

Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2013-22

Tax Collection, The Informal Sector, and Productivity

Julio C. Leal-Ordoñez
Banco de México
CIDE

December 2013

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively of the authors and do not necessarily reflect those of Banco de México.

Tax Collection, The Informal Sector, and Productivity*

Julio C. Leal-Ordoñez[†]

Banco de México
CIDE

Abstract: Some authors argue that informality is associated with distorted firm decisions and inefficiency. In this paper, I assess the quantitative effect of incomplete tax enforcement on aggregate output and productivity using a dynamic general equilibrium framework. I calibrate the model using data for Mexico and investigate the effects of introducing enforcement improvements. Under complete enforcement, labor productivity and output would be 19% higher under perfect competition and 34% higher under monopolistic competition. The source of this gain is the removal of distortions induced by incomplete enforcement of taxes which affect the economy in three ways: by reducing the capital-labor ratios of informal establishments; by allowing low-productive entrepreneurs to enter; and by misallocating resources towards low-productive establishments. I isolate the effects of pure factor misallocation, distorted occupational choices, capital accumulation, and complementarities.

Keywords: Tax enforcement, TFP, the informal sector.

JEL Classification: E23, E26, O17, O40.

Resumen: Algunos autores argumentan que la informalidad se asocia a distorsiones en las decisiones de las empresas y genera ineficiencia. En este artículo, investigo el efecto del cobro imperfecto de impuestos sobre el producto y la productividad usando un modelo de equilibrio general dinámico. Calibro dicho modelo usando datos de México y simulo una mejora en la capacidad de cumplir las leyes impositivas. Bajo cobro perfecto, la productividad y el producto aumentarían 19% en competencia perfecta, y 34% en competencia monopolística. La fuente de esta ganancia es la remoción de distorsiones inducidas por el cobro imperfecto, lo que afecta la economía de tres maneras: reduciendo los ratios capital-trabajo de los establecimientos informales; permitiendo entrar a los emprendedores de baja productividad; y asignando los recursos de manera ineficiente hacia establecimientos poco productivos. Aíslo los efectos de la asignación factorial ineficiente, de las decisiones ocupacionales distorsionadas, de la acumulación de capital y de las complementariedades.

Palabras Clave: Cobro de impuestos, productividad, sector informal.

*This paper is an updated version of a chapter of my PhD dissertation entitled “Informal Sector and Economic Development”. A previous version was entitled “Informal Sector, Productivity, and Tax Collection”. I would like to thank Richard Rogerson, Berthold Herrendorf and Ed Prescott for their advice. I would also like to thank David Lagakos, Roozbeh Hosseini, Galina Vershagina, Natalia Kovrijnykh, and Manjira Datta for their comments and suggestions. A version of this paper is currently “in press” for Review of Economic Dynamics, 2013.

[†] Dirección General de Investigación Económica. Email: jleal@banxico.org.mx.

1. Introduction

What are the implications of informality for economic development? Some authors have argued that firms operating in the informal sector are less regulated and less taxed than firms in the formal sector, which allows them to operate more efficiently. This represents a positive force in the economy (see Schneider & Enste, 2002). In contrast, other authors have highlighted distortions that might arise in the presence of a large informal sector and, specifically, of incomplete enforcement of taxes. For example, Lewis (2004) argues that informality distorts the “natural” competitive process, as informal firms enjoy an “unfair” cost advantage through tax avoidance; and Farrell (2004) reports that informal firms reduce their scale of operation in order to remain undetected by the government, which makes them less efficient.¹

In this paper, I am interested in quantifying the extent to which distortions associated with the way firms avoid taxes affect output and productivity. To do this, I start with a general equilibrium model of occupational choice and capital accumulation that includes a taxation policy with limited enforcement. In this framework, incomplete tax enforcement is the source of plant-idiosyncratic distortions similar to those studied by Restuccia & Rogerson (2008) and Hsieh & Klenow (2007), among others. Individuals have heterogeneous entrepreneurial abilities (as in Lucas, 1978) and each faces a discrete occupational choice: whether to be a formal entrepreneur, an informal entrepreneur or an employee (as in Rauch, 1991). If formal, the entrepreneur pays taxes; if informal, the entrepreneur faces a probability of being detected that depends positively on the amount of capital hired. Therefore, only small firms are able to evade taxes, because it is more difficult for the government to detect them.

The novelty in this paper is that it connects informal-sector data for a typical developing country with a general equilibrium model in which the consequences of incomplete enforcement of taxes can be studied. I calibrate the model using data for Mexico, an economy where 45% of the employees work in informal establishments. I then investigate the effects of improving enforcement. My main finding is that under complete enforcement, Mexico’s labor productivity and output would be 19% higher in the baseline model and 34% higher in a model with monopolistic competition. These gains come, to an important extent, from a novel mechanism through which better enforcement, and therefore higher taxes, increases capital, as the benefits of staying small fall.

A second contribution in the paper is that the gains from full enforcement are analyzed in the context of five leading papers in the literature on resource misallocation across plants. I classify these papers into three groups: a) Restuccia & Rogerson (2008), which assumes free entry to determine the mass of firms in equilibrium; b) Gollin (1995) and Guner *et al.* (2008), which emphasize occupational choices; and c) Hsieh & Klenow (2007) and Jones (2011), which focus on linkages and complementar-

¹ Similarly, Levy (2008) argues that informality constitutes an implicit subsidy for low-productive activities; and Bigio & Zilberman (2011) show that when screening of observable inputs occurs, the optimal monitoring strategy of the tax enforcement authority induces firms to distort their input demands.

ities using models with monopolistic competition. The relevance of this particular exercise is twofold: first, it provides a way to decompose the gains from full enforcement; second, it clarifies the differences among the papers above by comparing the results that alternative models provide for the same change in policy. Finally, the paper also makes contributions on the empirical side, which I detail below.

In Section 3, I present the baseline model. Two features of the equilibrium are key to understanding the distortionary aspects of incomplete enforcement. The first is that entrepreneurs with low productivity choose informality, while the more productive ones choose to operate in the formal sector. This occurs because any firm below a certain threshold can avoid detection and thereby operate informally and increase profits by avoiding taxes, at no additional cost. This feature induces two types of distortions: a misallocation of resources to establishments with low productivity and an increase in the number of unproductive entrepreneurs in the economy (i.e., a distortion of occupational choices). The second feature of the equilibrium is that informal establishments optimally reduce their scale to remain undetected by the government. This brings about a distortion in the capital of informal establishments, because the probability of being detected rises with the amount of capital hired. When full enforcement is introduced, these distortions disappear. However, since all entrepreneurs become formal, the tax burden increases for a group of previously untaxed entrepreneurs. At a basic level, improving enforcement involves a tradeoff between fewer distortions and more taxes.

My results suggest that when full enforcement is introduced, the gain of fewer distortions is greater than the cost of more taxes; however, when there are only partial improvements in enforcement, this is not necessarily the case. In particular, the effect of marginal changes in enforcement relies crucially on the size of the informal sector. If informality is very high, marginal improvements in enforcement reduce output because taxes increase for a high-productive set of marginal firms, while the gain from reducing distortions is relatively small. The opposite logic follows when informality is low. Thus, returning to the fundamental question regarding the implications of informality in the economy, the answer is: it depends. When I compute the equilibrium for a range of enforcement levels, I find that there is an inverted-U relationship between output and enforcement. Mexico is at the bottom of this curve where the positive and the negative effects roughly offset each other. If Mexico improves enforcement only marginally, output barely changes; but if Mexico goes all the way to full enforcement, the gains are sizable.

