CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.

THE NEXUS BETWEEN CORRUPTION, INFORMALITY AND MISALLOCATION IN MEXICO

TESINA

QUE PARA OBTENER EL GRADO DE

MAESTRO EN ECONOMÍA

PRESENTA

EDGAR ITAMAR AVALOS ALMANZA

DIRECTOR DE LA TESINA: DR. DAVID STRAUSS

CIUDAD DE MÉXICO JUNIO, 2018

Acknowledgments

I would like to thank all people who made this work possible.
To my sister, my mother, my father and the rest of my family for their continuous love and support.
To my friends and classmates with whom I spent the last two years working together for this degree.
To my advisor for his guidance and my thesis reader and seminar teacher for their valuable comments.
Finally, special thanks to Joaquin Mayorga and all staff from Microdatos INEGI for their help.

All the data from the Mexican Economic Censuses and the National Survey of Firms' Victimization (ENVE) in this thesis has been reviewed by INEGI to ensure no confidential information is disclosed.

The data from the National Survey of Victimization and Perception about Public Security (ENVIPE 2013) is available at `http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/envipe/2013/default.html`.

Finally, the data from the National Survey of Government Quality and Impact is available (ENCIG 2013) at `http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/encig/2013/default.html`.

Abstract

I investigate the relationship between the level of corruption in the federal entities of Mexico and the size of the informal sector within the states. I use a modified model à la Hsieh-Klenow with endogenous decision regarding the sector, formal or informal, in which firms operate and firm level data from the Mexican economic censuses.

Keywords: misallocation, informality, corruption, Mexico.
JEL classification: O17, O40, O47.

Content

List of tables

Chapter 1

Introduction

Informality, whether good or bad, has been pointed out as an important source of misallocation of economic resource among firms, which in turn is one of the main causes of total factor productivity (TFP) differences among countries. In this regard, Mexico has an arguably big informal sector; according to a preliminary report published by the National Institute of Statistics and Geography (INEGI) in December of 2017 they estimate that, during the year 2016, 22.6 percent of the total gross domestic product (GDP) was produced by informal economic units, with 56.7 percent of the labor force engaged in informality conditions. In at attempt of diminishing the above numbers, the federal government launched in 2014 a program called "Crezcamos Juntos" which aimed at bringing firms into formality by offering them discounts in social security benefits, monetary support for acquiring equipment and tools and credits with preferential rates, among others. Yet, one could argue that the program has been unsuccessful since, according to the INEGI, the strength of the informal sector has seen little change since then. So I ask: could it be that, despite the federal government's efforts, are there local factors or agents that ease the survival and growth of the informal sector?

In Mexico there is a saying "el que no transa no avanza" meaning that, if you do not cheat, you can not progress. Imagine the following: you are an entrepreneur wishing to open a business. You know that, according to the law, you need to register your new firm, pay government taxes

and your workers' social benefits; yet you also know that the probability of getting caught and face punishment if you opt to hide your firm and revenue from the government is small and that, even if you get visited by an inspector, there is a pretty big chance that you can bribe him and remain undetected. Still, it is not clear the direction that such a bribe would take in presence of a higher degree of corruption: would a more corrupt environment encourage Government employees to demand lower bribes in order to hide an informal firm since there could be many other corrupt agents who can offer a lower *price*? Or the other way around, would the corrupt agents collude in order to extort greater bribes from the firms?

So, in order to assess the impact of the distortions within the Mexican economy I will use the framework developed by Hsieh and Klenow (2009) with a twist: now, firms will have to choose if they want to operate in the formal or informal sector. If they choose the former, besides the distortions to output and capital present in the original work, firms will also face State taxes that will act as an additional distortion over output. On the other hand, firms that decide to produce in informality will not be subject to said taxes but instead will face an additional constraint over capital (in the form of a "tax"). It is usual within the literature to associate informality costs (such as the tax I just mentioned) yet I will take another approach. Motivated by the example form the previous paragraph, I will relate the cost of operating in the informal sector with the corruption level of the federal entity in which the firm lies. As so, I will link the size of the informal sector with the corruption level of the state. Finally, I use data from the Mexican economic census to calibrate the model and measure misallocation along three dimensions: general distortions in the Mexican economy, inefficiency of the informal sector and the degree of corruption of each state.

The structure of this work is as follows. In the remaining of this chapter I present some of the work that has been done prior and over which this thesis is constructed. In chapter 2 I give a summary of the structure of establishments for each year of the Mexican economic census. In chapter 3 I present the model with the corresponding modifications. In chapter 4 I explain how data is manipulated and used. Finally, in chapter 5 I present my findings.

1.1 Literature Review

There is no short literature on the role of misallocation of resources. Most famously, Hsieh and Klenow (2009) introduce a method for measuring the degree of misallocation among firms, as well as estimating the potential gains derived from a more efficient allocation. They calibrate their model using data from the manufacturing sector in India, China and the United States; employing the later as a "distortions free" comparison point, they find substantial potential gains: about 30 to 50 percent in China and 40 to 60 percent in India. Despite some simplifications in the model and possible measurement errors, this framework has proven a valuable starting point at measuring misallocation.

Again, Hsieh and Klenow (2014) describe how plants' life cycles dynamics could lower the aggregate manufacturing productivity of India and Mexico by 25 percent relative to the Untied States. Restuccia and Rogerson (2013) document how substantial differences in establishment size and productivity across countries arise in the presence of correlated distortions; they also find a strong positive association between average establishment size and GDP per capita. Bento and Restuccia (2016) focus on the reallocation of factors as a potential as an important measure of total factor productivity (TFP) differences among countries. However, the question remains: what is it that causes such differences and misallocation?

The above is not an easy question to fully answer, however several authors have pointed in a direction which has an important role: the informal sector. For example, Farrell (2004) reports how informal firms may choose to operate in a smaller inefficient scale in order to avoid detection by the government.

More specifically for Mexico, Leal-Ordoñez (2013) employs data from the country to assess the quantitative effect of incomplete tax enforcement on aggregate output and productivity. In his model, enforcement is represented as a capital level threshold that firms cannot trespass if they want to remain informal, otherwise they would be caught by the Government. He finds that, under perfect enforcement (all firms are formal), labor productivity and output would be 19 percent higher under perfect competition and 34 percent higher under monopolistic competition.

Also for Mexico, Busso, Fazio, and Levy (2012) make an important distinction between what they call illegal and informal firms. They measure legality as the ratio of total social security taxes to wages of salaried workers. On the other hand, informality is measured as the ratio of social security taxes to the remunerations of all workers. Putting it simple, a firm that employs one salaried worker to whom they pay social security taxes, but that also employs fifty other workers registered as non-salaried and thus exempt of social security benefits would be considered informal but legal. In fact, using data from the Mexican economic census, they claim that two thirds of all firms captured by the census are informal but legal, around 20 percent are illegal and informal, and only 3 percent are legal and formal. Then, employing the framework described in Hsieh and Klenow (2009), they derive the productivity losses due to misallocation of resources in Mexico and find that one peso would be worth 28 percent more if it were allocated in a formal and legal firm rather that in illegal and informal one, and 50 percent more that in a legal and informal firm.

Hence, it is clear that the informal sector plays a fundamental role in resources misallocation in Mexico. Yet, the aforementioned authors fail to answer the question posed in this thesis: what enables the option and motivates firms to operate in the informal sector, hence resulting in a misallocation of resources?

As economists we work under the fundamental concept that all agents are motivated through incentives. In that sense, we assume that firms compare the benefits of maneuvering in the informal sector (mainly being *exempt* of paying taxes) against the downsides of it (such as the legality of the operation, credit constraints, among others) before making a choice. Dabla-Norris, Gradstein, and Inchauste (2008) employ a general equilibrium model and find that the quality of the legal framework in a country is the most important factor in determining the size of the informal sector, even more so than taxes, regulations and financial constraints. D'Erasmo and Moscoso (2012) find that countries with a low degree of of debt enforcement and high costs of formality are characterized by a large share of output produced by low productivity firms in the informal sector. Goyette and Kouamé (2017), employing firm-level data of 38 African and Latin

American countries, find a positive correlation between distortions (policy ineffectiveness) and tax evasion.

