formal labor costs. More aggressive policies as a new social security scheme or a flat income tax can also be analyzed.

This paper also makes evident that the reduction of labor informality does not have an easy way out, and this explains many decades of unsuccessful efforts to reduce the problem. Consistent with the evaluations of the 2012 reduction on payroll taxes in

Colombia, the estimated model shows that the reduction in payroll taxes, amidst a complicated regulatory environment, has a moderate impact on informality and ties the fiscal accounts. The model simulations also show that tighter enforcement on the intensive margin could push formal firms towards informality, where all workers are hired informally, but the enforcement tends to be softer. Policies on an extensive margin are more effective, but do not necessarily increase labor formality. In fact, if policies are oriented towards small firms, the informality of labor might remain unchanged because small formal firms hire as many informal workers as an informal firm. This leads to the question of the importance of formalizing small businesses if formal small firms do not pay taxes and do not hire with a formal contract.

In sum, there is not a single policy that can magically reduce informality by itself. A combination of different policies focused on different types of informality and trying to avoid non-linearities and having in mind general equilibrium effects might be a better approach to solve the problem in the future.

This article makes an important contribution to understanding the interaction between informality of business and informality of labor in Colombia. However, is important to complete the informality panorama in Colombia with the self-employment analysis, which is one of the highest in Latin America.

## References

- Acosta-Henao, M. (2020). *Law Enforcement and the Size of the Informal Sector*. Retrieved from <https://ssrn.com/abstract=3588010> (Draft Paper)
- Alaimo, V., Bosch, M., Kaplan, D. S., Pagés, C., & Ripani, L. (2015). *Empleos para Crecer*. Washington, D.C.: Inter-American Development Bank.
- Almeida, R., & Carneiro, P. (2012). Enforcement of Labor Regulation and Informality. *American Economic Journal: Applied Economics*, 4(3), 64–89.
- Alvarez, J., & Ruane, M. C. (2019). *Informality and Aggregate Productivity: The Case of Mexico*. International Monetary Fund.
- Amarante, V., & Perazzo, I. (2013). Trabajo por Cuenta Propia y Monotributo en el Uruguay. *Revista internacional del trabajo*, 132(3-4), 623–641.
- Angulo, R. (2023). *Índice Multidimensional de Robustez de Micronegocios (IMICRO). Diseño y Resultados para una Estrategia de Focalización de la Secretaría Distrital de Desarrollo Económico de Bogotá*. (Tech. Rep.). Bogotá, Colombia: Inclusión SAS.
- Antón, A., & Rasteletti, A. (2018). *Imposición al Trabajo en Contextos de Alta Informalidad Laboral: Un Marco Teórico para la Simulación de Reformas Tributarias y de Seguridad Social* (Documento para Discusión No. 646). Washington, D.C.: Inter-American Development Bank.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5–32.
- Chacaltana, J. (2016). Peru, 2002-2012: Growth, Structural Change and Formalization. *Revista CEPAL*.

- Chisari, O. O., Romero, C. A., & Martínez, J. P. V. (2016). Economy-Wide Effects of Improving Small Firms Access to Capital Markets: An Applied General Equilibrium Assessment. *Revista de Análisis Económico–Economic Analysis Review*, 31(1), 15–36.
- Chowdhury, M. J. A. (2009). *Microcredit, Micro-enterprises, and Self-Employment of Women: Experience from the Grameen Bank in Bangladesh* (Vol. 209; Working Paper). Dhaka, Bangladesh: Center for Microfinance and Development.
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157–175.
- DANE. (2013). *General Methodology: Large Integrated Household Survey – GEIH*.
- DANE. (2020a). *General Methodology: Annual Services Survey – EAS*.
- DANE. (2020b). *General Methodology: Annual Trade Survey – EAC*.
- DANE. (2020c). *General Methodology: Survey Micro-establishments – EMICRON*.
- DANE. (2021). *Economic Census of Colombia: Economic Census Methodological Document*.
- Departamento Nacional de Planeación. (2019). *CONPES 3956 - Formalización Empresarial*. Retrieved from <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3956.pdf>
- De Paula, A., & Scheinkman, J. A. (2011). The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil. *Review of Income and Wealth*, 57, S8–S26.
- D’Erasmus, P. N., & Boedo, H. J. M. (2012). Financial Structure, Informality and Development. *Journal of Monetary Economics*, 59(3), 286–302.
- De Soto, H. (1989). *The Other Path: The Economic Answer to Terrorism*. New York, NY: Harper & Row.
- De Soto, H. (2000). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Basic Books.
- Dix-Carneiro, R., Goldberg, P. K., Meghir, C., & Ulyssea, G. (2021). *Trade and Informality in the Presence of Labor Market Frictions and Regulations* (Working Paper No. 28391). Cambridge, MA: National Bureau of Economic Research.
- Donovan, K., Lu, W. J., & Schoellman, T. (2023). Labor Market Dynamics and Development. *The Quarterly Journal of Economics*.
- Erosa, A., Fuster, L., & Martinez, T. R. (2023). Public Financing with Financial Frictions and Underground Economy. *Journal of Monetary Economics*, 135, 20–36.
- Eslava, M., Haltiwanger, J. C., & Pinzón, A. (2019). *Job Creation in Colombia vs the U.S.: “Up or Out Dynamics” Meets “The Life Cycle of Plants”* (Working Paper No. 25550). Cambridge, MA: National Bureau of Economic Research.
- Eslava, M., Meléndez, M., & Urdaneta, N. (2021). *Market Concentration, Market Fragmentation, and Inequality in Latin America* (Working Paper No. 11). NY, NY: United Nations Development Programme (UNDP).
- Estrada, D. A., & Hernández Rubio, A. (2019). *Situación Actual e Impacto del Microcrédito en Colombia*. Banco de la República de Colombia. Retrieved from <https://repositorio.banrep.gov.co/handle/20.500.12134/9723>
- Fairris, D., & Jonasson, E. (2016, February). *Determinants of Changing Informal Employment in Brazil, 2000–2010* (MPRA Paper No. 71475). Munich, Germany: University Library of Munich, Germany.
- Fajnzylber, P., Maloney, W. F., & Montes-Rojas, G. V. (2011). Does formality improve micro-firm performance? evidence from the brazilian simples program. *Journal of Development Economics*, 94(2), 262–276.
- Fernández, C. (2022). *Firms, Informality & Institutions. The Case of Colombia* (Documentos de Trabajo No. 020598). Bogotá, Colombia: Universidad del Rosario.
- Fernández, C., Villar, L., & Gómez, N. (2017). Taxonomía de la Informalidad en América Latina. *Coyuntura Económica*, 47(1 y 2), 137-167.
- Fernández, C. (2020). Informalidad Empresarial en Colombia. *Coyuntura Económica*, 50(1-2), 133-168.
- Fernández, C. (2023a). The Impact of Tax Asymmetries on Labor and Business informality. *Forthcomming*.
- Fernández, C. (2023b). *Informality, Productivity & Inequality: An Analysis from the Point of View of Firms and Workers*. (PNUD)
- Fernández, C. (2023c). Microbusiness and the Cost of Using Capital. *Forthcomming*.
- Fernández, C. (2023d). MSMEs and informality: A new employer-employee database for Colombia. *Forthcomming*.
- Fernández, C., & Villar, L. (2016). *The Impact of Lowering the Payroll Tax on Informality in Colombia* (Working Paper No. 015212). Bogotá, Colombia: Fedesarrollo.
- Granda, C., & Hamann, F. (2015). *Informality, Saving and Wealth Inequality in Colombia* (Working Paper No. 575). Washington, D.C.: Inter-American Development Bank.

- 'Haanwinckel, D., & Soares, R. R. (2021). 'workforce composition, productivity, and labour regulations in a compensating differentials theory of informality'. *The Review of Economic Studies*, 88(6), 2970-3010.
- Hamann-Salcedo, F. A., & Mejía, L. F. (2013). *Formalizando la Informalidad Empresarial en Colombia* (Vol. Capítulo 10. Pág.:399-427). Banco de la República de Colombia.
- Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, 60(1), 126.
- Jaramillo, F. (2020). *Informalidad – Margen Intensivo. Mimeo, note=Universidad del Rosario.*
- Jaramillo, M. (2013). *Employment Growth and Segmentation in Peru, 2001-2011: Country Case Study on Labour Market Segmentation* (Employment Working Paper No. 151). Geneva, Switzerland: International Labour Organization.
- Jenkins, S. P., & Kerm, P. V. (2007). *PARETOFIT: Stata module to fit a Type 1 Pareto distribution*. Statistical Software Components, Boston College Department of Economics.
- Kugler, A. D., Kugler, M. D., & Herrera-Prada, L. O. (2017). Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia's Reform. *Economía*, 18(1), 3–40.
- Levy, S. (2018). *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. Inter-American Development Bank.
- Levy, S., & Maldonado, D. (2021). *Misión del Empleo. Reporte Resumen.*
- Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(2), 139-191.
- Loayza, N. V. (2016). Informality in the Process of Development and Growth. *The World Economy*, 39(12), 1856–1916.
- Loayza, N. V., Ulyssea, G., & Utsumi, T. (2018). Informality and the Labor Market Effects of Mass Migration: Theory and Evidence from Syrian Refugees in Turkey. In *7th turkey labor market network meeting* (Vol. 9, p. 2020).
- Lora, E., Mejía, L. F., Benítez, M., Delgado, M. E., & Gutiérrez, D. (2021, April). *Reformas para una Colombia Post-COVID-19. Hacia un Nuevo Contrato Social* (Informes de Investigación No. 019238). Bogotá, Colombia: Fedesarrollo.
- Maloney, W. F. (2004). Informality Revisited. *World Development*, 32(7), 1159–1178.
- Medina, C., & Posso, C. (2010). *Technical Change and Polarization of the Labor Market: Evidence for Brazil, Colombia and Mexico* (Borradores de Economía No. 614). Bogotá, Colombia: Banco de la República.
- Mejía, D., & Posada, C. E. (2007). *Informalidad: Teoría e Implicaciones de Política* (Borradores de Economía No. 455). Bogotá, Colombia: Banco de la República.
- Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6), 1695–1725.
- OECD. (2007). *Eurostat-OECD Manual on Business Demography Statistics*.
- OECD. (2020). *Financing SMEs and Entrepreneurs 2020: An OECD Scoreboard*. OECD Publishing.
- OECD. (2021). *The Missing Entrepreneurs 2021*.
- OECD. (2022). *Financing SMEs and Entrepreneurs 2022: An OECD Scoreboard*. OECD Publishing.
- Paturot, D., Mellbye, K., & Brys, B. (2013). *Average Personal Income Tax Rate and Tax Wedge Progression in OECD Countries* (OECD Taxation Working Papers No. 15). Paris, France: OECD Publishing.
- Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., Mason, A. D., & Saavedra-Chanduvi, J. (2007). *Informality : Exit and Exclusion* (No. 6730). The World Bank Group.
- Rebolledo Abanto, P. A., & Soto Chavéz, R. (2004). Estructura del Mercado de Créditos y Tasas de Interés: Una Aproximación al Segmento de las Microfinanzas. *Revista Estudios Económicos*(11).
- Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8), 2015–2047.
- Ulyssea, G. (2019). *Formal and Informal Business Dynamics*.
- Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics*, 12, 525–546.
- Veyrune, M. R. M., Della Valle, G., & Guo, S. (2018). *Relationship Between Short-Term Interest Rates and Excess Reserves: A Logistic Approach*. International Monetary Fund.
- Vijlberger, W. P. (1992). Measuring income from family enterprises with household surveys. *Small Business Economics*, 4, 287-305.