In Section 7, I decompose the gains from full enforcement. I start by obtaining the “pure misallocation effect”. To do this, I use a simplified version of the model: a static endowment economy. In this simple economy, the only source of changes in aggregate output is the reallocation of existing resources across firms. When I introduce full enforcement, marginal products are equalized and the allocation of resources improves, which increases output by 2% relative to the benchmark.

Once marginal products are equalized, output can be expressed as a function of four factors: the capital stock, the supply of employees, the mass of firms, and the average entrepreneurial ability. The only way through which aggregate output can now be modified is by changing the levels of these

factors. I investigate the effects of full enforcement on the levels of each factor following the guidelines of the five leading papers mentioned previously.

I first bring the model closer to Restuccia & Rogerson (2008) (group “a” above) by shutting down the occupational choice channel and replacing it with a free-entry condition. This exercise teaches us about the importance of the effects of full enforcement on the mass of firms, while keeping the remaining three factors fixed. I find that output and productivity fall to a level that is 92.5% of the benchmark. This occurs due to higher taxes.

Next, I bring the model closer to Gollin (1995) and Guner *et al.* (2008) (group “b”), which teach us about the importance of occupational choices, while abstracting from changes in capital stock. When full enforcement is introduced, marginal entrepreneurs no longer find it attractive to remain in operation because they are forced to pay taxes and so become employees. Thus, the mass of entrepreneurs falls (which reduces output); however, average entrepreneurial ability and the supply of labor both rise (which increases output). As a result, output (and productivity) increases by 4.4% relative to the benchmark. Thus, occupational choices, in contrast to free entry, reverse the negative effect on the decrease in the mass of firms.

When I introduce capital accumulation bringing the model back to the baseline, output increases by 11%, suggesting that capital constraints on informal firms are an important source of the gains from full enforcement. This emphasizes the importance of the incentives to remain small that informality introduces.

There would also be a gain in the form of lower taxes. Under complete enforcement, the tax base is broadened, so even with a lower tax rate revenues would remain constant. This is precisely the core of the hypothesis in Lewis (2004), which argues that the combination of big government and incomplete enforcement creates the need to impose high taxes on the most productive part of the economy. I find that Mexico could lower taxes from a rate of 25% to one of 13% under full enforcement. This reduction would further increase output to a level 19% above the benchmark.

I perform one final exercise to investigate the importance of complementarity across varieties (and firms). This is done through a framework with monopolistic competition as in Dixit & Stiglitz (1977). This brings the model closer to Hsieh & Klenow (2007) and Jones (2011) (group “c”). As explained in Jones (2011), monopolistic competition amplifies resource misallocation, because complementarity puts extra weight on the “weak links.” More mechanically, there is amplification because distortions in one firm affect aggregate expenditure which, in turn, affects the optimal decisions of all firms in the market. I find that productivity gains from full enforcement under monopolistic competition triple, while output gains increase by a factor of 1.8.²

In summary, for the change in policy analyzed here, models such as Restuccia & Rogerson (2008),

²Hsieh & Klenow (2007) get much higher effects. However their distortions are more general which leaves unclear what policies or institutions are the source of them. Also, in Hsieh & Klenow (2007) the mass of firms is exogenous.

where the mass of firms is endogenous and the labor supply is exogenous, tend to obscure the effects of resource misallocation. In contrast, models with occupational choices, such as Gollin (1995) and Guner *et al.* (2008), allow the average ability and the supply of labor to move in the opposite direction of the mass of firms, which reverses the negative effect of the decrease in entry. Finally, models with monopolistic competition, such as those used by Hsieh & Klenow (2007) and Jones (2011), amplify the effects of misallocation, because distortions in one firm affect aggregate expenditure, which in turn affects the decisions of the remaining firms.

Full enforcement cannot be achieved without incurring extra administrative costs. These costs have to be subtracted from the gains in output. In Section 9.1, I perform an assessment of the magnitude of these costs and find that full enforcement would require a large increase in monitoring costs. Nonetheless, these costs would remain at a level that represents a small fraction of the model's GDP (1.3%). In the appendix, I also investigate the sensitivity of my results to the assumption of a step function for the probability of detection; I find that this assumption is not crucial to my quantitative results.

There are two more contributions on the empirical side. First, I document that the establishment-size distribution of employment in Mexico exhibits a "missing middle." Second, I construct the distribution of formal establishments and show that it closely resembles that of US establishments, suggesting that the presence of the informal sector is the main driver of the "missing middle" phenomenon. Second, using a micro-business survey, I document how the operating capital of informal firms does not vary with size. I argue that this behavior is not due to credit constraints, because the majority of informal entrepreneurs do not report the lack of credit as a problem; in fact, they use their own savings to finance their businesses. Consistent with Moll (2010), self-financing seems to undo capital misallocation from financial frictions. In contrast, I argue that incomplete tax enforcement offers an alternative explanation for capital distortions in informal firms, that is, that they reduce capital to remain small and undetected. Furthermore, most of these firms are not registered with the tax authority, suggesting that they are not concerned about audits and that the probability of getting caught is close to zero. This is consistent with the monitoring strategy that Bigio & Zilberman (2011) show is optimal, which is the same as that used in my baseline model. These authors highlight the fact that such an enforcement strategy generates equilibrium features that are consistent with three empirical regularities: 1) the presence of a "missing middle" (Tybout, 2000); 2) an inverse relationship between the amount evaded and the size of the firm (Dabla-Norris *et al.*, 2008); and 3) the set of evaders producing inefficiently (La Porta & Shleifer, 2008 and Farrell, 2004). I add one more regularity that is satisfied when enforcement depends on capital: 4) informal firms exhibit low capital-labor ratios.

My paper relates to the literature in the following way. First, the burdens on productivity associated with informality can be understood as a specific case of the type of idiosyncratic distortions studied in the literature on resource misallocation across heterogeneous plants and TFP, identified in the recent work of Restuccia & Rogerson (2008) and Hsieh & Klenow (2007). These papers study hypothetical distortions that affect the prices of individual establishments, but the source of the distortions remains

unclear. The source of those distortions is precisely what my paper addresses.

Perhaps the two papers closest to my work are Gollin (1995) and Guner *et al.* (2008). The case study by Gollin (1995) for Ghana analyzes the impact of taxes on large establishments on productivity; likewise, incomplete tax enforcement is similar to a size-dependent policy of the type used by Guner *et al.* (2008). One important difference between my work and that of Guner *et al.* (2008) and Gollin (1995) is that the policies they consider do not distort the capital-labor ratios of establishments. As mentioned previously, this margin is distorted by the incomplete enforcement of taxes and proves to be quantitatively important. Another important difference is that, by adding a free-entry condition and monopolistic competition, I am able to isolate the importance of firm entry, occupational choices, and complementarity. Moreover, I analyze the different effects of continuous vs. discontinuous distortions.

Second, there are models in which the informal sector arises from the incomplete enforcement of taxes and/or regulations: Rauch (1991), Amaral & Quintin (2006), Dabla-Norris *et al.* (2008), and de Paula & Scheinkman (2007), among others. However, these authors focus mainly on the determinants of informality rather than on the consequences of incomplete enforcement.

Finally, de V. Cavalcanti & Antunes (2007), Prado (2011) and Moscoso-Boedo & D'Erasmus (2009) study the aggregate effects of informality within the context of GE models. However, these papers focus not on the way firms avoid taxes, but rather on other distortions associated with informality, such as financial constraints and the regulation of entry. Regulation and entry costs, emphasized originally by de Soto (1989), can be important for innovation incentives (see for example Herrendorf & Teixeira, 2011), but these seem to be less important for informality (see Kaplan *et al.*, 2007).

The paper is organized as follows. After this Introduction, Section 2 presents data documenting relevant facts about the informal sector and resource allocation in Mexico; Section 3 presents the model, while Section 4 characterizes the steady-state equilibrium; in Section 5, I calibrate the model, and in Section 6, I present the results of the baseline model. In Sections 7, 8 and 9, I decompose the gains from enforcement using various models of resource misallocation. The last section contains my conclusions.