The papers mentioned before hint, unsurprisingly, to the following: the easier it is for a firm to *cheat*, the more likely it will do so. Now we ask ourselves: what is it that diminishes the Government's ability to enforce the law? Is it that monitoring is difficult and costly, or perhaps there are more factors at play? As mentioned in Acemoglu and Verdier (2000), involving of the government may give rise to a problem of excessive bureaucracy and corruption.

As in Mauro (1995), corruption has been shown to negatively affect investment as well as economic growth. Kurer (1993) mentions how *clients* can elect and support a political *patron* in return for economic benefits which can in turn lead to misallocation. Aidt (2003) reviews developments of corruption in the literature an already talks bout the possibility of a tax-collector agent whom can choose to take a bribe from a firm in order to avoid charging taxes.

Amirapu and Gechter (2014) investigate the costs associated with labor regulations in India that apply to firms with over ten workers. They estimate that, on average, regulations increase labor costs by 35 percent and, perhaps more strikingly, also detect considerable variation in this cost across states and are able to find a robust positive association between costs and the level of corruption in the states. Here, they employ two measures of corruption: a subjective perception measure reported by Transparency International, and the percentage of electricity that is lost in transmission and distribution, reported by the Reserve Bank of India. They justify said association and the direction it takes by allowing inspectors to extort firms threatening them with greater fines, consequence of India's labor regulations being unclear or outdated.

Finally, Bardhan (2006) mentions how corrupt activity may depend on how much corrupt activity is taking place around.

With the above in mind, I plan to investigate whether the individual *corruption level* of the states in Mexico has an influence over the easiness of operating a business under the informal sector. As stated in Acemoglu and Verdier (1998), preventing corruption is costly so, if we can not immediately aspire to a corruption *free* panorama, maybe the targeting of specific states can

help us breach the gap with developed countries.

Chapter 2

Summary Statistics

This thesis employs data from the Mexican economic censuses provided by the National Institute of Statistics and Geography (INEGI). Starting in 1993, this surveys are conducted every five years and encompass all urban, private and non-agricultural establishments with fixed locations plus a random sample of rural units of the same kind.

Following Busso et al. (2012) I distinguish firms by degree of formality depending on their ratio of social contributions to total remunerations; thus, I separate firms in three main groups depending on their ratio of social contributions to total remunerations. Formal firms are those for which their ratio is greater than 0.18, semi formal firms have a ratio between zero and 0.18 and informal firms have a ratio of zero (i.e., they do not pay social contributions at all). Additionally, I create a fourth category called *outsourcing* for firms whose payments to employees were all done through a third firm.

Next, I present some statistics regarding the distribution of characteristics between the three types of firms. First, the distribution of establishments by formality status among all years of the census is presented in Table 2.1. Notice how, each year, over 70 percent of all establishments surveyed were informal. Also, although the distribution of formal, semi formal and informal establishments has maintained relatively unchanged across all years, there was a big increase in outsourcing firms starting in 2008, mainly gaining terrain over informal firms.

Table 2.1: Percentage of Establishments by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*	**Total**
1993	6.16	7.33	86.51	0.00	1,240,028
1998	8.75	9.55	80.96	0.73	2,475,677
2003	6.09	8.07	84.20	1.63	2,658,440
2008	4.49	8.92	76.56	10.03	2,591,128
2013	5.45	7.04	73.88	13.62	3,467,106

Source: Author's calculations based on data from INEGI.
Note: Outsourcing in 1993 is zero because of the way this category was constructed.

Now, by looking at table 2.2 one can see that, each year, the greatest share of workers were employed in the informal sector. However, starting in 2008, there was a big chunk of workers that switched to outsourcing firms, particularly switching away from the formal sector.

Table 2.2: Share of Labor Usage by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	28.06	19.63	52.30	
1998	34.77	26.95	37.99	0.28
2003	30.68	23.12	40.89	5.28
2008	21.20	25.71	40.89	12.18
2013	23.67	22.65	37.33	16.33

Source: Author's calculations based on data from INEGI.

To give some more contrast I combine the information of tables 2.1 and 2.2 to present the average number of workers used each year by formality status in table 2.3. I see that, every year, semi formal establishments employ, on average, approximately six times more workers than informal establishments and formal firms nearly ten times as many as informal ones. On the other hand, despite the fact that outsourcing firms employed less average workers than informal establishments, in 2003 outsourcing firms presented and extraordinary increase (more than 1000 percent) in their average number of workers, even surpassing that of semi formal firms; the following year, this huge increase falls by more than 50 percent but still outsourcing firms start to average more than twice as many workers in comparison to informal firms.

In table 2.4 I present the shares of capital usage employed per formality status. In contrast

Table 2.3: Average Number of Workers by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	11.95	7.02	1.58	
1998	14.73	10.46	1.74	1.42
2003	21.01	11.94	2.03	13.48
2008	22.30	13.61	2.52	5.74
2013	18.04	13.37	2.10	4.98

Source: Author's calculations based on data from INEGI.

to the previous table, here the greatest share of capital is used in the formal sector. This and the previous observations are consistent with the intuition that the informal sector is more labor intensive relative to the formal one. Also, similarly to the previous table, there was an important change for the outsourcing sector in terms of capital usage where, in 2013, this kind of firms employed nearly one quarter of all capital of said year.

Table 2.4: Share of Capital Usage by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	57.16	21.61	21.22	
1998	58.12	27.22	14.29	0.34
2003	43.20	26.95	15.29	14.44
2008	43.89	23.64	12.56	19.89
2013	41.48	22.87	10.86	24.77

Source: Author's calculations based on data from INEGI.

As with average number of workers, I present the average capital employed by formality status in each year in table 2.5. Notice how, from 1993 to 2003, formal establishments used around 38 times more capital than informal ones and, starting in 2008, this relation changed to over 50 times more capital usage. For outsourcing establishments, as in average number of workers, there was a huge increase in 2003 on average capital usage by this kind of firms where they even employed, on average, more capital than formal firms and then again, starting in 2008, this number again diminished by over 50 percent to situate itself below that of semi formal establishments.

Lastly, in tables 2.6 and 2.7, I present the shares of value added contributed by each formality

Table 2.5: Average Capital Used by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	736.07	233.86	19.45	
1998	2062.59	885.54	54.87	146.95
2003	3690.22	1736.69	95.09	4596.82
2008	8117.48	2201.63	136.28	1646.92
2013	6565.65	2804.94	126.88	1569.47

Source: Author's calculations based on data from INEGI.
Note: Values are in thousands of Mexican pesos (MXN).

sector and the average value added by formality status, respectively. Very similar to table 2.4, the formal sector contributes the greatest share of value added while the informal sector contributes, in general, a small share. Again, there was an important change for outsourcing firms in 2003 and, since 2008, this sector has contributed a greatest share than that of the informal sector.

Table 2.6: Share of Value Added by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	46.50	28.37	25.11	
1998	49.12	35.46	15.13	0.27
2003	44.96	30.14	17.27	7.61
2008	35.67	33.24	15.20	15.87
2013	40.53	27.03	14.05	18.37

Source: Author's calculations based on data from INEGI.

Table 2.7: Average Value Added by Formality Status

Year	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	382.84	196.22	14.71	
1998	1345.47	890.31	44.82	88.81
2003	3050.95	1542.32	84.76	1924.44
2008	4601.28	2159.25	115.02	916.53
2013	4850.45	2506.57	124.16	880.16

Source: Author's calculations based on data from INEGI.
Note: Values are in thousands of Mexican pesos (MXN).

2.1 Distribution of age

In table 2.8 I present the distribution of age (measured in years) among establishments for each year of the census starting in 1998. Notice that, across all years, the informal category has the most share of younger firms when compared with the other categories. Also, in most years, the formal category has the most share of older firms; the only exception to this occurs in 2003 where the share of older firms in the outsourcing category surpasses that of the formal category, but even then the difference in share is small. This results are expected as, intuitively, one would expect the newly formed or younger firms to be small and thus elect to operate in the informal sector since they would not be constraint by the capital limitations that this entails. On the other hand, as a firm grows older it is likely that it wishes to also grow in size and hence decides to operate in the formal sector in order to gain better access to the capital market.

Notice as well that, in 1998, the distribution of age of outsourcing firms looked almost identical to that of the informal firms but it swifted over time to approach more that of semi formal establishments and, as of 2013, this two distributions look almost the same.