# Appendices

## A Sources of structural parameters and first guesses of estimated parameters

**Equilibrium wages for skilled and unskilled workers:** To estimate wages that are not affected by the composition of the labor force or the sector composition in the GEIH, a logarithmic regression of the wage was estimated with an adjusted variable for highly qualified workers and other control variables of the business and worker. To further minimize measurement error, the sample is restricted to employees who are between 18 and 69 years old and who have worked at least 20 hours in the reference week, but at most 84 hours (which is the 99th percentile). The lower limit of 20 hours is intended to exclude interns who are still in school and workers with very little connection to the labor market. The estimated regression is used to calculate the adjusted wage for high-skilled workers (COP\$973,097), unskilled workers (COP\$821,430) and unskilled informal workers (COP 688,978).

**Function parameters of the distributions:** The production function follows a Cobb-Douglas distribution, the firm distribution follows a Pareto distribution, and the mix of skilled and unskilled labor, a CES distribution. [Medina and Posso \(2010\)](#) estimated the substitution elasticity for unskilled / skilled labor in Colombia at  $\sigma = 1.47$  and  $\sigma = 1.31$ , if controlling for unemployment and minimum wage. This implies CES parameters of  $\rho = 0.32$ ,  $\rho = 0.24$ , respectively, given  $\sigma = 1/(1 - \rho)$ . Simple shares of skill workers in formal and informal firms (0.713 and 0.451, respectively) were used as initial values to estimate the CES shares. Workers with secondary education are considered skilled workers. The structural parameter of the Pareto location is calibrated at 7.3; and the initial shape parameter is 3, according to the fitting of the data for a firm of minimum size of 2, according to [Jenkins and Kerm \(2007\)](#). The initial parameter used for the Cobb-Douglas coefficient was  $\alpha = 0.4952$ , according to the Penn World tables.

**Exit probability of formal and informal firms:** The Colombian Central Bank estimates the exit rate of formal firms at 19%, using administrative data, which is the

parameter that we use in the estimates for the case of formal firms. For informal firms, I calculated the parameter at 24.8% (respectively) as the monthly average of the change in the number of entrepreneurs in the GEIH (as a proxy for the change in the number of firms), discounting the number of entrepreneurs who have been in business for less than 12 months for the period 2019/2018. It is important to bear in mind that in any case, these parameters not only contain the effect of entry and exit of the firm but also involve the effect of transitions between formality and informality.

**Taxes and contributions:** The income tax ( $\tau_y$ ) was estimated as the relationship between companies' taxes and gross profits (income minus direct or material inputs), according to the Colombian Corporate Supervisory Authority database. The resulting income tax is 28.8%. The VAT tax ( $\tau_{vat}$ ) used is the statutory rate (19%) net of income tax deductions (14.5%). The effective payroll tax rate is estimated using the statutory rates on wages observed in the GEIH for 2019 of formal workers in firms with 2 or more workers and includes: transportation subsidy, severance, interest on severance, additional annual wages (prima), vacations, employer's health, social security and 'Caja de compensacion,' ICBF and SENA contributions (when applicable), and occupational risk insurance. The total contribution amounts to 47.2%. The income tax threshold is established as the monthly equivalent of an annual income of COP 46 million (assuming no deductions<sup>14</sup>), and the threshold of the  $\tau_{vat}$  at COP\$100 millions, assuming no deductions.

**Cost of entry or minimum scale required to participate in the formal and informal sector,** denominated in units of production. This cost is estimated as the total monthly costs, other than materials, of a new firm (less than 1 year) with 2 workers (minimum size), including an entrepreneur (minimum size) at COP\$1'760,000 and COP\$378,000 for formal and informal firms, respectively.

**Annual fix cost:** Assuming that property costs are flexible, the annual fix cost is limited to the formalization cost, which is 0.34% as a percentage of unskilled wage for formal firms and close to 0.4% for informal firms (EMICRON, 2019). The estimation of the model for the Colombian case sets the informal costs as a structural parameter and the formal cost as the flexible, because the estimate for informal firms is more reliable

---

<sup>14</sup>This is an upper limit for the threshold. However, given that to obtain a deduction firms should use an "electronic bill", use electronic transactions, and prove that contractors pay their own social security, it is reasonable to assume this upper limit in the model.

and to avoid negative solutions, without setting additional restrictions.

**Other initial parameters:** Other initial parameters such as the enforcement parameters and the variance of the pre-productivity shocks, and structural parameters such as the mean productivity shocks used [Ulyssea \(2018\)](#) values. Also, individual company productivity is estimated using a grid  $\theta$  that affects income. This grid was created to save computational time and a non-binding high range was used so as not to limit the results. Additionally, a vector of transition probabilities is created for each point on the grid to calculate the expected post-entry values in each sector for each possible participant.

# MSMEs and the Cost of Using Capital

Cristina Fernández\*

## Abstract

Micro-business populate most of the demography of firms in developing countries. However very often this firms are discouraged to hire skilled workers and to accumulate capital, and therefore to be productive. The lack of incentives to accumulate capital occurs not only because interest rates are decreasing in capital, but also because they are subject to tax deductions that cannot be used by firms too small to be subject to taxes. This paper estimates a general equilibrium model that includes an endogenous extensive and intensive margin of informality, a tax structure that considers waivers to small firms and deductions to larger firms, capital and a cost of using it that follows a logistic function behavior, resembling what is found in the data. The model is calibrated for the case of the manufacturing sector in Colombia. Simulations suggest that policies oriented to increase the use of capital in one group, without any more policies to reduce the average cost of using capital, imply an increase of the cost of using capital for other groups. Therefore, policies to reduce the informal distortion of using capital can increase formal firm's productivity but reducing the size of the formal sector, thus harming overall productivity. On the other hand, policies oriented to reduce the size distortion of using capital can increase formality maintaining output and productivity relatively stable, because both tales of the productivity distribution are equally affected.

**Keywords:** Informality, Microbusiness, Business Informality, Cost of Using Capital, Interest Rates, Microcredit.

**JEL codes:** C68, D24, D58, J46, L26, L60

---

\*Economics PhD student at Universidad del Rosario and researcher of Fedesarrollo. I would like to thank the invaluable help of Andres García-Suaza, Fernando Jaramillo and Juan Miguel Gallego in developing this paper and the assistance of Hector Segura and Cecilia Suescún. Comments by Marcela Eslava, Luis Fernando Mejía and the assistants to EAFIT and los Andes seminars were also very helpful. This working paper is funded by the Colombia Científica- Alianza EFI Research Program, with code 60185 and contract number FP44842-220-2018, funded by The World Bank through the call Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology, and Innovation. I also would also like to thank Open Society that financed a previous version of this paper.

## A Introduction and literature review

Capital accumulation is crucial to promote long-term growth. However, factors such as information failures and informality create a distortion that impacts access to capital among small and/or informal firms. Therefore, analyzing capital accumulation in a setup of informality and other distortions inherent to small firms is crucial for developing countries. Although there are several theoretical frameworks studying the causes and consequences of informality, they usually do not consider capital accumulation. In contrast, there are some studies concentrated on capital accumulation, mainly in larger firms, that do not take into account informality.

One of the papers that better reproduces the main characteristics of informality in Latin America from the point of view of firms is [Ulyssea \(2018\)](#). This model, estimated for the case of Brazil, considers the intensive and extensive margin of informality as decisions taken by the firm after performing a cost-benefit exercise. [Alvarez and Ruane \(2019\)](#) estimated a similar model for the case of Mexico, but allowing some idiosyncratic shocks that are different for formal and informal firms, and finally, [Fernández \(2023a\)](#) also estimated a similar model for the case of Colombia, but considering the impact of profit taxes waivers for small firms, and labor cost deductions for larger firms. Most of this type of models do not include capital, which might affect inferences about the relevance of informality to shape public policy impacts.

However, there is some existing literature on capital costs and informality that assumes an extra cost of capital adjustment in the informal sector. For instance, [D’Erasmus and Boedo \(2012\)](#) developed a framework in which they estimate the impact of informality on consumption and wealth for different levels of development of the credit market, understood as the ability to use collateral to access credit. Similarly, [Loayza \(2016\)](#) highlights the problem of improper allocation generated by a distortion in relative costs: formal firms face higher labor costs, while informal firms face higher capital costs and lower productivity. [Chisari, Romero, and Martínez \(2016\)](#), through a general equilibrium model calibrated for Argentina, which includes a fixed extra financing cost for informal firms, estimated the gains in productivity that allow firms to become formal and access capital markets. [De Paula and Scheinkman \(2011\)](#) found that the Brazilian informal sector faces a 1.3 higher cost of funds than the informal sector, plus a higher enforcement. Therefore, according to the authors, there is a discontinuity in the level

of capital and labor employed at a fixed level of productivity. Consistently, a program oriented to reduce the capital cost of small, mostly informal firms, might increase the use of capital in small firms and their income level. [Erosa, Fuster, and Martinez \(2023\)](#), in their model calibrated for Brazil, find that financial frictions operate as a cost to high productivity firms, which reinforced with the possibility of operating informally, create an advantage to low productivity entrepreneurs.