2. Facts

There are four facts concerning informality in Mexico that are relevant to this paper: 1) the informal sector in Mexico is large, 2) the distribution of labor across establishment sizes in Mexico has a “missing middle,” 3) informal establishments are small, and 4) informal establishments operate with low capital-labor ratios.

I use establishment data classified by size taken from the Economic Census (2008) (henceforth, the census) and complement this with a household survey (ENAMIN, 2008) both compiled by the INEGI (the Mexican statistics bureau). The census does not capture establishments that lack a fixed structure, which represent a large proportion of the establishments in operation in Mexico; therefore, I use the ENAMIN survey to add this type of establishment to the census data. I focus on manufacturing, trade,

and services because in these sectors the unit of observation is the establishment. The other sectors in the census collect information from units that operate with more than one establishment.³ I also limit the analysis to establishments with at least one employee. Table 1 presents the full range of establishments used in this paper (henceforth, the census+enamin data).

Source	Establishments	Employment
Census	2,383,396	16,394,893
ENAMIN	1,141,592	2,833,618
Total	3,524,988	19,228,511

Table 1: Full range of establishments

2.1. *The size of the informal sector*

I construct two measures of informality, one obtained from the ENOE (a household survey) and the other from the census+enamin data. Both provide a similar measure of informality.

As in many other employment surveys, Mexico’s ENOE includes a question addressing whether or not the surveyed employee is registered with the Mexican Social Security Institute (IMSS). Following standard practice, I classify an employee as informal if she is not registered with the IMSS and as formal if she is. According to the ENOE, the employed population is around 40 million, so I work with the data in this survey to focus only on workers employed in the range of establishments captured in Table 1. After applying these filters, I find that 44.5% of employees are informal. This percentage corresponds to the ENOE (2008 Q2).

Another measure of informality can be obtained from the census+enamin data. The census does not classify employment according to IMSS registration status, but does include a classification of family workers and outsourced jobs, which typically are not registered at the IMSS. Using this definition, I obtain an informal share equal to 50% of employment.

2.2. *Informality and the distribution of employment*

In Table 2, I compare the employment distribution by establishment size in Mexico with that in the US, which I take as a relatively undistorted economy. From the first two columns, it is clear that in Mexico labor is concentrated in small establishments, whereas in the US, employment is concentrated in medium-sized and large ones. In particular, there is too little employment in medium-sized establishments in Mexico. This is a phenomenon previously documented in the literature, known as the “missing middle” (see Tybout, 2000 for a survey of developing countries). This phenomenon is also reflected in the mean size of establishments. In Mexico, the mean size is 5.5 workers, while in the US it is 17 workers (see also the last row of Table 2).

³With this fix, I keep 88% of census employment.

In the third column of Table 2, I present the distribution of employment in Mexico’s formal sector. To do so, I use data from the IMSS records. Notice that the IMSS records do not include information on the informal sector. Notice also that the distribution of the Mexico’s formal sector is very similar to that in the US. This suggests two things: 1) that the informal sector is the main driver of the “missing middle” phenomenon, and 2) that most informal employees work in small establishments or that informal establishments are small.

The size distribution of employment in the formal sector can also be obtained from the ENOE. In the case of the ENOE, the survey also includes information on the employment distribution of informal workers. Both distributions are presented in the last two columns of Table 2. The broad picture is similar to what we see using the IMSS records: most informal employees are working in small establishments and the distribution within the formal sector is similar to that in the US. See the Appendix for a detailed description of these distributions.

Table 2: Employment distribution: Mexico vs. the US

Size category	US	Mexico	Mex-Formal-1	Mex-Formal-2	Mex-Informal-2
Under 20	25.0	57.3	24.8	29.2	89.1
20 to 99	29.8	13.0	21.3	30.8	7.9
100 or more	45.3	29.8	53.9	40.0	3.0
Mean size	17	5.5	-	-	-

Sources: For the US, the Census Bureau’s *County Business Patterns* (2003). For Mexico, the INEGI, using the census+enamin data (2008). For Mex-Formal-1, the IMSS records (2008). For Mex-Formal-2 and Mex-Informal-2, the ENOE (2008 Q2). The mean size for the US is taken from Guner *et al.* (2008), while the mean size in Mexico corresponds to the census+enamin data.

2.3. Capital distortions

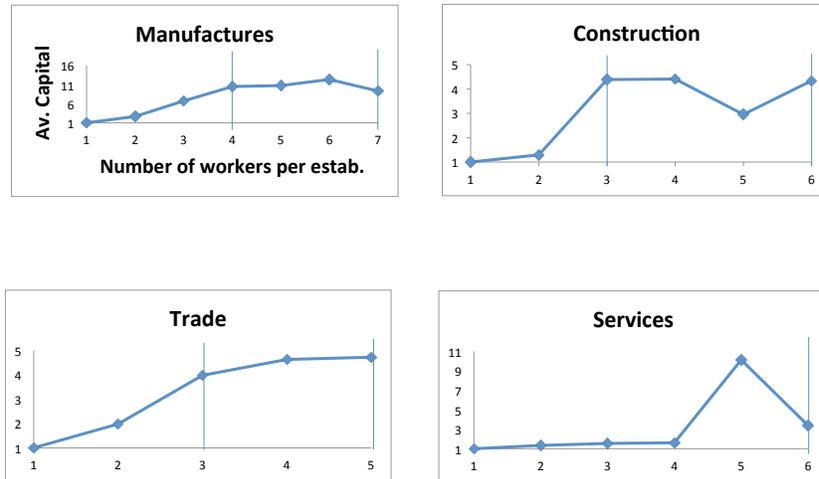
It is a well-established fact that informal firms operate with low capital per worker. For example, Thomas (1992) reports this for a survey of Peruvian establishments, de Paula & Scheinkman (2007) make the case for Brazilian firms, and Di-Giannatale (2011) for Mexico. However, my interest is on the particular shape of the capital distortions suffered by informal establishments. To investigate this, I use data from the ENAMIN, a survey that captures a large proportion of informality because its focus is on businesses with 6 or fewer workers (including the owner).

Figure 1 plots the capital (on the y-axis) and labor (on the x-axis) of informal establishments.⁴ There are four panels showing different sectors covered by the ENAMIN. In a typical model of heterogeneous firms without distortions, the capital-labor ratio is constant across size; this implies a linear relationship in a plot of capital vs. labor. Notice, however, that the relationship between capital and labor in Figure 1 is nonlinear; in particular, there is a range of labor for which capital is constant. This

⁴I use the following definition: formal if the business issues invoices, informal if not. In order to use invoices, the business must be registered with the tax authority.

range varies slightly from sector to sector, but is present in all. In other words, there is a capital level that informal establishments do not exceed. Also notice that this distortion on capital translates into low capital-labor ratios, as the literature reports.

Figure 1: Capital distortions of informal establishments



Sources: ENAMIN (2008). I group establishments according to size (number of workers). Then I compute the average capital (in pesos) for each group of establishments. Each panel shows: a) on the vertical axis, average capital relative to the smallest group of establishments (one worker) within each sector; and b) on the horizontal axis, the number of workers per establishment (see also the data appendix).

This type of distortion can be rationalized in a model in which the probability of being caught evading taxes increases with capital. In particular, assume that the probability of detection is zero if capital is less than or equal to a certain value, $b > 0$, and one if not. This introduces an incentive to remain small and to operate in the informal sector, hiring no more than b units of capital in order to benefit from tax evasion.