2.2 Distribution of size

Finally, by looking at the distribution of size (measured in number of workers employed) in table 2.9 it is seen that, each year, over 90 percent of all informal establishments only hired less than five workers each; on the other hand, the formal category has the greatest share of larger establishments. This is consistent with the recurrent idea present in the literature that informal establishments need to remain small in order to avoid detection by the government.

Table 2.8: Age Category Distribution Along Formality Status

Year	**Age category**	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1998	0 to 4	44.25	47.90	61.06	60.74
	5 to 9	27.50	25.78	20.12	21.70
	10 to 14	11.37	10.72	7.49	7.42
	15 to 19	6.45	6.08	4.55	4.30
	20 to 24	3.38	3.09	2.27	2.15
	25 or more	7.05	6.43	4.50	3.69
2003	0 to 4	31.80	36.53	55.89	41.59
	5 to 9	23.27	23.78	18.36	21.24
	10 to 14	17.64	16.19	10.50	13.52
	15 to 19	9.32	8.46	5.61	6.45
	20 to 24	6.15	5.21	3.42	3.91
	25 or more	11.82	9.84	6.22	13.30
2008	0 to 4	30.49	34.91	55.38	38.61
	5 to 9	21.35	22.43	18.92	21.48
	10 to 14	15.16	14.53	9.06	12.76
	15 to 19	12.47	11.33	6.94	10.02
	20 to 24	6.76	5.84	3.30	4.80
	25 or more	13.78	10.94	6.39	12.34
2013	0 to 4	28.00	31.83	46.77	30.97
	5 to 9	22.87	24.31	24.36	24.32
	10 to 14	15.03	14.95	11.32	14.73
	15 to 19	11.38	9.97	6.06	9.60
	20 to 24	8.58	7.78	4.60	7.42
	25 or more	14.13	11.15	6.88	12.97

Source: Author's calculations based on data from INEGI.
Note: Due to data access complications, at the time of writing I do not posses age information for the 1993 census.

Table 2.9: Size Category Distribution Along Formality Status

Year	**Size category**	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
1993	1 to 4	59.14	64.99	98.70	
	5 to 10	20.14	20.93	1.13	
	11 to 19	9.48	7.82	0.11	
	20 to 49	5.84	4.10	0.04	
	50 to 99	2.34	1.12	0.01	
	100 or more	3.06	1.02	0.00	
1998	1 to 4	55.88	54.20	97.07	98.12
	5 to 10	21.46	24.12	2.36	1.53
	11 to 19	10.52	11.53	0.41	0.24
	20 to 49	6.57	6.59	0.13	0.07
	50 to 99	2.44	1.88	0.02	0.04
	100 or more	3.12	1.67	0.01	0.01
2003	1 to 4	44.31	48.26	94.84	60.13
	5 to 10	24.17	27.15	4.11	22.14
	11 to 19	13.29	13.21	0.75	6.80
	20 to 49	9.38	7.26	0.24	5.67
	50 to 99	3.84	2.15	0.04	2.31
	100 or more	5.02	1.98	0.01	2.95
2008	1 to 4	39.16	42.55	90.84	86.17
	5 to 10	27.04	30.09	7.24	7.96
	11 to 19	15.06	14.62	1.40	2.21
	20 to 49	9.98	8.00	0.43	1.64
	50 to 99	3.67	2.46	0.07	0.92
	100 or more	5.09	2.28	0.02	1.10
2013	1 to 4	43.79	49.16	94.35	88.36
	5 to 10	25.26	25.94	4.30	6.44
	11 to 19	14.89	13.08	0.95	2.11
	20 to 49	9.40	7.36	0.33	1.49
	50 to 99	2.96	2.15	0.05	0.68
	100 or more	3.70	2.30	0.03	0.92

Source: Author's calculations based on data from INEGI.

2.3 Corruption

I use three different sources in measuring corruption, all of which come from different surveys conducted by INEGI. The first one is the National Survey of Victimization and Perception about Public Security (ENVIPE). The second one is the National Survey of Government Quality and Impact (ENCIG). The third one is the National Survey of Firms' Victimization (ENVE). The first two ones are conducted over households whereas the latter one is conducted over establishments. I detail the procedure and data used in obtaining the rankings in chapter 4.

Chapter 3

The Model

3.1 Basic Setup

As in Hsieh and Klenow (2009), I will consider a final good Y produced by a representative firm under perfect competition. This representative firm aggregates the outputs Y_s of S intermediate industries, so s can take on the values $\{1, \ldots, S\}$, using a Cobb-Douglas production function

$$Y = \prod_{s=1}^{S} Y_s^{\theta}, \text{ where } \sum_{s=1}^{S} \theta_s = 1. \tag{3.1}$$

Cost minimization of the above production function implies that

$$P_s Y_s = \theta_s P Y, \tag{3.2}$$

where P_s denotes the price charged by industry s and P the price of the final good. Substituting the condition of cost minimization (3.2) into the final good production function (3.1) yields the final good pricing rule

$$P = \prod_{s=1}^{S} \left(\frac{P_s}{\theta_s} \right)^{\theta_s}.$$

On a lower level, the output Y_s of industry s is a C.E.S. aggregate of M_s differentiated

products Y_{si}

$$Y_s = \left(\sum_{i=1}^{M_s} Y_{si}^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma}{\sigma-1}},$$

where σ is greater than one. Again, industry output Y_s is produced under perfect competitions so, from the first order condition, I obtain the demand for each differentiated product

$$Y_{si} = Y_s \left(\frac{P_s}{P_{si}} \right)^{\sigma}. \tag{3.3}$$

Finally, each good Y_{si} within an industry is produced by a single firm under monopolistic competition employing a Cobb Douglas production function

$$Y_{si} = A_{si} K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}. \tag{3.4}$$

Notice how the capital intensity α_s is common to all firms in an industry, yet we are allowing firm heterogeneity by letting physical productivity A_{si} be firm specific. It is at this level that I will be giving firms an extra choice (apart from choosing the capital and labor levels, as well as price fixation): whether to be formal or informal. The differences between formality and informality, including which kind of firms choose each option, will be explained in the following section.

3.2 Formality or Informality

Usually the main disadvantage for the firms when choosing formality over informality is having to pay government taxes, yet being informal must carry disadvantages of its own, otherwise all firms will choose to be informal and that is not what is observed in the real world.

Through the literature there are two main ways in which informal firms are constrained: by facing a probability of detection (and punishment) by the government, or through capital constraints (often related with credit constraints). For simplicity and in an effort of not distance ourselves to much from the original model, I opt for the latter option. Thus, informal firms

will face a higher price in the capital market, this, along some other restrictions, will yield two results that are consistent with the literature and the data: First, informal firms will have a lower capital to labor ratio than formal firms. Second, the most talented firms will choose to be formal. Moreover, I will relate this distortion on the capital price to the level of corruption present in each state.

Now, I describe the mathematical problem faced by each firm.

3.2.1 Formal Firms

In the case of choosing formality, firms face a State tax T_F over total output produced, thus the firm benefits are given by

$$\pi_{si}^F = (1 - T_F)(1 - \tau_{Ysi})P_{si}Y_{si} - (1 + \tau_{Ksi})RK_{si} - wL_{si}$$

where Y_{si} is produced as described by the Cobb Douglas production function (3.4). Here, τ_{Ysi} denotes output distortions (such as transportation costs or subsidies) and τ_{Ksi} denotes capital distortions (for instance, even between establishments in the formal sector there might be differences in credits accessibility). Here, the output distortion τ_{Ysi} raises the marginal product of capital and labor by the same amount whereas the capital distortion τ_{Ksi} only raises the marginal product of labor relative labor. I solve the maximization problem in two stages. First, cost minimization

$$\min_{K_{si},L_{si}} (1 + \tau_{Ksi})RK_{si} + wL_{si},$$
$$\text{subject to } Y_{si} = A_{si}K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}.$$

Yielding the capital to labor ratio

$$\frac{K_{si}}{L_{si}} = \frac{\alpha_s}{1 - \alpha_s}\frac{w}{(1 + \tau_{Ksi})R}, \tag{3.5}$$

and the standard expression for the cost function

$$c_{si} = \lambda_s A_{si}^{-1}(1+\tau_{Ksi})^{\alpha_s},$$

where

$$\lambda_s \coloneqq \left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}.$$

Now the profit maximization problem,

$$\begin{aligned}&\max_{P_{si}} (1-T_F)(1-\tau_{Ysi})\pi_{si}^F Y_{si}^d - c_{si}Y_{si}^d,\\ &\text{subject to } Y_{si}^d = \left(\frac{P_s}{P_{si}}\right)^{\sigma} Y_s.\end{aligned}$$