For the case of Colombia, [Hamann-Salcedo and Mejía \(2013\)](#) developed an endogenous informality model in which formal firms are allowed to be financed at the international interest rate, while the informal sector only has the ability to obtain limited domestic credit. [Granda and Hamann \(2015\)](#) replicate well the Colombian economy and the savings patterns of workers and firms by assuming interest rate differentials and labor and capital segmentation.

In order to partially fill the identified gap in the literature, the research question of this paper is how firms' decisions and outputs are affected by distortions in accessing to capital and the impact of removing those distortions. The paper advances on the same line as [Erosa et al. \(2023\)](#), but in this case the distortion takes two forms: One related to size, because small firms tend to be riskier, suffer information failures, and do not get tax incentives if they are below the exempt-tax bracket. The other distortion is related to informality since the business informality condition implies that often firms avoid getting in contact with financing institutions, and if they contact them, they cannot use their capital as collateral, increasing the cost of using capital. In addition, information failure tends to be larger in informal firms due to lack of documentation.

Accordingly, the main contribution of this paper is to develop a model with three main features. The first adopts the basic structure of [Ulyssea \(2018\)](#) that includes an extensive and intensive informality margin, small and large firms, and skilled and unskilled workers. The second incorporates the same tax structure of [Fernández \(2023a\)](#), which considers tax waivers to small firms and tax deductions oriented to larger firms. The third involves capital, and a cost of capital use, modeled by using a logistic function, following [Veyrune, Della Valle, and Guo \(2018\)](#). The use of this type of function allows to estimate interest rates in a dual-economic environment without going back to the type of models that assume different production functions for each sector.

The model is calibrated for the case of the manufacturing sector in Colombia. Sim-

ulations suggest that policies to redistribute capital across informal and formal firms might lower the overall cost of accessing capital, but generate a negative impact on overall productivity and informality. On the other hand, policies oriented to reduce the size distortion of using capital can increase formality maintaining output and productivity relatively stable, because both informal small firms and larger formal firms get affected.

The rest of this paper proceeds as follows: Section 2 presents the data sources and main facts for the case of Colombia, Section 3 illustrates the structure of the model, Section 4 is concentrated on calibration issues, Section 5 shows the results of policy simulations, and Section 6 concludes.

## **B Data sources and facts. The case of Colombia**

Microbusiness in Colombia accounts for approximately 93% of business, 30% of workers, and 8.3% of the value added in urban Colombia, excluding agriculture, government and self-employment (Fernández, 2023d). Some of this behavior could be encouraged by the tax system. 77% of these companies earn less than COP\$46 million (US\$14,000 per year at 2019 prices) the lower limit for paying profit taxes in the country (EEG,2019). As explained by Fernández (2023a), firms with income below this threshold have strong incentives to not hire workers formally, and in consequence find harder to hire skilled workers. This paper explains how these firms are also discouraged from accumulating capital. Therefore, Colombia's business demography ends up concentrated in very small unproductive firms because of the lack of incentives to use capital and/or hire skilled workers.

This section describes the data sources used to calibrate the model for the case of Colombia and provides some general characteristics of Colombian firm structure including labor and capital distributions of formal and informal firms, as well as some international comparisons. This overview is key to understanding the basic assumptions of the model and its policy implications.

The EEG (2019), the main source of data in this document, is a database that compiles different company-level surveys in Colombia (Fernández, 2023d). The scope of the analysis is restricted to the manufacturing industry, due to the lack of information on capital use in other sectors. The manufacturing sector in EEG (2019) is compiled in

the following way: 1- Firms with more than one worker (excluding self-employment) and less than ten workers are represented by the EMICRON, which is an employer-employee database linked to the household sector<sup>1</sup>; 2- Formal firms with more than 10 workers are collected from the EAM (2019) a manufacturing firm level census; 3- Informal firms with more than 10 workers are estimated by using the household survey (GEIH) questions directed to entrepreneurs, and by assuming that each entrepreneur represents one firm. The GEIH is a cross-sectional survey in which variables are representative at a monthly or quarterly level.

As shown in Table 1, after excluding self-employment, there are 104 thousand firms (7,770 observations)<sup>2</sup> of which 75 thousand are registered in the Chamber of Commerce. This amount can be compared with the total number of registered firms compiled by the Dane: 82 thousand (DEE, 2023), showing a raw representativeness of 91%.

Similarly, EEG (2020) accounts for 900,000 workers in the manufacturing sector 57% to the 1591 thousand manufacturing sector entrepreneurs estimated by the GEIH (excluding self-employment). More details about this database can be found in [Fernández \(2023d\)](#).

**Table 1. Representativeness of data source (thousands)**

	# Firms		# Workers
EMICRON	97	EMICRON	323
EAM	6	EAM	558
GEIH	1.4	GEIH	19
TOTAL EEG	104	TOTAL EEG	901
EEG Formal	75		
DEST	82	GEIH (no self)	1591
EEG Formal / DEST	0.91	EEG / GEIH	0.57

Source: Own calculations based on DEST (2023), GEIH, EEG.

One of the main characteristics of the firm structure of the manufacturing sector in Colombia is the large number of small companies. Within the manufacturing sector, even after excluding self-employment, 91% of the firms have equal or less than 10 workers, 5.4% between 10 and 50 workers, and 2.6% more than 50 workers.

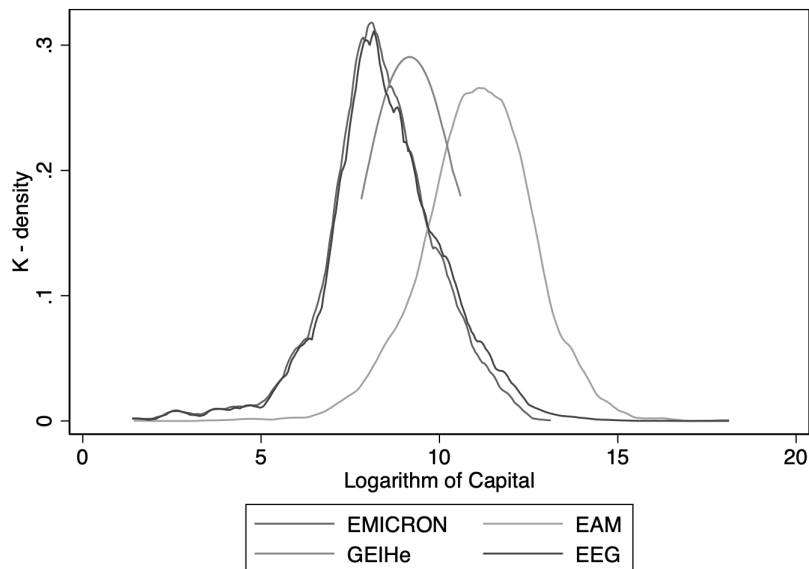
The capital stock in EMICRON and EAM is collected through the question that asks

<sup>1</sup>In line with the ILO criteria, this survey restricts the definition of firm to those with power of disposal over the production means.

<sup>2</sup>97 thousand were collected from EMICRON (1581 observations), 6 thousand from the EAM and 1.4 thousand from the GEIH

entrepreneurs to value the total assets used under the productive processes. The capital stock in informal firms with more than 10 workers and missing values (1.3% of the EEG) are imputed using the random forest technique to estimate the logarithm of the capital/labor relationship. Figure 1 shows the result of the imputation. As it can be shown, the imputed capital-to-labor ratio of informal firms with more than 10 workers lies between the distribution of the same relationship for micro-business and formal firms with more than 10 workers. This imputation does not significantly change the overall distribution of firms.

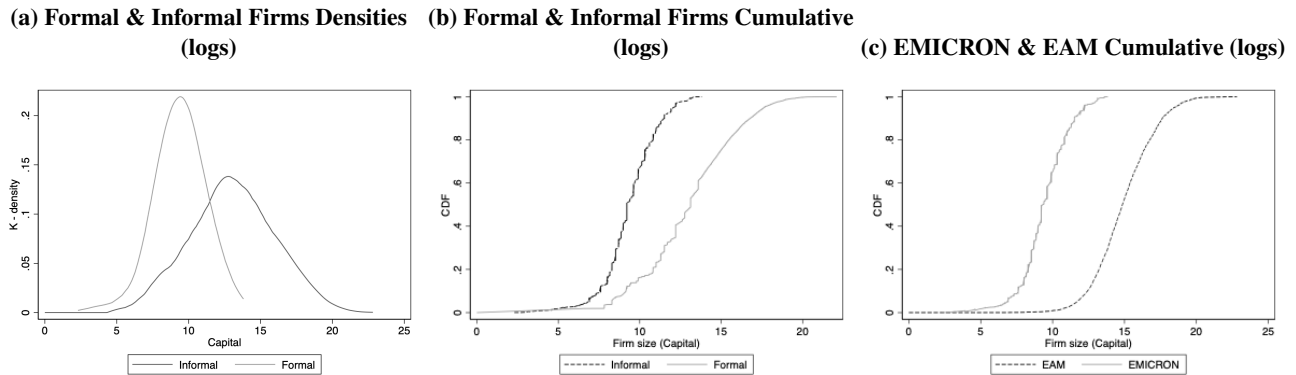
**Figure 1. Logarithm of capital to labor ratio**



Source: EEG (2019).

The capital stock distributions, shown in Figure 2, evidence some “missing middle” behavior in the case of Colombia, particularly among informal firms. This behavior is observed even if the survey is restricted to EMICRON. Figure 2a shows the density distribution of the capital logarithm for formal and informal firms, Figure 2b shows the respective cumulative distribution and Figure 2c shows the cumulative distribution of EMICRON and EAM. The differences between Figure 2b and Figure 2c suggest that the difference observed between informal and formal firms is not due to the source of the observations.

**Figure 2. Capital stock distributions**



Source: EEG (2019).