An alternative interpretation is that low capital-labor ratios reflect the borrowing constraints faced by informal entrepreneurs. Under this view, these individuals do not have access to formal credit because they lack the collateral needed to take out a loan. Since loans are used to finance capital, borrowing-constrained individuals operate with low capital.

2.4. Financial constraints or incomplete tax enforcement?

In the ENAMIN (2008), there is information that suggests that informal firms are not financially constrained. In the survey, the owners are asked how they financed their start-up costs: 50.5% stated

that they used their own savings; 19.4% said they borrowed from friends or relatives at zero interest; 13.5% stated that they did not need financing; 4.2% said they used a severance payment from a former job; and 2.2% used trade credit. Only 1.2% used credit from commercial banks. As discussed in Kaplan *et al.* (2007), this may mean that access to credit is difficult for small firms, or alternatively, that demand for credit is low. Fortunately, a second question in the ENAMIN is enlightening: the owners are asked what the main problem faced by their business is; only 2.39% answered that lack of credit was a problem; in contrast, 45% answered that low sales and excessive competition were the main problem, and 18% said that they had no problems. Therefore, informal entrepreneurs seem to be unaffected by credit constraints.

This evidence supports the research of Moll (2010), who argues that capital misallocation induced by borrowing constraints can be undone through self-financing because “...an entrepreneur without access to external funds can still accumulate internal funds over time to substitute for the lack of external funds.”⁵

Therefore, if financial frictions do not induce capital misallocation, then why do we see distortions in the capital-labor ratios of informal firms? I argue that the way taxes are enforced in developing countries induces these distortions on capital. I start with the observation that large establishments are more visible to tax authorities. In fact, a probability of detection that increases with firm size is a common assumption in the informal sector literature.⁶ Ihrig & Moe (2001) argue that capital in particular is harder to hide. Their empirical work on Asian economies shows that a rise in an index of tax enforcement is more effective in manufacturing than in services: manufacturing requires physical equipment and structures, whereas activities in the service sector are easier to hide. Following on from this idea, if the probability of detection depends on capital, we should see a larger proportion of informality in sectors that are less capital intensive. We can corroborate this using the ENOE: the proportion of informal employees in the trade sector is 53%, while in manufacturing it is 31%.

I also argue that the specific shape of the probability of detection is close to a step function: zero if capital is less than or equal to $b > 0$, and one if not. Although this assumption proves to be not crucial for the quantitative results, it is motivated by the following observations: 1) virtually no small businesses in the ENAMIN report taxes being a problem, whilst only 0.41% said that they were; 2) the majority of these businesses are not registered with any tax or municipal authority, which suggests that they are facing a probability of detection close to zero and explains why taxes are not a concern for these firms; 3) Figure 1 shows that capital flattens out for a certain labor range, consistent with the incentives of the step function.

⁵This possibility depends crucially on the evolution of the entrepreneur’s productivity over time. Moll shows that contrasting results in previous works in the literature on financial constraints and capital misallocation can be explained by differences in the assumption of the persistence of the productivity shock. Buera *et al.* (2011) and Midrigan & Xu (2010) are two examples with contrasting results.

⁶See for example, Rauch (1991), Fortin *et al.* (1997) and, more recently, de Paula & Scheinkman (2007)

Moreover, the research of Bigio & Zilberman (2011) favors the step function. The authors study the optimal monitoring strategy of a tax authority that maximizes revenue and faces a limited amount of resources, as well as costly monitoring. They show that when the tax authority screens observable inputs (such as capital, labor or land), the optimal strategy is a bang-bang solution: monitor small firms with zero probability up to a threshold and large firms with the highest possible probability. Moreover, the authors emphasize that the optimal policy in their paper is consistent with three empirical regularities: 1) the presence of a “missing middle” (Tybout, 2000); 2) an inverse relationship between the amount evaded and the size of the firm (Dabla-Norris *et al.*, 2008); and 3) some evaders producing inefficiently (La Porta & Shleifer, 2008). In addition to these three, I argue that when the probability of detection depends on observable capital, the model is also consistent with a fourth regularity: 4) informal firms exhibit a size range for which capital is constant (Figure 1).

In the baseline model, I assume that the probability of being caught depends on capital and takes the form of a step function. In the Appendix, I also explore how the quantitative results change if a continuous probability is assumed.

3. Baseline model

The previous section documented three facts. First, 45% of Mexican employees work in the informal sector; second, the distribution of Mexican employees by establishment size has a “missing middle”; third, most informal establishments are small; and fourth, informal firms operate with low capital per worker.

With this in mind, I build a model with heterogeneous entrepreneurial abilities and a taxation policy with limited enforcement. This policy links the probability of detection to the amount of capital hired by the tax evader. This will lead to an endogenously determined informal sector where establishments with low productivity sort into informality. This specification captures the fact that smaller establishments are more likely to be informal and that they also show a smaller capital-labor ratio.

There is a representative household in this economy that is populated by a continuum of individuals (members) of mass 1, as in Guner *et al.* (2008).⁷ At period zero, the household is endowed with K_0 units of capital while each member is endowed with entrepreneurial ability $z \in [\underline{z}, \bar{z}]$. This entrepreneurial ability is distributed according to PDF $g(z)$ and CDF $G(z)$ and does not evolve over time. Additionally, individuals have 1 unit of time each period.

The household has preferences over a sequence of consumption goods defined by:

$$\sum_{t=0}^{\infty} \beta^t u(C_t) \tag{1}$$

⁷Since there are no capital market imperfections, the assumption of a representative household is without loss of generality for the capital accumulation.

Where C_t is the consumption in period t . The household accumulates capital by making investments X_t , and as is standard; the accumulation is determined by the following rule:

$$K_{t+1} = X_t + (1 - \delta)K_t.$$

Each household member z can have one of three alternative occupations: entrepreneur in the formal sector, entrepreneur in the informal sector or employee in the formal or the informal sector.

Regardless of formality status, if an individual with ability z is an entrepreneur, she has access to the technology $f(z, k, l) = zk^{\theta_k}l^{\theta_l}$ and $0 < \theta_k + \theta_l < 1$, and I define $\gamma = \theta_k + \theta_l$. This technology exhibits decreasing returns to scale ensuring the coexistence of establishments with heterogeneous productivities (as in Lucas, 1978). If, on the other hand, the individual is an employee, the individual supplies 1 unit of labor that yields income w , independently of the value of z .

A government levies a tax τ_y on an establishment's output and the revenue is given back to the household as a lump-sum transfer. An output tax is equivalent to simultaneously taxing labor, capital, and entrepreneurial profits. The implicit assumption is that these three margins are taxed at the same rate. In Section 5 I provide a detailed discussion on the advantages of this choice.

Taxes can be avoided by operating in the informal sector, but tax avoidance comes at a cost. In particular, I assume that informal entrepreneurs face a probability of detection, and hence punishment. Once caught, the member will be given a fresh start in the next period and will be able to work in any of the three available occupations. The specification of the probability of detection will be referred to as the enforcement function. I focus on a function that depends on the amount of capital hired in the establishment. The following is assumed:

$$p(k(z)) = \begin{cases} 0, & k(z) \leq b \\ 1, & \text{else} \end{cases}, \quad (2)$$

where $k(z)$ is the capital hired by entrepreneur z and $b > 0$.

This enforcement policy gives informal entrepreneurs the opportunity of choosing to operate with a capital level equal to b or lower, that is, low enough not to get caught by the government, while still enjoying the benefits of tax avoidance. Note also that there is never going to be the case where, in equilibrium, an individual decides to operate informally and at the same time chooses capital greater than b , otherwise she will be caught and punished. This also means that no one gets caught in equilibrium (which, some might agree, describes government skills well), and furthermore, it means that we do not need to worry about the size of the punishment as long as its level is set high enough to reduce informal profits (if detected) to a level below formal profits. For purposes of completeness, the punishment is set equal to the current period earnings.