Again, I obtain the usual expression for the price setting rule

$$P_{si} = \lambda_s \frac{\sigma}{\sigma-1} A_{si}^{-1} \frac{(1+\tau_{Ksi})^{\alpha_s}}{(1-\tau_{Ysi})(1-T_F)}. \tag{3.6}$$

3.2.2 Informal Firms

As previously mentioned, informal firms do not pay government taxes T_F but instead face an additional (positive) *tax* γ_{si} over capital in order to operate in informality. Then, firm benefits for informal firms are given by

$$\pi_{si}^I = (1-\tau_{Ysi})P_{si}Y_{si} - (1+\tau_{Ksi})(1+\gamma_{si})RK_{si} - wL_{si}.$$

As for formal firms, τ_{Ysi} denotes output distortions and τ_{Ksi} capital distortions. Later on, I will try to link the additional distortion γ_{si} with the corruption level of the state in which the firm resides. Again, I solve the problem in two stages. From the first order conditions of the cost

minimization problem

$$\min_{K_{si},L_{si}} (1+\tau_{Ksi})(1+\gamma_{si})RK_{si} + wL_{si},$$
$$\text{subject to } Y_{si} = A_{si}K_{si}^{\alpha_s}L_{si}^{1-\alpha_s},$$

yields the capital to labor ratio

$$\frac{K_{si}}{L_{si}} = \frac{\alpha_s}{1-\alpha_s}\frac{w}{(1+\tau_{Ksi})(1+\gamma_{si})R}, \tag{3.7}$$

and the cost function

$$c_{si} = \lambda_s A_{si}^{-1}(1+\tau_{Ksi})^{\alpha_s}(1+\gamma_{si})^{\alpha_s},$$

with λ_s as before.

Notice how the capital to labor ratio of informal firms in equation (3.7), is lower than that of formal firms, equation (3.5), as it is usual in the literature and observed in the data.

Solving then the profit maximization problem for informal firms

$$\max_{P_{si}} (1-\tau_{Ysi})\pi_{si}^I Y_{si}^d - c_{si}Y_{si}^d$$
$$\text{subject to } Y_{si}^d = \left(\frac{P_s}{P_{si}}\right)^{\sigma} Y_s,$$

yields the price setting rule for informal firms

$$P_{si} = \lambda_s \frac{\sigma}{\sigma-1} A_{si}^{-1} \frac{(1+\tau_{Ksi})^{\alpha_s}(1+\gamma_{si})^{\alpha_s}}{1-\tau_{Ysi}}. \tag{3.8}$$

3.2.3 Which Firms Choose What?

When deciding whether being formal or informal, firms will choose whatever option that yields them the greatest profit given their technology. First, from the production function (3.4) and the

demand for differentiated product Y_{si} in equation (3.3),

$$L_{si} = \left(\frac{P_s}{P_{si}}\right)^{\sigma} \frac{Y_s}{A_{si}} \left(\frac{K_{si}}{L_{si}}\right)^{-\alpha_s}. \tag{3.9}$$

Then, substituting the capital to labor ratio (3.5) and the price setting rule (3.6) for formal firms into equation (3.9) I obtain the optimal amount of labor usage for formal firms,

$$L_{si}^F = \mu_s \frac{A_{si}^{\sigma-1}(1-\tau_{Ysi})^{\sigma}(1-T_F)^{\sigma}}{(1+\tau_{Ksi})^{\alpha_s(\sigma-1)}},$$

where μ_s is defined as

$$\mu_s \coloneqq Y_s P_s^{\sigma} \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \left(\frac{1-\alpha_s}{w}\right)^{\sigma(1-\alpha_s)+\alpha_s} \left(\frac{\alpha_s}{R}\right)^{\alpha_s(\sigma-1)}.$$

With the above expression and the capital to labor ratio in equation (3.5) I can compute the benefits of firm A_{si} in the formal market,

$$\pi_{si}^F = \mu_s \frac{A_{si}^{\sigma}(1-\tau_{Ysi})^{\sigma}(1-T_F)^{\sigma}}{(1+\tau_{Ksi})^{\alpha_s(\sigma-1)}} \Bigg\{ (1-T_F)(1-\tau_{Ysi}) \cdot A_{si} \Big[\frac{\alpha_s}{1-\alpha_s}\frac{w}{(1+\tau_{Ksi})R}\Big]^{\alpha_s} - \frac{w}{1-\alpha_s} \Bigg\}.$$

Now, I do the same for informal firms. Substituting the capital to labor ratio (3.7) and the price setting rule(3.8) for informal establishments into equation (3.9) I obtain the optimal labor usage for informal firms,

$$L_{si}^I = \mu_s \frac{A_{si}^{\sigma-1}(1-\tau_{Ysi})^{\sigma}}{(1+\tau_{Ksi})^{\alpha_s(\sigma-1)}(1+\gamma_{si})^{\alpha_s(\sigma-1)}},$$

with μ_s defined as before. Then, I can compute the benefits for firm A_{si} when operating in the

informal market,

$$\pi_{si}^{I} = \mu_s \frac{A_{si}^{\sigma}(1-\tau_{Ysi})^{\sigma}}{(1+\tau_{Ksi})^{\alpha_s(\sigma-1)}(1+\gamma_{si})^{\alpha_s(\sigma-1)}} \Bigg\{ (1-\tau_{Ysi}) \cdot A_{si} \Big[\frac{\alpha_s}{1-\alpha_s} \frac{w}{(1+\tau_{Ksi})(1+\gamma_{si})R} \Big]^{\alpha_s} - \frac{w}{1-\alpha_s} \Bigg\}.$$

Thus, a firm A_{si} will choose formality over informality if its benefits in the formal market, π_{si}^{F}, are greater than in the informal market, π_{si}^{I}. Then, I obtain the following result.

Proposition 1 (Positive Sorting)**.** *Fixing the distortions level. If*

$$1 - T_F > \frac{1}{(1+\gamma_{si})^{\alpha_s\sigma}} \text{ and } (1+\gamma_{si})^{\alpha_s(\sigma-1)} > 1,$$

then the model presents positive sorting in the formality choice, this is, firms will decide to operate in the formal sector if and only if their physical productivity A_{si} is above a certain threshold.

Prueba. As previously mentioned, a firm will choose to be formal if and only if that yields it more benefit than operating in the informal sector. This is, firm A_{si} will be formal if and only if

$$\pi_{si}^{F} \geq \pi_{si}^{I}.$$

Replacing both benefits with the previously derived expressions the above condition becomes,

$$\begin{aligned}(1-T_F)(1-\tau_{Ysi})A_{si} \left[\frac{\alpha_s}{1-\alpha_s} \frac{w}{(1+\tau_{Ksi})R} \right]^{\alpha_s} - \frac{w}{1-\alpha_s} \\ \geq \frac{(1-\tau_{Ysi})A_{si}}{(1+\gamma_{si})^{\alpha_s(\sigma-1)}} \left[\frac{\alpha_s}{1-\alpha_s} \frac{w}{(1+\tau_{Ksi})(1+\gamma_{si})R} \right]^{\alpha_s} - \frac{w}{(1+\gamma_{si})^{\alpha_s(\sigma-1)}(1-\alpha_s)},\end{aligned}$$

if and only if

$$\begin{aligned}A_{si} \left[(1-T_F) - \frac{1}{(1+\gamma_{si})^{\alpha_s\sigma}} \right] \\ \geq \frac{w}{(1-\alpha_s)(1-\tau_{Ysi})} \left[1 - \frac{1}{(1+\gamma_{si})^{\alpha_s(\sigma-1)}} \right] \left[\frac{1-\alpha_s}{\alpha_s} \frac{(1+\tau_{Ksi})R}{w} \right]^{\alpha_s}.\end{aligned}$$

Now, since by assumption

$$1 - T_F > \frac{1}{(1+\gamma_{si})^{\alpha_s \sigma}} \tag{3.10}$$

or equivalently,

$$1 - T_F - \frac{1}{(1+\gamma_{si})^{\alpha_s \sigma}} > 0$$

then, a firm will choose to be formal if and only if

$$A_{si} \geq \frac{w}{(1-\alpha_s)(1-\tau_{Ysi})} \left[1 - \frac{1}{(1+\gamma_{si})^{\alpha_s(\sigma-1)}}\right] \left[\frac{1-\alpha_s}{\alpha_s} \frac{(1+\tau_{Ksi})R}{w}\right]^{\alpha_s} \cdot \left[1 - T_F - \frac{1}{(1+\gamma_{si})^{\alpha_s \sigma}}\right]^{-1}.$$

Additionally, since again by assumption

$$(1+\gamma_{si})^{\alpha_s(\sigma-1)} > 1, \tag{3.11}$$

the above condition is positive. □

The next question is, how restrictive are the assumptions of the above proposition? Since σ (the elasticity of substitution) is greater than one, α_s (industry s capital share) is greater tan zero and I'm assuming that γ_{si} is positive (otherwise firms will have zero incentives to choose formality over informality), condition (3.11) is always fulfilled. Condition (3.10) needs more examination. In chapter 4 I discuss how how σ and α_s are chosen, but I will briefly make use of their values here; σ is set to two whereas α_s has values ranging from around 0.1 to nearly one. Also, Leal-Ordoñez (2013) estimates a T_F of 0.25. This translates to that, on average, γ_{si} will need to be greater than 0.4 and, in the most extreme case, it will need to be greater than five in order to fulfill the condition.