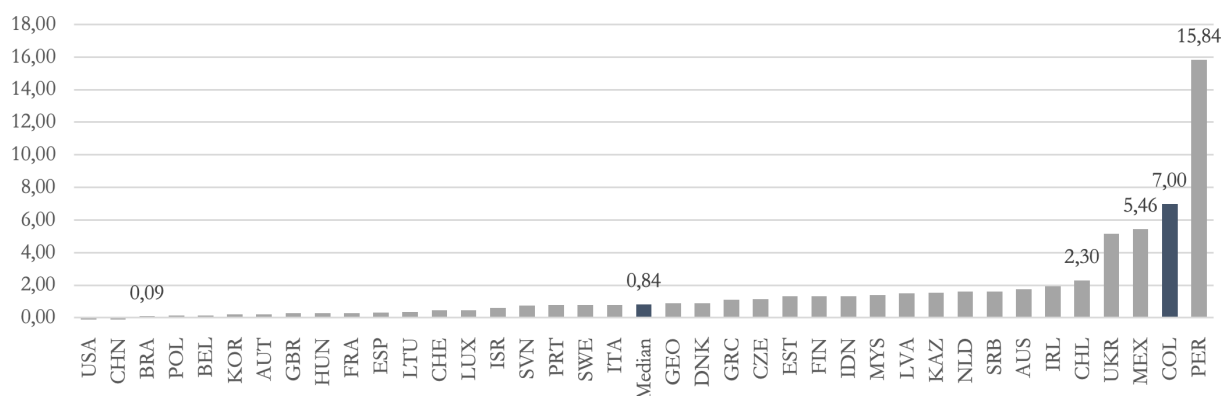
Differences in capital costs are widely known in developing countries, where small and particularly small informal firms have last-resort access to the black-market finance sector. According to EMICRON, most Colombian microbusinesses recognize that they get funds whenever they ask for them. However, 14% of the informal firms get their credits through informal means (friends 6%, black market 9%), and even 8% of the formal micro business get their access to credit among informal markets (friends 7%, black market 1%). The interest rates in the black market can be several times higher than the formal market interest rates<sup>3</sup>.

The lack of incentives faced by microbusiness to accumulate capital takes two forms: the first one is related to size. Interest rates tend to decrease in capital, since capital guarantees credits. In line with this, there is a spread between the interest rates charged to SMEs and larger firms, which is particularly large in Colombia (7 pp per year vs. a median of 0.84 in the OECD countries), as seen in Figure 3.

On top of this, Colombia has an additional spread for MSMEs that, according to the OECD (2020) averaged 9pp between 2010 and 2018, as shown in Figure 4. This spread covers both formal and informal microbusiness, that tend to show higher costs for informal informal business, as it is more risky to lend to informal firms due to lack of documentation.

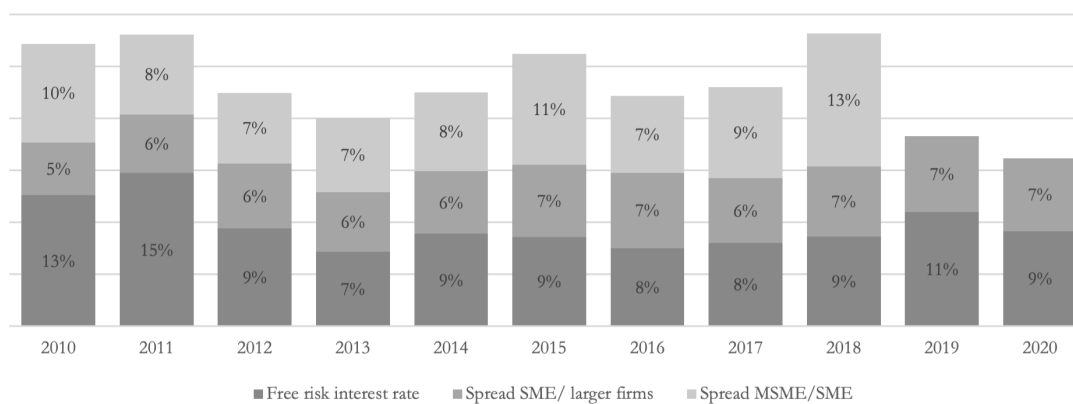
<sup>3</sup>Nevertheless, it is also possible that some informal small firms do not ask for means to finance capital because they do not need it, or they think that they do not need it. In fact, this is the main reason for 40% of employers. As explained above, this would imply a different production function for informal and formal firms. Another particularity is an important self-selection process in the sense that many employers do not ask for credit in formal markets because they think that they will not get access to it. Therefore, businesses might do better with capital, but they prefer to act as if they don't.

**Figure 3. Interest rates spreads SMEs/larger firms in the OECD (2020)**



Source: OECD (2022)

**Figure 4. MSMEs/SMEs spread in Colombia**



Source: Own calculations based on OECD (2020) and OECD (2022)

## C The model structure

The main challenge of this paper was to estimate a model that accounts for different distributions of capital among formal and informal firms, without going back to the type of model that assumes different production functions for each sector, and therefore dualism.

The model also adopts the same tax structure of Fernández (2023a) with waivers and deductions, which resembles that of many developing countries. This specification is

key, as it shapes the use of capital and labor in small and large firms. In fact, as shown in Fernández (2022), firms that pay taxes have high incentives to hire formal workers, since they can tax-deduct interest payments and all the formal labor costs, including wages and social security; meanwhile, small firms do not have access to those incentives since they do not pay taxes. The tax structure also affects capital accumulation.

The model also includes the main features of Ulyssea (2018) such as an intensive and extensive margin of informality, small and large firms, and skilled and unskilled workers. The inclusion of intensive and extensive margins is important since both types of informality do not always react in the same direction. The same is true for large and small firms, because the former pay taxes but have incentives to hire formally and to accumulate capital, which is not the case of small firms; and for skilled and unskilled workers, since large firms need a larger number of qualified workers.

In the model, formal and informal firms face the same production function that uses a Cobb-Douglas specification to mix labor and capital. In turn, labor is a CES's mix of skilled and unskilled labor. Formal firms pay taxes if they are above the tax threshold, and social security for those workers that are hired formally, or a cost associated to the inherent risk of informal hires, otherwise. Meanwhile, informal firms do not pay taxes or social security but face a cost associated with the risk of operating informally. Both cost of informality in the intensive and extensive margin are increasing in capital, so larger firms tend to run and hire formally. The model also considers a cost of using capital that is increasing in capital and faces a surcharge if the firms operate informally. More specifically, the structure of the model is the following:

**Production function.** The basic model assumes one good, the same production function with decreasing returns, and the same prices for formal and informal firms (normalized to 1). This feature is key to avoid assuming a complete dual market. Formal firms are subject to taxes if their gross income is above the respective thresholds.

$$y_S(\theta) = (1 + \tau_y)(1 + \tau_{VAT})\theta L_S^\alpha K_S^\beta,$$

*with  $(\alpha + \beta) < 1$  and  $s = I, F$*

$$\tau_y = 0 \text{ if } \theta L_S^\alpha K_S^\beta < Y_Y \text{ or } S = I, \tau_{VAT} = 0 \text{ if } \theta L_S^\alpha K_S^\beta < Y_{VAT} \text{ or } S = I$$

where I refers to informal and F refers to formal in terms of the extensive margin,  $y_S$  is the income of the formal/informal sector,  $\theta$  denotes ex-ante-productivity,  $\tau_{au_y}$  and

$\tau_{VAT}$  income tax and VAT, and  $Y_Y$  and  $Y_{VAT}$  the respective thresholds.

**Fix and extensive margin costs.** Firms face fix costs of production, larger on informal firms ( $FC_F > FC_I$ ) that are tax-deductible. Additionally, informal firms face enforcement costs based on the probability and cost of being caught by the authorities that is proportional to labor costs, since larger firms are easier to be detected. Parameters  $b_I$  and  $d$  represent how strong enforcement is and how much it changes with size.

$$T(L_I) = \left(1 + \frac{L_I^d}{b_I}\right)$$

**Cost of using capital.** Capital costs face two frictions: one related to size and another related to informality. In order to model interest rate, a logistic function is introduced. This allows the cost of capital use to decrease in the amount of capital, in line with Estrada and Hernández Rubio (2019), Chowdhury (2009), and Rebolledo Abanto and Soto Chavéz (2004), but allows an upper and a lower bound following Veyrune et al. (2018). The function can be written as follows:

$$c_{KS} = \left(i + \gamma_I + \frac{\gamma_{MS}}{1 + e^{(a_S K_S + c_S)}}\right), \text{ with } S = I, F$$

Where  $i$  is the free risk interest rate,  $\gamma_I$  is the spread face by informality,  $\gamma_{MS}$  is the spread face by micro-business, and  $a_S, c_S$  are the shape and location parameters of the logistic function, respectively. Capital costs are assumed to be tax-deductible.

**Worker's qualification.** Two levels of qualification (skilled ( $l_1$ ) and unskilled ( $l_2$ )) are considered and mixed through a CES function, with parameter  $\rho$  and elasticity of substitution  $\eta_S$  in the following way:

$$L_S = \left(\eta_S l_1^\rho + (1 - \eta_S) l_2^\rho\right)^{\frac{1}{\rho}}, S = I, F$$

Workers perform the same activities within the firm, subject to their level of qualification. They are hired according to their relative costs, but there is a critical threshold  $\tilde{l}$  below which only informal workers are hired.