Finally, note that in terms of the equilibrium characterization of the occupational choices, this

specification and any other that includes a strictly increasing probability of being caught are equivalent. Both will characterize occupational choices with two thresholds in the range of entrepreneurial abilities z (see Section 4). However, the stepwise specification does affect the distortion suffered by informal establishments in their capital-labor ratios (this will be clear in a moment). Thus, in AppendixC, I explore how the quantitative results change when I use a continuous probability of detection.

3.1. Earnings for alternative occupations

I will now analyze the earnings for alternative occupations in more detail. As mentioned previously, an individual can have one of three possible occupations: employee, informal entrepreneur or formal entrepreneur. I assume employees are free to move across sectors and therefore a member working as an employee will simply earn wage w , regardless of the sector in which she is employed.

A formal entrepreneur with entrepreneurial ability z maximizes profits according to:

$$\pi_F(z; w, r) = \max_{\{l_F, k_F\}} \left\{ (1 - \tau_y) z k_F^{\theta_k} l_F^{\theta_l} - w l_F - r k_F \right\}, \quad (3)$$

where w is the wage rate and r is the price of capital. $k_F(z, w, r)$ and $l_F(z, w, r)$ denote the optimal choices of capital and labor respectively in the problem above.

An entrepreneur in the informal sector maximizes expected profits by taking into account the probability of detection mentioned in the previous section:

$$\pi_I(z; w, r) = \max_{\{l_I, k_I\}} \left\{ (1 - p(k_I)) \left(z k_I^{\theta_k} l_I^{\theta_l} - w l_I - r k_I \right) \right\}.$$

$k_I(z, w, r)$ and $l_I(z, w, r)$ denote the optimal choices of capital and labor respectively. Note that, as mentioned above, it is not optimal for any informal entrepreneur to operate with capital greater than b (otherwise her profits will be zero). However, she could choose to operate with capital equal to b , just low enough to prevent detection by the government, while still enjoying the benefits of tax avoidance. Therefore, the profits of an entrepreneur in the informal sector can also be expressed as:

$$\pi_I(z; w, r) = \max_{\{l_I, k_I\}} \left\{ z k_I^{\theta_k} l_I^{\theta_l} - w l_I - r k_I \right\} \quad s.t. \quad k_I \leq b \quad (4)$$

Once occupations are defined for each z (the occupational choices are described below), total household income is given by:

$$\begin{aligned}
E(w, r) &= \int_{\underline{z}}^{\bar{z}} I(z) \pi_I(z; w, r) dG(z) \\
&+ \int_{\underline{z}}^{\bar{z}} F(z) \pi_F(z; w, r) dG(z) + \int_{\underline{z}}^{\bar{z}} w(1 - I(z) - F(z)) dG(z)
\end{aligned} \tag{5}$$

where $F(z)$ and $I(z)$ are index functions and equal to 1 if the occupation for individual z is either formal or informal entrepreneur respectively, and zero otherwise. Similarly, let the index function, $I^c(z)$ equal 1 for the case where an informal entrepreneur is constrained (i.e., $k_I(z, w, r) = b$), and zero otherwise.

3.2. Government

In the present model, the government obtains revenue from two different sources: taxes and enforcement penalties. Given the nature of the enforcement policy, penalty revenues will be zero in equilibrium. I assume a balanced budget for the government in every period so that all proceeds from government activities are given back to the household in the form of a lump-sum transfer. The government budget balance condition is:

$$R_t = T_t, \quad \forall t \tag{6}$$

where R_t is tax revenue.

3.3. Representative household problem

The household chooses sequences of consumption, capital, and each member's occupation, taking as given the price sequences $\{w_t, r_t\}$, taxes τ_y , transfers $\{T_t\}$, and enforcement parameter b to maximize lifetime utility. The problem is:

$$\max_{\{C_t, K_t, I_t(z), F_t(z)\}} \left\{ \sum_{t=0}^{\infty} \beta^t u(C_t) \right\} \tag{7}$$

Subject to the following budget constraint:

$$C_t(z) + K_{t+1} - (1 - \delta)K_t = r_t K_t + E(w_t, r_t; \tau_y, b) + T_t, \quad \forall t \tag{8}$$

where K_0 is given and $E(w_t, r_t; \tau_y, b)$ is the same as in 5. I use $I(z; w, r)$ and $F(z; w, r)$ to represent optimal occupational decisions.

I focus on the steady-state (SS) equilibrium of this economy. As is standard, the first-order conditions of this problem in the steady state imply that:

$$r = \frac{1}{\beta} - (1 - \delta) \quad (9)$$

3.4. Market clearing

The market-clearing condition for the labor market will equate the aggregate labor demand from the two sectors to the labor supply:

$$\int_{\underline{z}}^{\bar{z}} I(z; w_t, r_t) l_I(z; w_t, r_t) dG(z) + \int_{\underline{z}}^{\bar{z}} F(z; w_t, r_t) l_F(z; w_t, r_t) dG(z) = \int_{\underline{z}}^{\bar{z}} W(z; w_t, r_t) dG(z) \quad (10)$$

where $W(z; w_t, r_t) = 1 - I(z; w_t, r_t) - F(z; w_t, r_t)$. Market clearing for the capital and good markets are, respectively:

$$\int_{\underline{z}}^{\bar{z}} I(z; w_t, r_t) k_I(z; w_t, r_t) dG(z) + \int_{\underline{z}}^{\bar{z}} F(z; w_t, r_t) k_F(z; w_t, r_t) dG(z) = K_t,$$

and,

$$C_t + K_{t+1} - (1 - \delta)K_t = \int_{\underline{z}}^{\bar{z}} I(z; w_t, r_t) y_I(z; w_t, r_t) dG(z) + \int_{\underline{z}}^{\bar{z}} F(z; w_t, r_t) y_F(z; w_t, r_t) dG(z).$$

3.5. Equilibrium definition

An equilibrium for this economy is sequences $\{C_t, K_{t+1}, w_t, r_t\}$ and $\{I_t(z), F_t(z)\} \forall z \in [\underline{z}, \bar{z}]$, whereby taking factor prices $\{w_t, r_t\}$, policies parameters τ_y and b , and transfers $\{T_t\}$, the household solves its problem, firms maximize their profits $\forall t$, and markets clear $\forall t$.

3.6. Steady state

In the following section, I will focus on the steady-state equilibrium. Since I define time-invariant taxation and enforcement policies, the dynamic part of this economy is no different from that in the standard growth model. In the steady state, factor prices, occupational decisions, aggregate capital and output are constant over time.

4. Properties of the model

4.1. Occupational choices

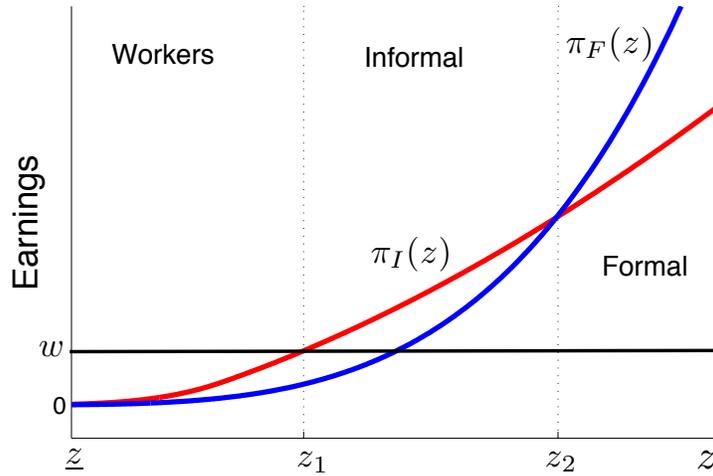
In this section, I analyze a number of properties of the model. The steady-state equilibrium is characterized by three thresholds, $\{z_1, z_c, z_2\}$, which summarize the occupational decisions of the agents

and whether the capital choices of informal entrepreneurs are constrained or unconstrained. Figure 2 provides a graph of the optimal occupational choices. I study the determination of these thresholds next.