3.3 Total Factor Productivity

3.3.1 Total Factor Productivity of Revenue

The total factor productivity of revenue, or TFPR, of a firm is defined as price times physical productivity A_{si}, this is, for formal firms

$$\begin{aligned}\text{TFPR}_{si}^{F} &\coloneqq P_{si}^{F} A_{si} \\ &= \left(\frac{\text{MRPK}_{si}^{F}}{\alpha_s}\right)^{\alpha_s} \left(\frac{\text{MRPL}_{si}^{F}}{1-\alpha_s}\right)^{1-\alpha_s}\end{aligned}$$

where,

$$\begin{aligned}\text{MRPK}_{si}^{F} &\coloneqq \alpha_s \frac{P_{si}^{F} Y_{si}^{F}}{K_{si}^{F}} \\ &= \frac{\sigma}{\sigma-1} R \frac{1+\tau_{Ksi}}{(1-\tau_{Ysi})(1-T_F)},\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}\text{MRPL}_{si}^{F} &\coloneqq (1-\alpha_s) \frac{P_{si}^{F} Y_{si}^{F}}{L_{si}^{F}} \\ &= \frac{\sigma}{\sigma-1} w \frac{1}{(1-\tau_{Ysi})(1-T_F)}.\end{aligned} \tag{3.13}$$

Similarly for the informal firms,

$$\begin{aligned}\text{TFPR}_{si}^{I} &\coloneqq P_{si}^{I} A_{si} \\ &= \left(\frac{\text{MRPK}_{si}^{I}}{\alpha_s}\right)^{\alpha_s} \left(\frac{\text{MRPL}_{si}^{I}}{1-\alpha_s}\right)^{1-\alpha_s}\end{aligned}$$

where now,

$$\begin{aligned}\mathrm{MRPK}_{si}^{I} &\coloneqq \alpha_s \frac{P_{si}^{I} Y_{si}^{I}}{K_{si}^{I}} \\ &= \frac{\sigma}{\sigma-1} R \frac{(1+\tau_{Ksi})(1+\gamma_{si})}{1-\tau_{Ysi}},\end{aligned} \tag{3.14}$$

and

$$\begin{aligned}\mathrm{MRPL}_{si}^{I} &\coloneqq (1-\alpha_s) \frac{P_{si}^{I} Y_{si}^{I}}{L_{si}^{I}} \\ &= \frac{\sigma}{\sigma-1} w \frac{1}{1-\tau_{Ysi}}.\end{aligned} \tag{3.15}$$

3.3.2 Total Factor Productivity of an Industry

I will now obtain the total factor productivity of each sector in an industry by aggregating the previously obtained expressions for firm level productivity. For each given industry we will aggregate separately the informal firms and the formal firms that compose it. Thus, aggregating the marginal revenue product of capital for formal firms (3.12),

$$\begin{aligned}RK_s^F &= R \sum^{M_s^F} K_{si}^F \\ &= \frac{\sigma}{\sigma-1} \alpha_s (1-T_F) \sum^{M_s^F} \frac{1-\tau_{Ysi}}{1+\tau_{Ksi}} P_{si}^F Y_{si}^F,\end{aligned}$$

where M_s^F denotes the number of formal firms in industry s. Therefore, I can express the marginal revenue product of capital of formal firms at the industry level as

$$\begin{aligned}\overline{\mathrm{MRPK}}_s^F &\coloneqq \alpha_s \frac{P_s Y_s}{K_s^F} \\ &= \frac{\sigma}{\sigma-1} \frac{R}{1-T_F} \left(\sum^{M_s^F} \frac{1-\tau_{Ysi}}{1+\tau_{Ksi}} \frac{P_{si}^F Y_{si}^F}{P_s Y_s} \right)^{-1}.\end{aligned}$$

Analogously, aggregating the marginal revenue product of labor for formal firms (3.13),

$$\begin{aligned} wL_s^F &= w\sum^{M_s^F} L_{si}^F \\ &\frac{\sigma}{\sigma-1}(1-\alpha_s)(1-T_F)\sum^{M_s^F}(1-\tau_{Ysi})P_{si}^F Y_{si}^F. \end{aligned}$$

I then define

$$\begin{aligned} \overline{\mathrm{MRPL}}_s^F &\coloneqq (1-\alpha_s)\frac{P_s Y_s}{L_s^F} \\ &= \frac{\sigma}{\sigma-1}\frac{w}{1-T_F}\left(\sum^{M_s^F}(1-\tau_{Ysi})\frac{P_{si}^F Y_{si}^F}{P_s Y_s}\right)^{-1}. \end{aligned}$$

Combining the above two expressions, I get the total factor productivity of revenue of the formal sector within an industry,

$$\begin{aligned} \overline{\mathrm{TFPR}}_s^F &\coloneqq \frac{P_s Y_s}{{K_s^F}^{\alpha_s}{L_s^F}^{1-\alpha_s}} \\ &= \left(\frac{\overline{\mathrm{MRPK}}_s^F}{\alpha_s}\right)^{\alpha_s}\left(\frac{\overline{\mathrm{MRPL}}_s^F}{1-\alpha_s}\right)^{1-\alpha_s}. \end{aligned}$$

Next, I do the same for informal firms. From the marginal revenue product of capital for informal firms (3.14),

$$\begin{aligned} RK_s^I &= R\sum^{M_s^I} K_{si}^I \\ &= \frac{\sigma}{\sigma-1}\alpha_s\sum^{M_s^I}\frac{1-\tau_{Ysi}}{(1+\tau_{Ksi})(1+\gamma_{si})}P_{si}^I Y_{si}^I, \end{aligned}$$

where M_s^I denotes the number of informal firms in industry s. Then,

$$\begin{aligned}\overline{\text{MRPK}}_s^I &\coloneqq \alpha_s \frac{P_s Y_s}{K_s^I} \\ &= \frac{\sigma}{\sigma - 1} R \left(\sum^{M_s^I} \frac{1 - \tau_{Ysi}}{(1 + \tau_{Ksi})(1 + \gamma_{si})} \frac{P_{si}^I Y_{si}^I}{P_s Y_s} \right)^{-1}.\end{aligned}$$

Aggregating the marginal revenue product of labor for informal firms (3.15),

$$\begin{aligned}wL_s^I &= w \sum^{M_s^F} L_{si}^I \\ &\frac{\sigma}{\sigma - 1}(1 - \alpha_s) \sum^{M_s^I} (1 - \tau_{Ysi}) P_{si}^F Y_{si}^F.\end{aligned}$$

And,

$$\begin{aligned}\overline{\text{MRPL}}_s^I &\coloneqq (1 - \alpha_s) \frac{P_s Y_s}{L_s^I} \\ &= \frac{\sigma}{\sigma - 1} w \left(\sum^{M_s^I} (1 - \tau_{Ysi}) \frac{P_{si}^I Y_{si}^I}{P_s Y_s} \right)^{-1}.\end{aligned}$$