**Intensive margin costs.** Formal firms can hire informally or formally. If they hire informally, they face an enforcement cost  $T(l_{FI(1,2)}) = \left(1 + \frac{l_{FI(1,2)}^d}{b_{1,2}}\right)$ , where  $b_{1,2}$  is the parameter decreasing on enforcing, and  $d$  a parameter that indicates the intensity of the relationship. If they hire formally, they pay social security and a minimum wage  $\max(w_{1,2}, w_{min})$ , where  $w_{1,2}$  is the equilibrium wage for formal and informal firms and  $w_{min}$  the minimum wage, but have access to a tax deduction on labor costs  $(1 + \tau_w)(1 - \tau_y)$ . The threshold below which firms hire only informal workers ( $\tilde{l}_{1,2}$ ) is obtained by equalizing the marginal cost of hiring formally and informally.

$$w_{1,2} \left(1 + (d+1) \frac{l_{FI(1,2)}^d}{b_{1,2}}\right) = (1 + \tau_w)(1 - \tau_y) \max(w_{1,2}, w_{min})$$

$$\tilde{l}_{FI(1,2)} = \left[ \frac{b_{1,2}}{(d+1)} \left\{ (1 + \tau_w)(1 - \tau_y) \frac{\max(w_{1,2}, w_{min})}{w_{1,2}} - 1 \right\} \right]^{\frac{1}{d}}$$

**Maximization problems.** Given that the threshold does not change with size, it is possible to reduce the number of variables by rearranging the maximization problem to four types of firms, as follows:

A. Informal firms

$$\max_{l_{i1}, l_{i2}} \pi_I = \theta L_I^\alpha K_I^\beta - \left(1 + \frac{l_{I1}}{b_I}\right) w_1 l_{I1} - \left(1 + \frac{l_{I2}}{b_i}\right) w_2 l_{I2} - cK_I - FC_I$$

$$s.t. L_I = \left(\eta_I l_{I1}^\rho + (1 - \eta_I) l_{I2}^\rho\right)^{\frac{1}{\rho}}, L_I > 0, \alpha + \beta < 1, w_1 > w_2$$

B. Formal firms that hire only informal ( $l_1 < \tilde{l}_1$  &  $l_2 < \tilde{l}_2$ )

$$\max_{l_{i1}, l_{i2}} \pi_F = (1 - \tau_y - \tau_{VAT}) \theta L_F^\alpha K_F^\beta - \left(1 + \frac{l_{FI1}}{b_1}\right) w_1 l_{FI1} - \left(1 + \frac{l_{FI2}}{b_2}\right) w_2 l_{FI2} - cK_F - FC_F$$

$$s.t. l_F = \left(\eta_f l_{F1}^\rho + (1 - \eta_f) l_{F2}^\rho\right)^{\frac{1}{\rho}}, l_F > 0, \alpha + \beta < 1, w_1 > w_2 \text{ \& } b_2 > b_1 > 0$$

C. Formal firms that hire some formal skilled workers ( $l_1 > \tilde{l}_1$  &  $l_2 < \tilde{l}_2$ )

$$\begin{aligned} \max_{l_{f1}, l_{i2}} \pi_F &= (1 - \tau_y - \tau_{VAT})\theta L_F^\alpha K_F^\beta - \left(1 + \frac{\tilde{l}_1}{b_1}\right)w_1\tilde{l}_1 - \left(1 + \frac{l_{FI2}}{b_2}\right)w_2l_{FI2} \\ &\quad - (1 - \tau_w)(1 - \tau_y)w_1l_{FF1} - cK_F - FC_F \\ s.t. \quad l_F &= \left(\eta_i l_{F1}^\rho + (1 - \eta_f)l_{F2}^\rho\right)^{\frac{1}{\rho}}, \quad l_F > 0, \quad \alpha + \beta < 1, \quad w_1 > w_2, \quad b_2 > b_1 > 0, \\ l_{FFs} &= l_{Fs} - \tilde{l}_s, \quad s = 1 \end{aligned}$$

D. Formal firms that hire some formal skilled and unskilled workers ( $l_1 > \tilde{l}_1$  &  $l_2 > \tilde{l}_2$ )

$$\begin{aligned} \max_{l_{f1}, l_{i2}} \pi_F &= (1 - \tau_y - \tau_{VAT})\theta L_F^\alpha K_F^\beta - \left(1 + \frac{\tilde{l}_1}{b_1}\right)w_1\tilde{l}_1 - \left(1 + \frac{\tilde{l}_{FI2}}{b_2}\right)w_2\tilde{l}_2 \\ &\quad - (1 - \tau_w)(1 - \tau_y)(w_1l_{FF1} + \max(w_2, w_{min})l_{FF2}) - cK_F - FC_F \\ s.t. \quad l_F &= \left(\eta_i l_{F1}^\rho + (1 - \eta_f)l_{F2}^\rho\right)^{\frac{1}{\rho}}, \quad l_F > 0, \quad \alpha + \beta < 1, \quad w_1 > w_2, \quad b_2 > b_1 > 0, \\ l_{FFs} &= l_{Fs} - \tilde{l}_s, \quad s = 1, 2 \end{aligned}$$

**Entry and exit.**  $M$  firms enter the market and observe  $\nu$  (iid), which is a parameter that distorts the true productivity ( $\theta$ ), observed after entering the market. To enter the formal sector, firms must pay a fixed cost:  $E_F > E_I$ . If firms face unexpectedly low productivity, they exit the market before they start producing.

**Consumers.** Representative household inelastically supplies  $L$  and derives utility solely from consuming the final good,  $welfare = wL + \pi$  (*profits*) +  $T$  (*taxes*), assuming taxes are returned to consumers with an efficiency rate of 100%.

## D Calibration of the model

The model is calibrated through a two-stage minimum distance estimator (SMD) that uses the value of the structural parameters, the guess values of the estimated parameters, and the equations and functional forms of the model to generate moments of the distribution functions of some key variables of the economy. Moments of simulated data distributions are compared with the same moments obtained from the real data distributions, and the estimated parameters are adjusted until the model converges.

## D.1 Structural and estimated parameters

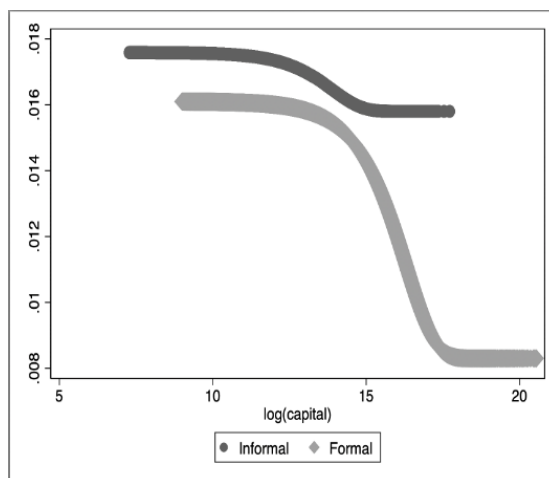
The structural variables of the economy and their respective sources are shown in Table 2. These parameters are very similar to the ones used in Fernández (2023a) but adjusted to the manufacturing sector. The cost of using capital for formal and informal firms is calibrated to fit capital, and according to Figure 5, this cost is always higher for informal firms, but decreases for capital in both formal and informal firms. Table 3 shows the estimated parameters that are in line with what was expected for the manufacturing sector<sup>4</sup>.

Table 2. Structural parameters

		Parameters	Source
$w_1$	Equilibrium wage of skilled workers	1232	GEIH
$w_2$	Equilibrium wage of unskilled informal workers	759	GEIH
$i$	Free risk interest rates (yearly)	10.4%	OECD
$\tau_w$	Payroll/Payroll Contributions	0.47	Fernández & Mejía (2019)
$\tau_y$	Income taxes	0.29	Statutory
$V_0$	Pareto parameter, ex ante productivity	7.00	Calibrated
$k_f$	Exit probability of formal firms	0.11	GEIH (firms)
$\frac{C_I}{w_2}$	Cost of operating in the informal sector / unskilled wage	0.2	Calibrated
$\tau_{VAT}$	VAT tax (net of income tax deduction)	0.145	Statutory
$y_{ty}$	Monthly threshold for paying income tax (40% deductions)	3833	Statutory
$y_{tVAT}$	Monthly threshold for paying VAT	8333	Statutory

<sup>4</sup>The source of the initial estimated parameters in the model is the following: cost of the intensive and extensive margin: Fernández (2022), but the final values are lowered since the specification of the cost function varies. Exit probability of firms: calculated as the change in the number of entrepreneurs minus the number of new entrepreneurs : GEIH (2018 and 2019). Cost of operating in the formal sector, Pareto shape parameter, cost of entering the formal and informal sector, Cobb-Douglas parameter, ex post productivity: Fernández (2021). CES shares and Beta: EEG(2019). However, all these parameters were modified to fit the data distributions

**Figure 5. Cost of using capital**



	Formal	Informal
$i_{inf}$	10%	21%
$i_{sup}$	22%	23%
a	$-1 \times e^{-6}$	$-1 \times e^{-8}$
c	-3	-5

Source: Own calculations.

**Table 3. Estimated parameters**

		Parameter
$b_f$	Intensive margin cost, skilled workers	10.8
$b_f$	Intensive margin cost, unskilled workers	12.6
$b_i$	Cost of extensive margin	1.6
$k_f$	Exit probability of formal firms	0.2
$\frac{C_F}{w_2}$	Cost of operating in the formal sector / unskilled wage	0.2
$x_i$	Pareto shape parameter	2.3
$E_f$	Cost of entering the formal sector	1271
$E_i$	Cost of entering the informal sector	1024
$\alpha$	Cobb Douglas function parameter	0.4
$\sigma$	Productivity variance ex post	0.7
$\rho$	CES elasticity parameter	0.3
$\eta_I$	Skilled worker CES share. Informal firms	0.2
$\eta_F$	Skilled worker CES share. Formal firms	0.3
$\beta$	Beta	0.1

Source: Own estimations based on EEG (2019) and initial values based on Fernández (2021 and 2022), EEG(2019) and GEIH (2018 and 2019) and according footnote 3.