In a steady-state equilibrium with positive formal and informal sectors, there exist thresholds $\{z_1, z_c, z_2\}$ such that:

1. $\forall z \in [z, z_1)$, individuals decide to be employees;
2. $\forall z \in [z_1, z_2)$, individuals are informal entrepreneurs;
3. $\forall z \in [z_2, \bar{z}]$, individuals are formal entrepreneurs;
4. when $z_c > z_1$, individuals $z \in (z_c, z_2)$ are constrained informal entrepreneurs; and when $z_c \leq z_1$, all informal entrepreneurs are constrained.

Figure 2: Characterization of occupational decisions



The determination of thresholds z_1 and z_2 is quite intuitive. For z_1 , note that employees earnings do not vary with entrepreneurial ability (the black line in Figure 2), while entrepreneurial profits increase with z (the red and blue lines). Not surprisingly, there exists a threshold z_1 such that $w = \pi_M(z_1; w, r)$, where $\pi_M(z; w, r) = \max \{\pi_I(z; w, r), \pi_F(z; w, r)\}, \forall z$. It follows that all agents with $z < z_1$ will become employees and the rest entrepreneurs. For z_2 , one must recall that informal establishments cannot operate with capital greater than b ; this restriction is more costly for entrepreneurs with a greater value of z , because they would naturally prefer to operate at a large scale given their high productivity. As a result, there exists a threshold z_2 such that $\pi_I(z_2; w, r) = \pi_F(z_2; w, r)$. It follows that entrepreneurs with ability $z < z_2$ would prefer the informal sector and the rest the formal sector.

The optimal choices of formal entrepreneurs are quite standard:

$$k_F(z, w, r) = ((1 - \tau_y)z)^{\frac{1}{1-\gamma}} \left(\frac{\theta_l}{w}\right)^{\frac{\theta_l}{1-\gamma}} \left(\frac{\theta_k}{r}\right)^{\frac{1-\theta_l}{1-\gamma}}, \quad (11)$$

$$l_F(z, w, r) = ((1 - \tau_y)z)^{\frac{1}{1-\gamma}} \left(\frac{\theta_l}{w}\right)^{\frac{1-\theta_k}{1-\gamma}} \left(\frac{\theta_k}{r}\right)^{\frac{\theta_k}{1-\gamma}}, \quad (12)$$

and, therefore, maximum profits can be expressed as a function of prices and parameters:

$$\pi_F(z, w, r) = (1 - \gamma)((1 - \tau_y)z)^{\frac{1}{1-\gamma}} \left(\frac{\theta_l}{w}\right)^{\frac{\theta_l}{1-\gamma}} \left(\frac{\theta_k}{r}\right)^{\frac{\theta_k}{1-\gamma}}. \quad (13)$$

One less standard feature of the model is related to the presence of the informal sector. Again, some entrepreneurs in the informal sector will be better off hiring capital equal to b , just low enough to avoid detection. The threshold z_c is defined so that all informal entrepreneurs with $z < z_c$ operate unconstrained with $k_I(z, w, r) < b$, while all those with $z \geq z_c$ operate constrained, i.e., $k_I(z, w, r) = b$. To illustrate this, consider an entrepreneur z in the informal sector for whom $k_I(z, w, r) < b$. The optimal capital demand for this entrepreneur will be identical to that given by equation (11) but replacing $\tau_y = 0$. Note that the monotonicity of this demand function with respect to z ensures the existence of the threshold z_c , as defined above. Hence, the optimal informal profits are expressed in terms of prices and parameters by:

$$\pi_I(z; w, r) = \begin{cases} (1 - \gamma)z^{\frac{1}{1-\gamma}} \left(\frac{\theta_l}{w}\right)^{\frac{\theta_l}{1-\gamma}} \left(\frac{\theta_k}{r}\right)^{\frac{\theta_k}{1-\gamma}}, & k_I(z, w, r) < b \\ (1 - \theta_l)z^{\frac{1}{1-\theta_l}} \left(\frac{\theta_l}{w}\right)^{\frac{\theta_l}{1-\theta_l}} b^{\frac{\theta_k}{1-\theta_l}} - rb, & k_I(z, w, r) = b \end{cases}. \quad (14)$$

With this in hand, we can provide a more formal argument for the existence of threshold z_2 . Note first that both the informal and formal entrepreneurs' profits are increasing convex functions of z (the exponent of the entrepreneurial ability is $\frac{1}{1-\gamma} > 1$). Second, note that by comparing equation (13) and the top case of equation (14), it is clear that at least for all $z \leq z_c$, informal profits are higher than formal profits. This is trivially true for other entrepreneurs to the right of z_c . Finally, note that $\frac{1}{1-\gamma} > \frac{1}{1-\theta_l}$, hence as $z \rightarrow \infty$, $\pi_F(z; w, r) > \pi_I(z; w, r)$. This implies the existence of a threshold z_2 such that $\pi_I(z_2; w, r) = \pi_F(z_2; w, r)$, provided that $b > 0$ and $\tau_y > 0$ is not too large.

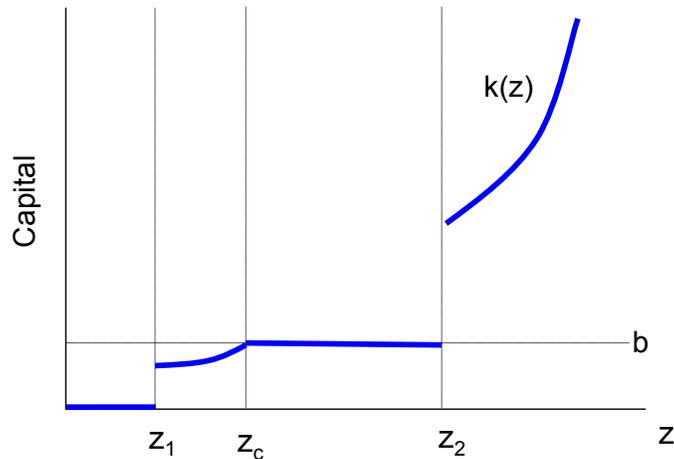
Furthermore, notice that in order to have a steady-state equilibrium where both the informal and formal sectors are positive, $b > 0$ must not be too small and $\tau_y > 0$ must not be too large. When τ_y is large, the profits in the formal sector remain below the profits in the informal sector across the full range of existing entrepreneurial abilities $[z, \bar{z}]$. This means all entrepreneurs would become informal. For example, in the case $\tau_y = 1$, formal sector profits are zero for all $z \in [z, \bar{z}]$ and therefore when $b > 0$, all entrepreneurs are informal. Similarly, when $b = 0$, profits in the informal sector are zero regardless of ability level and all entrepreneurs become formal if $\tau_y < 1$. For intermediate cases, the informal

sector size will be positive provided that $b > 0$ is not too small, otherwise profits in the informal sector would remain low for all agents when compared to both employee earnings or formal profits. Finally, note that if in equilibrium both the informal and the formal sectors are positive, it must follow that not all of the informal entrepreneurs are unconstrained, otherwise the threshold z_2 would not exist.