Again, combining this expressions I get the aggregated total factor productivity of revenue of the informal sector within an industry,

$$\begin{aligned}\overline{\text{TFPR}}_s^I &\coloneqq \frac{P_s Y_s}{{K_s^I}^{\alpha_s} {L_s^I}^{1 - \alpha_s}} \\ &= \left(\frac{\overline{\text{MRPK}}_s^I}{\alpha_s} \right)^{\alpha_s} \left(\frac{\overline{\text{MRPL}}_s^I}{1 - \alpha_s} \right)^{1 - \alpha_s}.\end{aligned}$$

Then, I can express the joint marginal product of capital and labor as

$$\begin{aligned}\overline{\mathrm{MRPK}}_s &\coloneqq \alpha_s \frac{P_s Y_s}{K_s} \\ &= \alpha_s \frac{P_s Y_s}{K_s^F + K_s^I} \\ &= \left(\frac{1}{\overline{\mathrm{MRPK}}_s^F} + \frac{1}{\overline{\mathrm{MRPK}}_s^I} \right)^{-1}\end{aligned}$$

and,

$$\overline{\mathrm{MRPL}}_s = \left(\frac{1}{\overline{\mathrm{MRPL}}_s^F} + \frac{1}{\overline{\mathrm{MRPL}}_s^I} \right)^{-1}.$$

Thus, I define the total factor productivity of revenue of industry as

$$\begin{aligned}\overline{\mathrm{TFPR}}_s &\coloneqq \frac{P_s Y_s}{K_s^{\alpha_s} L_s^{1-\alpha_s}} \\ &= \left(\frac{\overline{\mathrm{MRPK}}_s}{\alpha_s} \right)^{\alpha_s} \left(\frac{\overline{\mathrm{MRPL}}_s}{1-\alpha_s} \right)^{1-\alpha_s}.\end{aligned}$$

Hence, from the above expression and the prices aggregation equation for industry s, I derive the total factor productivity of an industry,

$$\begin{aligned}\mathrm{TFP}_s &\coloneqq \frac{Y_s}{K_s^{\alpha_s} L_s^{1-\alpha_s}} \\ &= \left[\sum^{M_s^F} \left(A_{si} \frac{\overline{\mathrm{TFPR}}_s}{\mathrm{TFPR}_{si}^F} \right)^{\sigma-1} + \sum^{M_s^I} \left(A_{si} \frac{\overline{\mathrm{TFPR}}_s}{\mathrm{TFPR}_{si}^I} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}.\end{aligned}$$

3.4 Gains from Reallocation

I now will derive an expression for aggregated output and compare it with an efficient (no distortions) benchmark. Denoting the aggregated amount of capital and labor used across all industries as K^j and L^j with j in $\{F, I\}$. Then, from the definitions of marginal revenue products

of capital and labor at the industry level and equation (3.2) I obtain

$$K_s^j = K^j \frac{\alpha_s \theta_s (\overline{\mathrm{MRPK}}_s^j)^{-1}}{\sum_{s'}^{S} \alpha_{s'} \theta_{s'} (\overline{\mathrm{MRPK}}_{s'}^j)^{-1}},$$

and

$$L_s^j = L^j \frac{(1-\alpha_s) \theta_s (\overline{\mathrm{MRPL}}_s^j)^{-1}}{\sum_{s'}^{S} (1-\alpha_{s'}) \theta_{s'} (\overline{\mathrm{MRPL}}_{s'}^j)^{-1}}.$$

From the above expressions it is seen that if the marginal revenue products change at the firm level but remain constant at the industry level , then the allocation of capital and labor across industries remain the same. I will use this to compare the observed output against the one of an efficient economy.

Equalizing distortions within industries and eliminating the distinction between formal and informal firms implies that

$$\mathrm{TFPR}_{si}^F = \mathrm{TFPR}_{si}^I = \overline{\mathrm{TFPR}}_s,$$

in which case efficient industry TFP will depend only on the physical productivity of all firms belonging in each industry,

$$\bar{A}_s = \left[\sum_{i=1}^{M_s} A_{si}^{\sigma-1} \right]^{\frac{1}{\sigma-1}} \tag{3.16}$$

Finally, due to the observation that total capital and labor used in each industry does not change, the gains from reallocations in output terms can be derived from the expressions of TFP_s and $\bar{A}_s$,

$$\frac{Y}{Y_{eff}} = \prod_{s=1}^{S} \left[\frac{\overline{\mathrm{TFPR}}_s}{\bar{A}_s} \right]^{\theta_s}. \tag{3.17}$$

Chapter 4

Empirical Implementation

4.1 Estimation Procedure

The first parameters that I need to set are the rental rate of capital, the elasticity of substitution and the industries' capital shares. Following Hsieh and Klenow (2009), I set the rental rate of capital R to 0.10, assuming a 5 percent real interest rate and inflation rate. Additionally, in said paper they set the elasticity of substitution σ to 3. Due to data availability I am required to work with industries defined at the NAICS 3-digit level in contrast to the ISIC 4-digit level employed in the original paper. Since industries are defined more broadly, one could argue that the degree of substitution between them will be lower and thus set σ to a value of 2.

Then, with industries defined at the NAICS 3-level I can use public data from the United States' Bureau of Economic Analysis to construct the industries' capital shares in the manufacturing, retailing and services sectors. For each year, I calculate industry s capital share as

$$\alpha_s = 1 - \frac{Remuneration_s}{Value\ added_s}.$$

Next, I want to estimate the distortions as functions of observable variables. From equations

(3.12) and (3.13), if firm si is formal,

$$(1-\tau_{Ysi})(1-T_F) = \frac{\sigma}{\sigma-1}\frac{1}{1-\alpha_s}\frac{wL^F_{si}}{P^F_{si}Y^F_{si}},$$
$$1+\tau_{Ksi} = \frac{\alpha_s}{1-\alpha_s}\frac{wL^F_{si}}{RK^F_{si}}.$$

On the other hand, if firm si is informal then, from equations (3.14) and (3.15),

$$1-\tau_{Ysi} = \frac{\sigma}{\sigma-1}\frac{1}{1-\alpha_s}\frac{wL^I_{si}}{P^I_{si}Y^I_{si}}$$
$$(1+\tau_{Ksi})(1+\gamma_{si}) = \frac{\alpha_s}{1-\alpha_s}\frac{wL^I_{si}}{RK^I_{si}}.$$

From the above equations I recover the capital distortion for formal firms and the output distortions for informal firms from the available data. Yet, in order to disentangle the remaining effects I need to make an additional assumption.

Assumption 1. *The fiscal burden over formal firms is constant across all industries and all states within each year. This is, every formal firm in a given year face the same formality tax T_F.*

The above assumption allows me to asses the fiscal burden on formal firms, this is $1-T_F$, by following the procedure done in Leal-Ordoñez (2013) and taking for each year, the ratio of total tax revenue from the formal sector to value added associated to these firms.

The remaining relevant variables are computed in the same way as in Hsieh and Klenow (2009). Industry s capital distortion and output distortion are

$$1+\bar{\tau}_{Ks} = \frac{\alpha_s}{1-\alpha_s}\frac{wL_s}{RK_s}$$

and

$$1-\bar{\tau}_{Ys} = \frac{\sigma}{\sigma-1}\frac{1}{1-\alpha_s}\frac{wL_s}{P_sY_s},$$

respectively. Total factor productivity of revenue of a firm, a sector or an industry can be computed as presented in section 3.3. For instance, total factor productivity of revenue of firm si is given by

$$\text{TFPR}_{si} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}}$$

and so on. Next, from the production function of Y_{si} in equation (3.4) and the demand for Y_{si} in equation (3.3), the production technology of firm si can be written as

$$A_{si} = \kappa_s \frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha_s} L_{si}^{1-\alpha_s}},$$

where $\kappa_s = (P_s Y_s)^{-\frac{1}{\sigma-1}}/P_s$. As noted by the authors, relative productivities and reallocation gains are unaffected by setting κ_s to one for each industry s.

Lastly, the value added share of intermediate good s is calculated as

$$\theta_s = \frac{P_s Y_s}{Y}.$$

With the above valued we can compute efficient TFP and gains from reallocation using equations (3.16) and (3.17).