Table 4 shows the model fit in terms of adjustment to some moments of the labor, capital, and firm's distributions. As shown in this table, the model is accurate to esti-

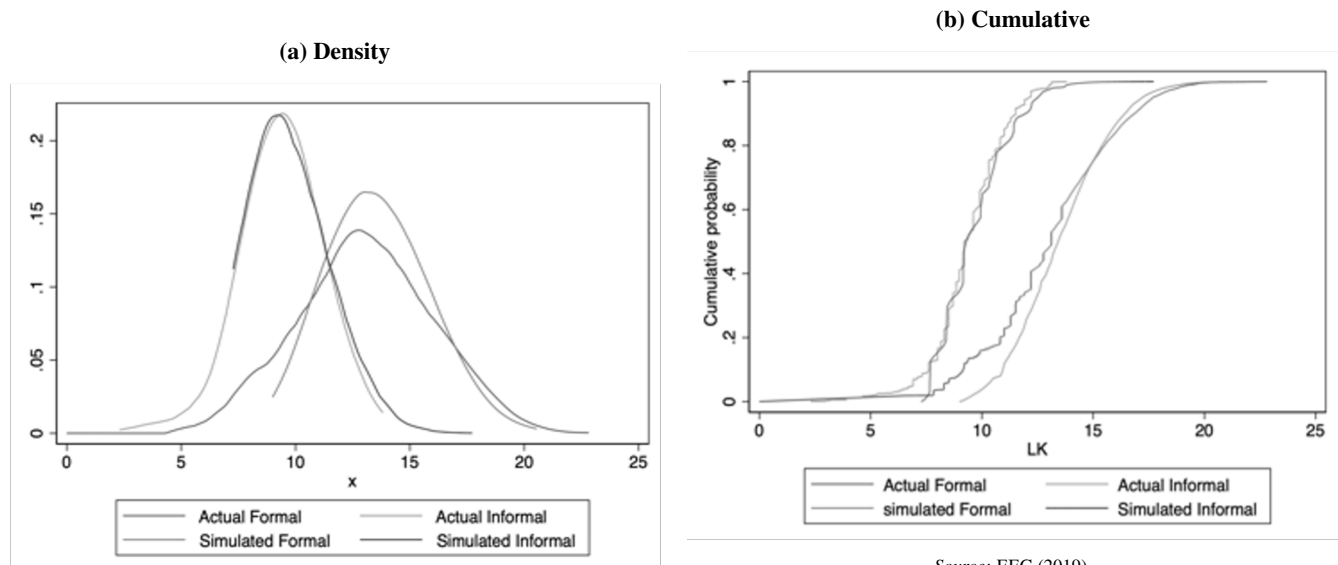
mate distributions in most variables, but tends to overestimate the intensive margin and underestimate the number of formal firms with less than 10 workers. Figure 6 shows the fit of the capital stock distributions for formal and informal firms.

**Table 4. Model fit**

	<b>Actual</b>	<b>Simulated</b>
Informal workers (%)	0.31	0.31
Unskilled	0.37	0.34
Skilled	0.12	0.14
% of capital used by microbusiness	0.01	0.01
% of capital used by informal firms	0.02	0.04
Ln K/L formal firms	10.3	10.8
Ln K/L informal firms	8.3	8.3
% informal firms	0.88	0.87
<10 workers	0.94	0.93
Intensive margin	0.31	0.39
Formal firms <10 workers	0.49	0.40
Formal firms 10-20 workers	0.13	0.12
Formal firms 20-40 workers	0.12	0.12
Formal firms 40+ workers	0.27	0.35

Source: Own estimations based on EEG (2019).

**Figure 6. Distribution of capital fit**



Source: EEG (2019).

Source: EEG (2019).

## E Policy simulations

Two sets of policies oriented to increase access to capital were simulated: Eliminating the friction of the informal firm and reducing the friction of the small firm.

### E.1 Eliminating the informal firm’s friction in using capital

In this simulation, the same parameters for formal and informal logistic functions were used, as shown in Table 5. It means that informal firms will face a lower cost of using capital, but not as low as the free-risk interest rate. An important consideration in analyzing this scenario is how the interest rate reduction will be implemented. Indeed, it should be considered that the total amount of capital is fixed, and therefore, any attempt to reduce the cost of using capital for a group might mean an increase in the cost of capital for other groups. Not only that, but also this type of policy is often costly and implies an increase in taxes. Consequently, an alternative scenario is proposed, in which there is an increase in profit taxes, which encompasses the reduction of finance costs for informal firms.

**Table 5. Baseline and simulation parameters for informal firms**

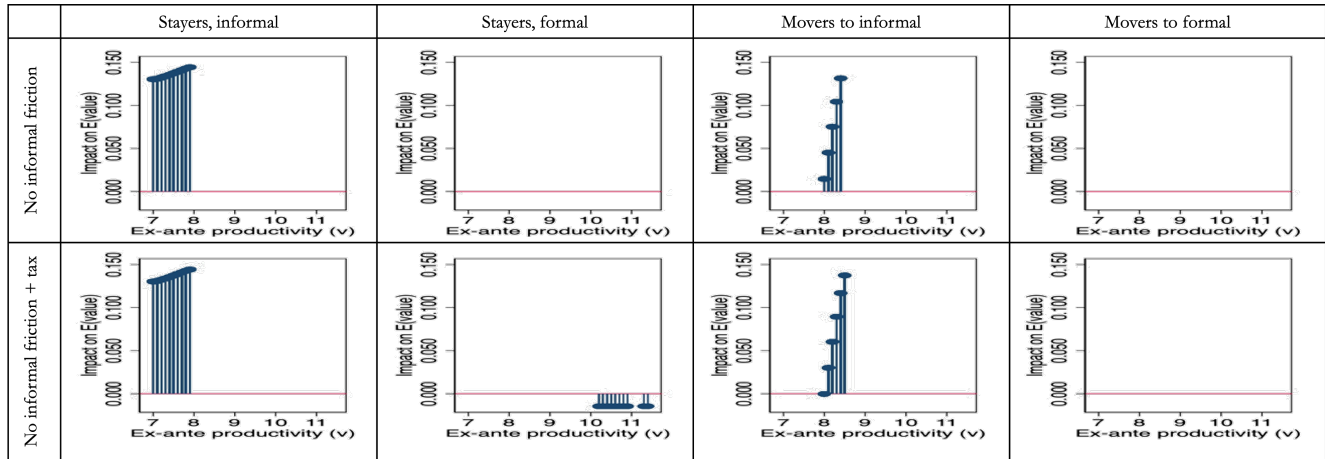
	<b>Formal</b>	<b>Informal baseline</b>	<b>Informal simulation</b>
$i_l$	10%	21%	10%
$i_h$	22%	23%	22%
<b>a</b>	$-1 \times e^{-6}$	$-1 \times e^{-8}$	$-1 \times e^{-6}$
<b>c</b>	-3	-5	-3

Source: Own estimations based on EEG (2019).

Figure 7 shows the initial impact of removing the distortion of the informal firm’s interest rate on the present value of future profits according to their ex ante productivity and for the following groups of firms: i) firms that are informal in the baseline and policy scenarios or ”stayers”, informal, ii) firms that are formal in the baseline and policy scenarios or ”stayers formal”, iii) firms that are formal in the baseline and informal in the policy scenario or ”movers to informal”, and, iv) firms that are informal in the baseline and formal in the base scenarios ”movers to formal”. In the figures, the firms were sorted by their exogenous productivity. According to Figure 8, in the short run ( $t=0$ ) and assuming no change in taxes, being informal becomes more productive (column 1), creating incentives for some formal firms to become informal (column 3), at a point

where small formal firms disappear. If taxes are charged, these policies also harm larger formal firms (column 1 row 2).

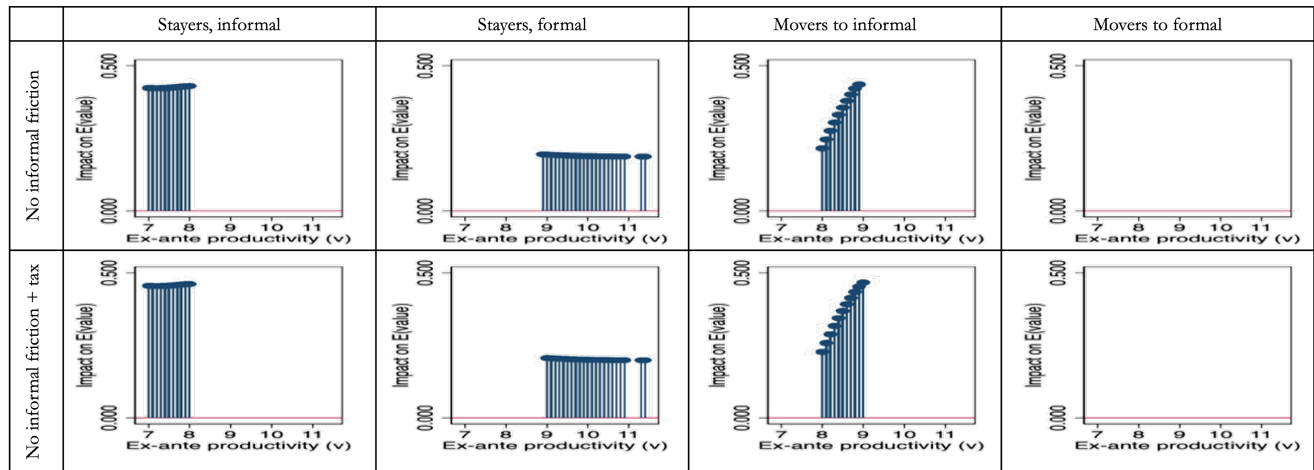
**Figure 7. Micro effects of reducing the informal firms' distortion in accessing to capital (t=0)**



Source: Simulated data based on EEG (2019)

In a more medium-term scenario (t=1), the risk-free interest rate decreases, since firms that move to the informal market use less capital being informal than being formal. In this scenario, even formal firms benefit due to the lower cost of capital use (column 2).

**Figure 8. Micro effects of reducing the informal firms' distortion in accessing to capital (t=1)**



Source: Simulated data based on EEG (2019)

Table 6 shows the overall macro-impact of removing the restriction of the informal

firm on accessing capital. In the proposed scenario, the capital gets cheaper for informal businesses (from 1.8% to 1.2%), and they become more capital intensive (the capital share of informal firms increases from 4% to 26%). On the other hand, the formal sector is now concentrated in larger, more productive capital-intensive firms, and the share of formal microbusiness capital among total capital of formal firms tends to zero.

In this scenario, there is also an increase in business informality from 88% to 99%. In fact, small formal firms disappear, moving the economy towards dualism (small informal firms and large formal firms). Consistently, the share of informal output increases from 20% to 60%. This shows the importance of including capital in the informality analysis.

The lower costs of formal firms (from 1.3% to 0.8%) encourage some formal firms to become informal, but also welcome the entry of a number of informal firms (informality increases from 88% to 99%). Informal firms show lower productivity and do not taxes, and therefore general productivity and tax indicators of the economy worsen (tax revenues decrease from 38% to 19%) . The income effect on informal GDP is positive, but the substitution effect is negative and larger, therefore, the output worsens.