4.2. Capital-labor ratios

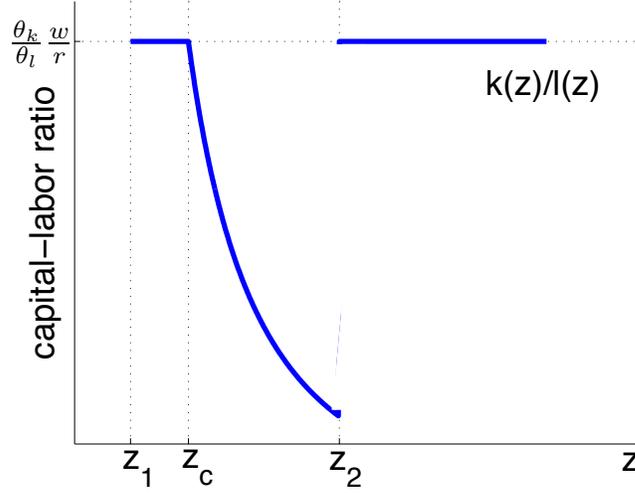
Given the nature of tax enforcement, the capital-labor ratios of constrained informal entrepreneurs are distorted in equilibrium. First, notice that the capital demand schedule has a discontinuity (see Figure 3). As noted above, if both sectors are positive, it must follow that the more capable informal entrepreneurs are constrained. To illustrate this, consider the entrepreneur z_2 , who is indifferent between the two sectors. If informal, she would hire b capital; if formal, she would hire an amount strictly greater than b . To see why, note that optimal decisions of formal entrepreneur z_2 are the same as those of a hypothetical entrepreneur that operates unconstrained and pays no taxes; this entrepreneur is $z_h = (1 - \tau_y)z_2$. Entrepreneur z_2 hires capital strictly greater than b as long as $z_h > z_c$, and this inequality holds, because, as shown in the bottom case of equation (14), $\pi_I(z; w, r)$ is strictly increasing.

Figure 3: Capital profile



Notice that, as a corollary of this property, the labor demand schedule is strictly increasing with respect to z , in equilibrium. Finally, notice that the discontinuity in the capital schedule translates into an informal sector that appears less capital intensive. The capital-labor ratio is smaller for constrained establishments (see Figure 4), as is the capital-output ratio.

Figure 4: Capital-labor ratios



5. Calibration

My calibration strategy is different from that followed in works that focus on developed economies, such as Restuccia & Rogerson (2008) and Guner *et al.* (2008). These papers assume that the US has small distortions and the distortion-free scenario in the model is used as a benchmark to study how deviations affect equilibrium variables. In this study, however, the distortions characteristic of the Mexican economy prevent me from taking the same approach.

The parameters to calibrate are the tax rate paid by the formal sector, τ_y , the technology parameters, θ_k , γ , and depreciation δ , the discount rate β , the enforcement policy parameter, and the entrepreneurial ability distribution parameters. The enforcement policy used as a benchmark is described in equation (2), where the probability of detection depends on capital; therefore, only one parameter needs to be calibrated, i.e., b . Entrepreneurial ability is assumed to follow a truncated Pareto distribution with parameters z_{min} , z_{max} , and s . More specifically, I assume that z has CDF:

$$G(z) = \frac{1 - \left(\frac{z_{min}}{z}\right)^s}{1 - \left(\frac{z_{min}}{z_{max}}\right)^s},$$

where $s > 0$ is the shape parameter and $z \in [z_{min}, z_{max}]$, with $0 < z_{min} < z_{max}$. I make this choice for two main reasons. The first is that the firm-size distribution in the US has been reported to be well-described by a Pareto distribution (Axtell, 2001). The second is more practical: a truncated Pareto is fully defined on an interval that I can link directly to the model objects \underline{z} and \bar{z} .

Next, I continue with the value of the parameters for which I am able to provide an independent calibration: the exponent of capital in the production function θ_k , the depreciation rate δ , and the tax rate τ_y . I choose $\theta_k = 0.33$ for the following reasons: first, because it is the standard value used by a number of studies focusing on Mexico; for example, Bergoeing *et al.* (2001) use $\theta_k = 0.33$ to compute

TFP series for Mexico, Solimano *et al.* (2005) perform growth accounting using $\theta_k = 0.35$ for several Latin American economies including Mexico, and Restuccia (2008) uses a value of $\theta_k = .28$ for a production function with decreasing returns to scale; second, this value is consistent with the estimates of Garcia-Verdu (2005). I choose $\delta = .05$ following Solimano *et al.* (2005) and Bergoeing *et al.* (2001), who use the same value for the depreciation rate in articles using Mexican data; additionally, as I will explain below, this value is roughly consistent with data on investment and on consumption of fixed capital in Mexico. To calibrate the tax rate, an assessment of the fiscal burden on formal firms is needed. I take the total tax revenue from the formal sector, which amounted to 11% of GDP in 2008; then, I make an assessment of the value added associated to these firms, this value amounts to 44% of GDP.⁸ The ratio of these two numbers is $\tau_y = 11/44 = 0.25$.

A discussion of the advantages of using an output tax is needed. In principle, an output tax makes the model much more tractable. More importantly, this choice is consistent with a common notion of doing business in Mexico, which is that entering the formal sector implies facing all taxes, while staying informal implies facing none. One way to corroborate this notion is by looking at the revenue collected from corporations that generate most of the formal sector's output. An examination of official tax revenue statistics indicates that corporations pay 92% of the taxes collected by the SAT (Mexico's tax authority); these include VAT, corporate taxes, and income tax withheld from employees. The key observation is that there are not many differences in the tax burden on corporations when we look at different taxes: for example, corporations pay 97% of VAT revenue and also pay 94% of total income tax (this includes corporate and personal income). The same point can be made from a different perspective; Fuentes-Castro *et al.* (2011) calculate tax evasion among small businesses in 2008 at 96%. This means that only 4% of potential revenue is collected from these firms. There are no differences between income and VAT, because small businesses face a single sales tax that substitutes for both. Thus, I believe that summarizing all taxes into a single grabbing-hand type of tax is a reasonable assumption.

Given the choices of θ_k , δ , and τ_y , I proceed to calibrate the remaining parameters in the model. In order to do this, I solve for the equilibrium as a function of these parameters and set their value so that the model replicates a number of features of the Mexican economy. These features include various moments of the size distribution of employment, the size of the informal sector, and the aggregate capital-output ratio. The data for the moments of the size distribution of employment and the size of the informal sector were described in Section 2. The data for the capital-output ratio have not been described previously. An assessment of the magnitude of the capital-output ratio is needed. To do this, I use data on the consumption of fixed capital (as a proportion of GNI) from Indicators (2005) and take the average since 1980 (which I call d). This average is around 10%. The model

⁸The value added by the formal sector consists of the value added by financial corporations (3.5%), the value added by public non-oil, non-financial corporations (2%), the value added by the general government (8%), and the value added by formal private non-financial corporations and quasi-corporations (30.5%).

counterpart of d is $\delta K/Y$. Since δ and d are known, I solve for K/Y from this equation and obtain $K/Y = d/\delta = .10/.05 = 2$.

This value of the capital-output ratio is close to that found in two independent studies that estimate capital stock in Mexico. Hofman (2000) performs a disaggregated estimation by type. The implied capital-output ratio in his work is around 1.7. Restuccia (2008) uses data from the Penn World Tables to estimate the capital-output ratios of a number of Latin American countries. He finds a value of around 1.9 for Mexico.⁹

Note also that the discount rate β cannot be calibrated in the usual way, which would be to obtain the value of r from the FOC of the firms and then use this value in the Euler equation to determine β . In principle, one would think that the FOC of formal establishments could be used to find the value of r ; however, to do so I would need an estimation of the capital-output ratio in the formal sector, which is not available. Mexico's national accounts include the informal sector, and since I used national account data to estimate the K/Y ratio, I think of it as a ratio that includes the capital and the output from both sectors.