4.2 Imputation of Missing Wages and Sample Definition

Recall that, in order to estimate the distortions levels and gains from reallocation, I need to know three key components from the firms: value added $P_{si}Y_{si}$, capital used K_{si} and total remunerations wL_{si}. The later one poses a problem since, in the data collected by the Mexican economic census, roughly three quarters of all firms report employing workers but paying zero wages or do not report it at all. In order to solve this setback I take on two complementary routes to estimate the wages of said firms. First, I take advantage that, within the questions asked in the census, there is some data related to the use of outsourcing such as “How much did you pay to another

company that hired and provided personal to this establishment" or "How much did you pay for professional services?". For firms that report it, I use the expenses found in this questions to input wages for firms that reported having hired workers but paying zero wages. Then, following the procedure done in Busso et al. (2012), by computing the average wage paid by (no outsourcing) firms with up to ten workers in the corresponding 3-digit sector in each year and use this estimate to impute wages for missing firms based on their number of workers.

Finally, I drop observations reporting non positive value added, capital usage or number of workers and, following Hsieh and Klenow (2009), I trim the top one percent tails of $\log(A_{si}M_s^{\frac{1}{\sigma-1}}/\bar{A}_s)$ and $\log(\mathrm{TFPR}_{si}/\overline{\mathrm{TFPR}}_s)$. The final sample consists of 12,432,379 establishments. The distribution of said establishments was among years and formality status was reported in table 2.1.

4.3 Corruption Indices

As previously mentioned, I use three distinct measures of corruption. The first one comes from the National Survey of Victimization and Perception about Public Security (ENVIPE); in particular, I employ two questions from this survey: "Is corruption on of the three top problems that worries you the most" and "In your state, are the people who commit robbery, fraud and extortion punished by the authorities". I create a ranking using the mean of responses from the above questions. Given its nature, I expect this index to be the farthest away from explaining the issue that I attempt to investigate for two main reasons: First, given the broadness of the former question, it might be the case that a person does perceive her state as very corrupt but there are other issues that she thinks are more problematic (insecurity for instance). Second, the latter question is more related to corruption among agents involved in offenses against property (like police officers) and this might not directly translate into the type of corruption which affects establishments.

The second one comes from the National Survey of Government Quality and Impact (ENCIG).

Here, I employ responses to the questions “In your opinion, how frequent do you think corruption practices occur among businessmen?”, “In your opinion, how frequent do you think corruption practices occur among your state government?” and “In your opinion, how frequent do you think corruption practices occur among your local government?” As with the previous survey, I create the ranking using the frequencies of responses from this questions. Assuming peoples perception are correct, this index may be better at measuring the phenomenon that I intend to study if: a higher degree of corruption among businessman implies more briber or extortions are paid; or if more frequent corruption practices among local governments implies more corrupt Government employees.

It is important to notice that the previous two surveys are conducted over households, as so, their perception of corruption may vary significantly in comparison to an establishment’s point of view.

The final one derives from the National Survey of Firms’ Victimization (ENVE). This survey is conducted over establishments and ask them, among other things, if a Government employee has demanded any sort of favor or bribes. As with the others, I take the frequencies of responses to create the ranking. Given its nature, I expect this index to be the best suited in measuring corruption between establishment and Government agents but still an important limitation with thus survey is that, of all establishments surveyed, none report having employed more that ten workers; as so, the incidents and environment of this small firms may no reflect that of bigger establishments.

Due to a lack of information, I am only capable of obtaining this indices for the year 2013. As so, I will no be available to examine the misallocation dynamic across time. Still, I hope that this one time period is enough to link, at least partially, the degree of misallocation with the corruption level of the states. I present the ranking of all three indices in table 4.1.

Table 4.2 shows the correlation between the three indices. As one can see, there is significant variation between them, this might be due to the fact that there is not a single absolute way of measuring corruption and the three indices can be measuring different components of the bigger

phenomenon that do not necessarily translate into each other.

Table 4.1: Corruption Rankings Among the States of Mexico

State	*ENCIG*	*ENVIPE*	*ENVE*
Aguascalientes	6	6	19
Baja California	8	26	11
Baja California Sur	11	25	13
Campeche	13	8	17
Chiapas	2	20	7
Chihuahua	4	18	2
Ciudad de México	31	32	29
Coahuila	19	5	15
Colima	17	13	6
Durango	20	21	4
Guanajuato	30	16	14
Guerrero	12	15	20
Hidalgo	3	11	28
Jalisco	27	17	9
México	21	30	27
Michoacán	29	22	10
Morelos	32	27	31
Nayarit	7	1	16
Nuevo León	1	29	24
Oaxaca	24	12	23
Puebla	28	24	21
Querétaro	10	7	3
Quintana Roo	16	28	26
San Luis Potosí	26	4	1
Sinaloa	14	14	25
Sonora	25	10	22
Tabasco	23	31	8
Tamaulipas	5	9	18
Tlaxcala	22	23	32
Veracruz	15	19	30
Yucatán	9	2	12
Zacatecas	18	3	5

Source: Author's calculations based on data from INEGI.

Table 4.2: Correlation Between Corruption Indices

	ENCIG	*ENVIPE*	*ENVE*
ENCIG	1.00		
ENVIPE	0.14	1.00	
ENVE	0.28	0.33	1.00

Source: Author's calculations based on data from INEGI.

Chapter 5

Results

5.1 Productivity and Distortions

From now on I will denote $\log(A_{si}M_s^{\frac{1}{\sigma-1}}/\bar{A}_s)$ as ***TFPQ*** and $\log(\mathrm{TFPR}_{si}/\overline{\mathrm{TFPR}}_s)$ as ***TFPR***. Additionally, I denote the *scaled total capital distortion* of a firm as the logarithm of the observed capital distortion divided by the aggregated capital distortion, this is

$$\log\left(\frac{1+\tau_{Ksi}}{1+\bar{\tau}_{Ks}}\right)$$

for formal firms and

$$\log\left(\frac{(1+\tau_{Ksi})(1+\gamma_{si})}{1+\bar{\tau}_{Ks}}\right)$$

for informal firms. Analogously, I denote the *scaled total output distortion* as

$$\log\left(\frac{(1+\tau_{Ksi})(1+T_F)}{1+\bar{\tau}_{Ys}}\right)$$

for formal firms and

$$\log\left(\frac{1+\tau_{Ysi}}{1+\bar{\tau}_{Ys}}\right)$$

for informal ones. The interpretation of this coefficients is the following: An establishment presenting a $TFPQ$ of 1.10 (0.90) is ten percent more (less) productive that the average firm of its industry. An establishment presenting a $TFPR$ of 1.10 (0.90) comprises a share of its industry value added that is ten percent greater (smaller) in the efficient economy than in the actual data. An establishment displaying a scaled total capital distortion of 1.10 (0.90) faces a cost of capital that is 10 percent greater (smaller) than that of the average firms in its industry. Finally, an establishment displaying a scaled total output distortion of 1.10 (0.90) faces restrictions to output that are 10 percent more beneficial (harmful) than those of the average firm in its industry.

I present the average degree of distortion by formality status for each year in table 5.1. As expected, on average the scaled total capital distortion is increasing as one moves from formality to informality. This is consistent with the idea that informal firms face capital restriction and thus have to pay a higher price in the capital market. One can also notice that the degree of this distortion has increased overtime for each of the formal, informal and semi formal establishments. Interestingly, outsourcing firms seem to have, on average, extraordinary low capital costs when compared to other firms in the same industry.

On the other hand, the scaled total output distortion is also increasing from formality to informality. This again is consistent with the idea that formal firms have to pay Government taxes which pose an additional *harmful* restriction. Contrary to the capital distortion, the harmful output distortion seems to have, in general, decreased over time.

Next, by looking at the average *TFPQ* in table 5.2 it is seen that, on average, formal firms tend to be more productive that semi formal firms and this in turn are more productive than informal firms. It is also remarkable to note that between formal, semi formal and informal firms the tendency across time has been to become less productive whereas outsourcing establishment, although they started very unproductive, have become more and more productive in each period. Finally, formal establishment tend to have a greater *TFPR* than informal establishments. On the other hand, outsourcing firms are the only ones that seem to be, at least in recent years, near their efficient share of value added relative to the efficient benchmark.

Table 5.1: Distortions by Formality Status

Variable	**Year**	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
Scaled	1993	0.41	0.39	0.89	
Total	1998	0.39	0.51	0.87	−0.05
Capital	2003	0.52	0.66	1.05	0.13
Distortion	2008	0.75	0.84	1.18	−1.35
	2013	0.83	1.01	1.58	−1.33
Scaled	1993	−0.30	−0.32	0.89	
Total	1998	−0.18	−0.15	0.74	−0.08
Output	2003	−0.14	−0.11	0.67	−0.05
Distortion	2008	0.07	0.09	0.71	−1.43
	2013	0.00	0.02	0.87	−1.53

Source: Author's calculations based on data from INEGI.