**Table 6. Macro effects of reducing the informal firm’s distortion in accessing to capital (t=1)**

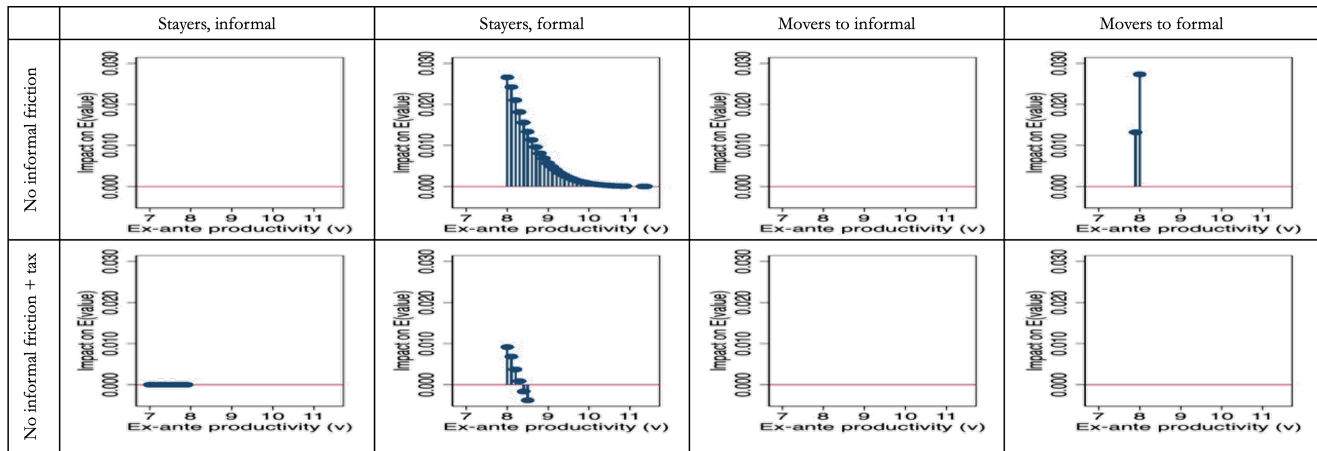
	Baseline	Capital	
		Friction informal	Friction informal + increase profit tax
<b>Informal firms</b>	87.5%	99.3%	99.5%
<b>Informal GDP</b>	20.2%	59.0%	63.3%
<b>Mass of firms (millions)</b>	1.38	1.36	1.36
<b>Informal workers</b>	29.0%	65.9%	69.36%
<b>Skilled informal workers</b>	13.3%	37.9%	41.83%
<b>Unskilled informal workers</b>	31.8%	70.9%	74.26%
<b>Skilled wage</b>	1,302	881	837
<b>Unskilled wage</b>	774	653	646
<b>Interest rate</b>	0.8%	0.6%	0.6%
<b>Capital cost formal firms</b>	1.3%	0.8%	0.7%
<b>Capital cost informal firms</b>	1.8%	1.1%	1.1%
<b>Capital share of informal firms</b>	0.04	0.26	0.29
<b>Capital share of formal microfirms</b>	0.01	0.00	0.00
<b>ln (K/L) formal</b>	10.87	11.35	11.42
<b>ln (K/L) informal</b>	8.29	8.97	9.02
<b>TFP (%baseline)</b>	9.92	9.73	9.70
<b>Olye &amp; Parks</b>	1.77	1.57	1.59
<b>Output (mm)</b>	165.22	144.36	142.63
<b>Tax revenues</b>	38.04%	19.42%	17.51%
<b>Welfare</b>	1348	1135	1133

Source: Simulated data based on EEG (2019).

## E.2 Reducing size firm's friction in using capital

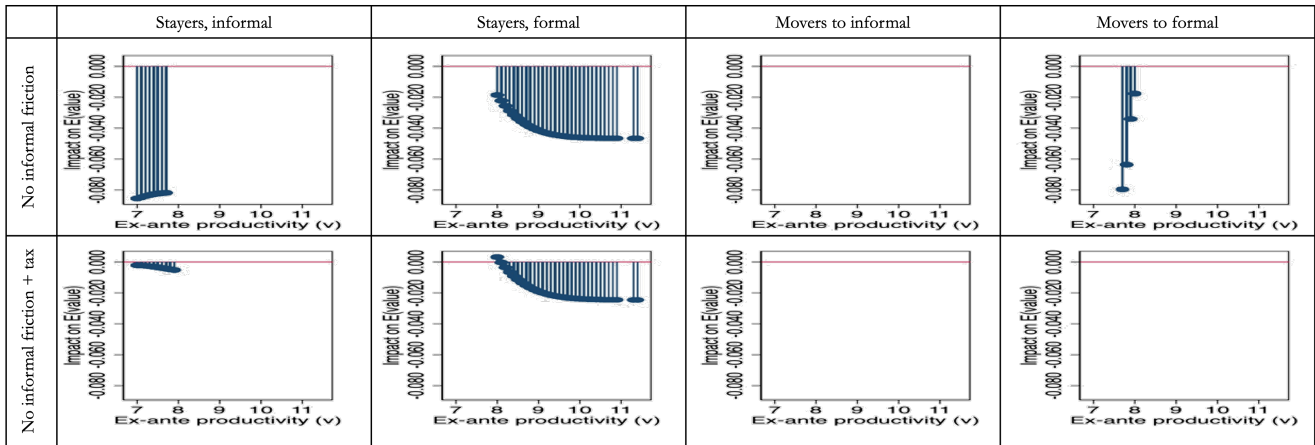
In this simulation, the parameter C of the logistic function is increased by 0.5 pp. In Figure 5 the cost of using capital for formal firms (light line) is moved inward, which means increasing the amount of formal firms for which the preferential rate is available. According to Figure 9, lowering the cost of accessing capital for micro-business is positive in the short run for formal firms, at a point where some informal firms become formal (column 4, row 1). However, if the policy is tax-financed, the impact disappears (column 4, row 2) and the impact on formal firms is lower. In the longer term (Figure 10), since there is an incentive to use more capital, interest rates increase, and large formal firms are negatively affected (column 2) particularly if the policy is financed by taxes. In contrast, when policies are financed but taxes alleviate the informal sector, because of competition.

**Figure 9. Micro effects of reducing the microbusiness distortion in accessing to capital (t=0)**



Source: Simulated data based on EEG (2019)

**Figure 10. Micro effects of reducing the microbusiness distortion in accessing to capital (t=1)**



Source: Simulated data based on EEG (2019)

Regarding macro effects, Table 7 shows how the economy adjusts by increasing risk-free interest rates (from 0.8% to 0.9%), but the adjustment is lower if taxes are increased. GDP and the informality of business and labor decrease (from 87% to 76%), but if it is financed through taxes, the reduction is compensated, because in overall formal firms get hurt.

**Table 7. Macro effects of reducing the microbusiness distortion in accessing to capital (t=1)**

	<b>Baseline</b>	<b>Micro friction</b>	<b>Micro friction + increase in profit tax</b>
<b>Informal firms</b>	87.5%	76.8%	86.2%
<b>Informal GDP</b>	20.2%	16.5%	19.9%
<b>Mass of firms (miilons)</b>	1.38	1.43	1.39
<b>Informal workers</b>	29.0%	21.30%	27.83%
<b>Skilled informal workers</b>	13.3%	9.98%	12.68%
<b>Unskilled informal workers</b>	31.8%	23.31%	30.51%
<b>Skilled wage</b>	1,302	1,401	1,318
<b>Unskilled wage</b>	774	839	780
<b>Interest rate</b>	0.8%	0.9%	0.8%
<b>Capital cost formal firms</b>	1.3%	1.2%	1.1%
<b>Capital cost informal firms</b>	1.8%	1.8%	1.8%
<b>Capital share of informal firms</b>	0.04	0.03	0.04
<b>Capital share of formal microfirms</b>	0.01	0.01	0.01
<b>ln (K/L) formal</b>	10.87	10.76	10.84
<b>ln (K/L) informal</b>	8.29	8.32	8.31
<b>TFP (%baseline)</b>	9.92	9.89	9.91
<b>Olye &amp; Parks</b>	1.77	1.91	1.80
<b>Output (mm)</b>	165	161	164
<b>Tax revenues</b>	38%	44%	39%
<b>Welfare</b>	1348	1220	1198

Source: Simulated data based on EEG (2019).

## **F Conclusions and policy implications**

Colombia is a micro-business country. One of the possible explanations for this fact is the tax threshold below which firms are not obliged to pay taxes. Micro-businesses, and most likely unskilled, have incentives to hire workers informal, whereas larger firms have incentive to hire informal because they have a tax deduction plus high enforcement. On top of that, micro-businesses also have low incentives to get capital, because they face a friction due to size and another due to informality. Therefore, the demography of Colombian companies is polluted by small firms with few incentives to hire capital and skilled workers and therefore unproductive.

This paper analyzed the impact of policies oriented to reduce the distortion in capital accumulation caused by size - smaller firms tend to face higher cost of using capital - and by informality - informal firms tend to face higher interest rates. The methodology used to estimate this impact is the estimation of a General Equilibrium model for the case of Colombia, that considers formal/informal and small/large firms, as well as a tax structure that is typical of developing countries, and the accumulation of capital, that obeys a cost of using capital that is assumed to follow a logistic function behavior.

Programs oriented to reduce the informality distortion on capital increase informality, at a point that informal small firms disappear. Productivity and GDP also decrease because informal firms are less productive. Actions directed at reducing the size distortion increase formality without changing much output and productivity, but as they need to be financed, some groups might end up affected. Online banking represents a huge opportunity to improve in this field, since it can reduce the size and informal distortions in accessing credit and increase access to international finance without additional cost. It can also speed up credit access, which is one of the features that matter the most for small business.

## References

- Acosta-Henao, M. (2020). *Law Enforcement and the Size of the Informal Sector*. Retrieved from <https://ssrn.com/abstract=3588010> (Draft Paper)
- Alaimo, V., Bosch, M., Kaplan, D. S., Pagés, C., & Ripani, L. (2015). *Empleos para Crecer*. Washington, D.C.: Inter-American Development Bank.
- Almeida, R., & Carneiro, P. (2012). Enforcement of Labor Regulation and Informality. *American Economic Journal: Applied Economics*, 4(3), 64–89.
- Alvarez, J., & Ruane, M. C. (2019). *Informality and Aggregate Productivity: The Case of Mexico*. International Monetary Fund.
- Amarante, V., & Perazzo, I. (2013). Trabajo por Cuenta Propia y Monotributo en el Uruguay. *Revista internacional del trabajo*, 132(3-4), 623–641.
- Angulo, R. (2023). *Índice Multidimensional de Robustez de Micronegocios (IMICRO). Diseño y Resultados para una Estrategia de Focalización de la Secretaría Distrital de Desarrollo Económico de Bogotá*. (Tech. Rep.). Bogotá, Colombia: Inclusión SAS.
- Antón, A., & Rasteletti, A. (2018). *Imposición al Trabajo en Contextos de Alta Informalidad Laboral: Un Marco Teórico para la Simulación de Reformas Tributarias y de Seguridad Social* (Documento para Discusión No. 646). Washington, D.C.: Inter-American Development Bank.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5–32.
- Chacaltana, J. (2016). Peru, 2002-2012: Growth, Structural Change and Formalization. *Revista CEPAL*.
- Chisari, O. O., Romero, C. A., & Martínez, J. P. V. (2016). Economy-Wide Effects of Improving Small Firms Access to Capital Markets: An Applied General Equilibrium Assessment. *Revista de Análisis Económico–Economic Analysis Review*, 31(1), 15–36.
- Chowdhury, M. J. A. (2009). *Microcredit, Micro-enterprises, and Self-Employment of Women: Experience from the Grameen Bank in Bangladesh* (Vol. 209; Working Paper). Dhaka, Bangladesh: Center for Microfinance and Development.
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. *Ensemble machine learning: Methods and applications*, 157–175.
- DANE. (2013). *General Methodology: Large Integrated Household Survey – GEIH*.
- DANE. (2020a). *General Methodology: Annual Services Survey – EAS*.
- DANE. (2020b). *General Methodology: Annual Trade Survey – EAC*.
- DANE. (2020c). *General Methodology: Survey Micro-establishments – EMICRON*.
- DANE. (2021). *Economic Census of Colombia: Economic Census Methodological Document*.
- Departamento Nacional de Planeación. (2019). *CONPES 3956 - Formalización Empresarial*. Retrieved from <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3956.pdf>
- De Paula, A., & Scheinkman, J. A. (2011). The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil. *Review of Income and Wealth*, 57, S8–S26.
- D’Erasmo, P. N., & Boedo, H. J. M. (2012). Financial Structure, Informality and Development. *Journal of Monetary Economics*, 59(3), 286–302.
- De Soto, H. (1989). *The Other Path: The Economic Answer to Terrorism*. New York, NY: Harper & Row.
- De Soto, H. (2000). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. Basic Books.
- Dix-Carneiro, R., Goldberg, P. K., Meghir, C., & Ulyssea, G. (2021). *Trade and Informality in the Presence of Labor Market*

- Frictions and Regulations* (Working Paper No. 28391). Cambridge, MA: National Bureau of Economic Research.
- Donovan, K., Lu, W. J., & Schoellman, T. (2023). Labor Market Dynamics and Development. *The Quarterly Journal of Economics*.
- Erosa, A., Fuster, L., & Martinez, T. R. (2023). Public Financing with Financial Frictions and Underground Economy. *Journal of Monetary Economics*, 135, 20–36.
- Eslava, M., Haltiwanger, J. C., & Pinzón, A. (2019). *Job Creation in Colombia vs the U.S.: “Up or Out Dynamics” Meets “The Life Cycle of Plants”* (Working Paper No. 25550). Cambridge, MA: National Bureau of Economic Research.
- Eslava, M., Meléndez, M., & Urdaneta, N. (2021). *Market Concentration, Market Fragmentation, and Inequality in Latin America* (Working Paper No. 11). NY, NY: United Nations Development Programme (UNDP).
- Estrada, D. A., & Hernández Rubio, A. (2019). *Situación Actual e Impacto del Microcrédito en Colombia*. Banco de la República de Colombia. Retrieved from <https://repositorio.banrep.gov.co/handle/20.500.12134/9723>
- Fairris, D., & Jonasson, E. (2016, February). *Determinants of Changing Informal Employment in Brazil, 2000–2010* (MPRA Paper No. 71475). Munich, Germany: University Library of Munich, Germany.
- Fajnzylber, P., Maloney, W. F., & Montes-Rojas, G. V. (2011). Does formality improve micro-firm performance? evidence from the brazilian simples program. *Journal of Development Economics*, 94(2), 262–276.
- Fernández, C. (2022). *Firms, Informality & Institutions. The Case of Colombia* (Documentos de Trabajo No. 020598). Bogotá, Colombia: Universidad del Rosario.
- Fernández, C., Villar, L., & Gómez, N. (2017). Taxonomía de la Informalidad en América Latina. *Coyuntura Económica*, 47(1 y 2), 137-167.
- Fernández, C. (2020). Informalidad Empresarial en Colombia. *Coyuntura Económica*, 50(1-2), 133-168.
- Fernández, C. (2023a). The Impact of Tax Asymmetries on Labor and Business informality. *Forthcomming*.
- Fernández, C. (2023b). *Informality, Productivity & Inequality: An Analysis from the Point of View of Firms and Workers*. (PNUD)
- Fernández, C. (2023c). Microbusiness and the Cost of Using Capital. *Forthcomming*.
- Fernández, C. (2023d). MSMEs and informality: A new employer-employee database for Colombia. *Forthcomming*.
- Fernández, C., & Villar, L. (2016). *The Impact of Lowering the Payroll Tax on Informality in Colombia* (Working Paper No. 015212). Bogotá, Colombia: Fedesarrollo.
- Granda, C., & Hamann, F. (2015). *Informality, Saving and Wealth Inequality in Colombia* (Working Paper No. 575). Washington, D.C.: Inter-American Development Bank.
- ’Haanwinckel, D., & Soares, R. R. (2021). ‘workforce composition, productivity, and labour regulations in a compensating differentials theory of informality’. *The Review of Economic Studies*, 88(6), 2970-3010.
- Hamann-Salcedo, F. A., & Mejía, L. F. (2013). *Formalizando la Informalidad Empresarial en Colombia* (Vol. Capítulo 10. Pág.:399-427). Banco de la República de Colombia.
- Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, 60(1), 126.
- Jaramillo, F. (2020). *Informalidad – Margen Intensivo*. Mimeo, note=Universidad del Rosario.
- Jaramillo, M. (2013). *Employment Growth and Segmentation in Peru, 2001-2011: Country Case Study on Labour Market Segmentation* (Employment Working Paper No. 151). Geneva, Switzerland: International Labour Organization.
- Jenkins, S. P., & Kerm, P. V. (2007). *PARETOFIT: Stata module to fit a Type 1 Pareto distribution*. Statistical Software Components, Boston College Department of Economics.
- Kugler, A. D., Kugler, M. D., & Herrera-Prada, L. O. (2017). Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia’s Reform. *Economía*, 18(1), 3–40.
- Levy, S. (2018). *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. Inter-American Development Bank.
- Levy, S., & Maldonado, D. (2021). *Misión del Empleo. Reporte Resumen*.
- Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School*, 22(2), 139-191.
- Loayza, N. V. (2016). Informality in the Process of Development and Growth. *The World Economy*, 39(12), 1856–1916.
- Loayza, N. V., Ulyssea, G., & Utsumi, T. (2018). Informality and the Labor Market Effects of Mass Migration: Theory and Evidence from Syrian Refugees in Turkey. In *7th turkey labor market network meeting* (Vol. 9, p. 2020).
- Lora, E., Mejía, L. F., Benítez, M., Delgado, M. E., & Gutiérrez, D. (2021, April). *Reformas para una Colombia Post-COVID-19. Hacia un Nuevo Contrato Social* (Informes de Investigación No. 019238). Bogotá, Colombia: Fedesarrollo.
- Maloney, W. F. (2004). Informality Revisited. *World Development*, 32(7), 1159–1178.
- Medina, C., & Posso, C. (2010). *Technical Change and Polarization of the Labor Market: Evidence for Brazil, Colombia and Mexico* (Borradores de Economía No. 614). Bogotá, Colombia: Banco de la República.
- Mejía, D., & Posada, C. E. (2007). *Informalidad: Teoría e Implicaciones de Política* (Borradores de Economía No. 455). Bogotá, Colombia: Banco de la República.

- Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6), 1695–1725.
- OECD. (2007). *Eurostat-OECD Manual on Business Demography Statistics*.
- OECD. (2020). *Financing SMEs and Entrepreneurs 2020: An OECD Scoreboard*. OECD Publishing.
- OECD. (2021). *The Missing Entrepreneurs 2021*.
- OECD. (2022). *Financing SMEs and Entrepreneurs 2022: An OECD Scoreboard*. OECD Publishing.
- Paturot, D., Mellbye, K., & Brys, B. (2013). *Average Personal Income Tax Rate and Tax Wedge Progression in OECD Countries* (OECD Taxation Working Papers No. 15). Paris, France: OECD Publishing.
- Perry, G. E., Maloney, W. F., Arias, O. S., Fajnzylber, P., Mason, A. D., & Saavedra-Chanduvi, J. (2007). *Informality : Exit and Exclusion* (No. 6730). The World Bank Group.
- Rebolledo Abanto, P. A., & Soto Chavéz, R. (2004). Estructura del Mercado de Créditos y Tasas de Interés: Una Aproximación al Segmento de las Microfinanzas. *Revista Estudios Económicos*(11).
- Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8), 2015–2047.
- Ulyssea, G. (2019). *Formal and Informal Business Dynamics*.
- Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics*, 12, 525–546.
- Veyrune, M. R. M., Della Valle, G., & Guo, S. (2018). *Relationship Between Short-Term Interest Rates and Excess Reserves: A Logistic Approach*. International Monetary Fund.
- Vijberger, W. P. (1992). Measuring income from family enterprises with household surveys. *Small Business Economics*, 4, 287-305.