Table 5.2: *TFPQ* and *TFPR* by Formality Status

Variable	**Year**	*Formal*	*Semi formal*	*Informal*	*Outsourcing*
TFPQ	1993	−0.10	−0.18	−3.12	
	1998	−0.64	−0.53	−2.90	−2.31
	2003	−0.55	−0.53	−2.75	−1.76
	2008	−1.05	−1.10	−3.26	−1.79
	2013	−0.90	−1.00	−3.33	−1.69
TFPR	1993	0.38	0.40	−0.68	
	1998	0.35	0.36	−0.40	0.08
	2003	0.35	0.38	−0.26	0.11
	2008	0.22	0.22	−0.26	0.96
	2013	0.35	0.39	−0.23	1.02

Source: Author's calculations based on data from INEGI.

5.2 Gains from Reallocation

Next, in table 5.3 I present the gains from reallocation (in terms of percentage of output) for each sector of the economy by year as computed with equation (3.17). One can see that, each year, the potential gains have increased and that the retailing sector seems to be the most affected by inefficiencies.

Table 5.3: Gains from Reallocation

Sector	1993	1998	20003	2008	2013
Manufacturing	95.23%	96%	93%	105%	165%
Retailing	-	173%	182%	241%	284%
Services	128%	134%	142%	184%	228%

Source: Author's calculations based on data from INEGI.

5.3 Corruption

Finally, I wish to link the degree of scaled total capital distortions with the level of corruption of the states. To this end, I run a OLS model with the scaled capital distortion as dependent variable and size category, age category, formality, *TFPQ* and the three distinct indices of corruption among states as explicative variables for the year 2013. I present the results of this regressions in tables 5.4, 5.5 and 5.6 for the ENCIG, ENVIPE and ENVE indices, respectively. The first column of each table stands for the regression employing the full 2013 sample, in the second column I excluded outsourcing establishments and in the third and last column y considered only formal and semi formal establishments. Note that, since the dependent variable is a $\log$, all coefficients should be interpreted as semi-elasticities.

Notice that, when considering the full 2013 sample, moving one place up in the ENCIG ranking implies, *ceteris paribus*, an increase of one percent in scaled total capital distortion; doing the same in the ENVIPE or ENVE rankings implies a two percent increase in said variable. Although this effects are small, when compared to the mean of the dependent variable (which translates into an average distortion of twelve percent), this effects represent an increase of eight and sixteen percent relative to the mean, respectively. Also notice that, when removing outsourcing firms from the sample, the effect of every index falls nearly by 40 percent.

Table 5.4: Effect of Corruption on Scaled Total Capital Distortion, ENCIG

	All	*No Outsourcing*	*Informal and Semi formal*
Size category	-0.200***	-0.035***	-0.064***
	(0.002)	(0.002)	(0.002)
Age category	-0.136***	-0.072***	-0.070***
	(0.001)	(0.001)	(0.001)
Formality status	-0.784***	0.425***	0.556***
	(0.002)	(0.002)	(0.004)
TFPQ	-0.031***	0.045***	0.042***
	(0.000)	(0.000)	(0.000)
index_ENCIG	0.011***	0.007***	0.008***
	(0.000)	(0.000)	(0.000)
Constant	3.669***	0.497***	0.118***
	(0.007)	(0.008)	(0.014)
Mean dept. var.	1.12	1.49	1.53
Obs.	3,357,972	2,914,746	2,732,843
R-sq.	0.07	0.02	0.02
F-stat.	49,214.23	14,571.42	8,886.84

Source: Author's calculations based on data from INEGI.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 5.5: Effect of Corruption on Scaled Total Capital Distortion, ENVIPE

	All	*No Outsourcing*	*Informal and Semi formal*
Size category	-0.204***	-0.038***	-0.066***
	(0.002)	(0.002)	(0.002)
Age category	-0.134***	-0.071***	-0.069***
	(0.001)	(0.001)	(0.001)
Formality status	-0.785***	0.420***	0.549***
	(0.002)	(0.002)	(0.004)
TFPQ	-0.035***	0.043***	0.039***
	(0.000)	(0.000)	(0.000)
index_ENVIPE	0.019***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)
Constant	3.530***	0.435***	0.069***
	(0.007)	(0.008)	(0.014)
Mean dept. var.	1.12	1.49	1.53
Obs.	3,357,972	2,914,746	2,732,843
R-sq.	0.07	0.03	0.02
F-stat.	52,424.35	15,646.96	9,799.52

Source: Author's calculations based on data from INEGI.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 5.6: Effect of Corruption on Scaled Total Capital Distortion, ENVE

	All	*No Outsourcing*	*Informal and Semi formal*
Size category	-0.200***	-0.035***	-0.064***
	(0.002)	(0.002)	(0.002)
Age category	-0.135***	-0.072***	-0.069***
	(0.001)	(0.001)	(0.001)
Formality status	-0.785***	0.417***	0.543***
	(0.002)	(0.002)	(0.004)
TFPQ	-0.031***	0.045***	0.041***
	(0.000)	(0.000)	(0.000)
index_ENVE	0.020***	0.012***	0.012***
	(0.000)	(0.000)	(0.000)
Constant	3.511***	0.427***	0.069***
	(0.007)	(0.008)	(0.014)
Mean dept. var.	1.12	1.49	1.53
Obs.	3,357,972	2,914,746	2,732,843
R-sq.	0.07	0.03	0.02
F-stat.	53,103.67	15,963.88	10,117.68

Source: Author's calculations based on data from INEGI.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Chapter 6

Conclusion

Following previous works I assess the losses in productivity associated to the informal sector in Mexico while adding three things: First, I modify the original Hsieh-Klenow model to allow for endogenous decision regarding in which sector, formal or informal, do firms wish to operate. Second, I give a possible explication for the problem of capital distortions persistent in the firms where, if an establishment belonging to one of the bottom five least corrupt states were to change to one of the top five most corrupt states it would face, *ceteris paribus*, over fifty percent more capital distortions. Third, I document how outsourcing firms should be more carefully examined given their behavior and characteristics.

References

Acemoglu, D., & Verdier, T. (1998). "Property rights, corruption and the allocation of talent: A general equilibrium approach." *The Economic Journal*.

Acemoglu, D., & Verdier, T. (2000). "The choice between market failures and corruption." *The American Economic Review*.

Aidt, T. S. (2003). "Economic analysis of corruption: A survey." *The Economic Journal*.

Amirapu, A., & Gechter, M. (2014). *Labor regulations and the cost of corruption: Evidence from the indian firm size distribution.*

Bardhan, P. (2006). "The economist's approach to the problem of corruption." *World Developmnet*.

Bento, P., & Restuccia, D. (2016). "Misallocation, establishment size, and productivity." *NBER Working Paper*.

Busso, M., Fazio, M. V., & Levy, S. (2012). "(in)formal and (un)productive: The productivity costs of excessive informality in mexico." *IDB Working Paper Series*.

Dabla-Norris, E., Gradstein, M., & Inchauste, G. (2008). "What causes firms to hide output? the determinants of informality." *Journal of Development Economics*.

D'Erasmo, P. N., & Moscoso, H. J. (2012). "Financial structure, informality and development." *Journal of Monetary Economics*.

Farrell, D. (2004). "The hidden dangers of the informal sector." *McKinsey Quarterly*.

Goyette, J., & Kouamé, W. A. (2017). *Distorsion, policy inefectiviness and tax evasion.*

Hsieh, C.-T., & Klenow, P. J. (2009). "Misallocation and manufacturing tfp in china and india."

The Quarterly Journal of Economics.

Hsieh, C.-T., & Klenow, P. J. (2014). "The life cycle of plants in india and mexico." *The Quarterly Journal of Economics*.

Kurer, O. (1993). "Clientelism, corruption and the allocation of resources." *Public Choice*.

Leal-Ordoñez, J. C. (2013). "Tax collection, the informal sector, and productivity." *Banco de México Working Paper*.

Mauro, P. (1995). "Corruption and growth." *The Quarterly Journal of Economics*.

Restuccia, D., & Rogerson, R. (2013). "Misallocation and productivity." *Review of Economic Dynamics*.