5.1. Joint Calibration

The remaining parameters are γ , z_{min} , z_{max} , s , b , and β . The choice of z_{min} is arbitrary. This is because all individuals with entrepreneurial ability below the z_1 threshold become identical employees (i.e., their ability is transformed into 1 unit of labor). Therefore, what matters in equilibrium is the mass of individuals to the left of z_1 . Once z_{min} is set, this mass is fully determined by the parameters that describe the distribution of entrepreneurial abilities.

The rest of the parameter values are calibrated jointly by replicating moments of the plant-size distribution, the capital-output ratio, and the size of the informal sector. In the model there is a weakly monotonic equilibrium relationship between the size of a productive unit in terms of the labor employed and its entrepreneurial ability. I take advantage of this feature to calibrate the parameters of the entrepreneurial-ability distribution, using the employment distribution of establishments across size categories, as well as information on the average size of the units in each category.¹⁰ The moments targeted are:

1. the average size of establishments in the economy,
2. the average size of establishments with more than 100 workers,

⁹As a check, I used the capital accumulation equation in the balanced growth path combined with data on investment and capital consumption to jointly calculate the capital-output ratio and the depreciation rate. Specifically, I take yearly data on gross fixed capital formation (%GDP) and the consumption of fixed capital (%GNI) from Indicators (2005), and take averages since 1980; then I solve the following system of equations: (1): $(1+n)(1+g)(K/Y) = (1-\delta)(K/Y) + (I/Y)$, and (2) $\delta(K/Y) = d$. Where n and g are the annual growth rates of population and technology respectively, and $d = 0.105$. I set $n = 0.02$ and $g = 0.025$, again using data since 1980. The two unknowns are (K/Y) and δ . I obtain $K/Y = 1.9$ and $\delta = .059$.

¹⁰This procedure is close to those in Guner, Ventura, & Xu (2008) and Rubini (2009)

3. the proportion of workers in establishments with more than 100 workers,
4. the size of the informal sector, and
5. the capital-output ratio

Note that by targeting the first three moments, I can also match their complements: the share of workers and the average size of establishments with 100 workers or fewer. How well I can match similar moments for more disaggregated size categories will depend solely on the structure imposed by the Pareto distribution. As I show below, the calibration yields parameter values that replicate the data fairly well, even in highly disaggregated size categories, despite the fact that I do not target such moments. I present a summary of the calibration targets in Table 3.

Table 3: Calibration targets summary

Parameter	Target	Source
θ_k	capital share	Gollin (2002); Garcia-Verdu (2005)
δ	gross capital formation; consumption of fixed capital;	WDI, Solimano <i>et al.</i> (2005) and Bergoeing <i>et al.</i> (2001)
τ_y	tax revenue from private formal sector	own assessment
z_{min}	arbitrary	-
γ β	moments of distribution;	
z_{max} s b	size of informal sector; capital-output ratio	joint calibration

5.2. Calibration properties

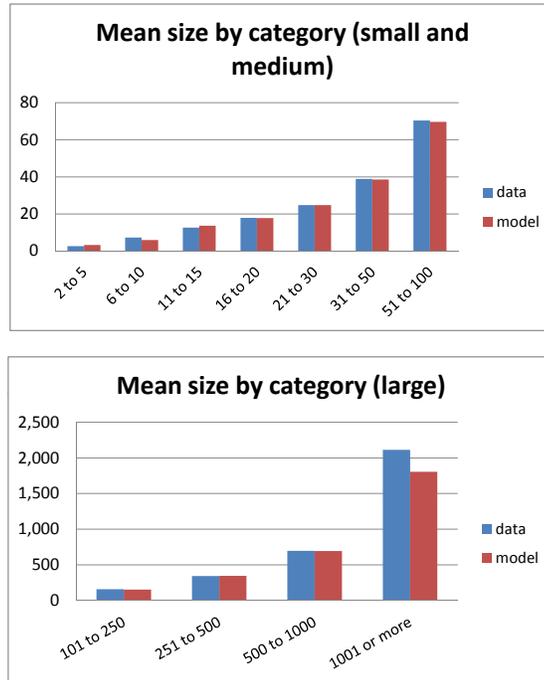
The targeted moments are well matched as can be confirmed in Table 4, which presents data and model values.

Table 4: Calibration targets

Targeted Variables	Data	Model
K/Y	2.0	2.1
mean size	5.5	5.8
informal size	0.45	0.45
mean size by employment size category:		
more than 100	362	360
worker share by employment size category:		
more than 100	.30	.28

Perhaps more interesting is the fact that the calibration yields parameters that replicate well a number of moments that were not targeted explicitly. In Figure 5, the model is shown to replicate the mean size for a number of highly disaggregated size categories.

Figure 5: Calibration properties: Non-targeted moments



The calibrated parameter values are presented in Table 5. Note that the value of γ is 0.76, relatively low compared to that found in studies focused on the United States. In particular, Atkeson & Kehoe (2005) estimate a value of 0.85 for US manufactures. γ controls the returns to scale at the establishment level. The closer γ is to 1, the lower the degree of decreasing returns and the more efficient it is to concentrate production in large establishments. In the limiting case of $\gamma = 1$ (constant returns to scale), efficient output is reached by concentrating all resources in a single unit, i.e., the most productive one. The low value of γ I find implies that the distortion-free allocation for Mexico is to have more workers in small units than would be the case in countries where the degree of decreasing returns is smaller (i.e., a larger γ).

That the distortion-free allocation in Mexico is different than that in the US is not necessarily a bad result. It is not the thesis of this paper that the differences between Mexico and US distributions are due solely to tax enforcement differences. It could be argued that Mexico's distortion-free skewed

distribution is merely the result of its early stage of development. A number of authors have documented the steady rise in average firm size in the US during the 19th and 20th century (for a short bibliography, see Desmet & Parente, 2009). Furthermore, when one looks at the distribution of US of the past, at a point in time during which the US had the same GDP per capita as modern-day Mexico (around the 1930s), it is clear that it was not as highly concentrated in small establishments as it is in today's Mexico.¹¹

Finally, one more feature of the calibrated economy is that all informal entrepreneurs are constrained (i.e. that $z_c \approx z_1$), which is roughly consistent with the evidence in Figure 1.¹²

Parameter	Calibrated value
θ_k	.33
δ	.05
β	0.94
γ	0.76
z_{min}	1
z_{max}	13.05
s	4.25
b	10.50
τ_y	.25

6. Results of the baseline model

Once the model is calibrated to the Mexican economy, I can investigate the effects of incomplete enforcement policies. To do this, I use this calibrated economy as a benchmark and perform three exercises: one that focuses on short-run effects and two more that explore long-run effects. In these exercises, I introduce complete tax enforcement by making $b = 0$ in the model; this implies that all establishments would pay a uniform tax (τ_y).

As a first step, I look at the equilibrium in the first period, immediately after the new enforcement policy is introduced. In this period, the capital stock is the same as in the economy with incomplete enforcement, because accumulation has not occurred yet. Table 6 shows the value of aggregate variables in this context. Since capital and employment (employees+entrepreneurs) are no different from the benchmark economy, the only reason that output increases is that these resources are used more

¹¹Granovetter (1984) documented the fact that the proportion of employees in US manufacturing establishments with fewer than 20 employees was 10% in 1933, while the proportion of employees in Mexican manufacturing establishments with fewer than 15 workers was 37.5% in 2005. Notice that the size category is capped at a smaller size for Mexico than for the US, but the proportion allocated is still larger. Similarly, for the same size categories I find that for the retail and wholesale sectors, the figures are 63.8% and 44.4% for the US in 1939, and 72% and 48% for Mexico in 2005.

¹²This is clearer for construction, trade and services, which are the sectors where most of the informal entrepreneurs operate.

efficiently. Not surprisingly, output increases proportionally with TFP. Also, notice that in this first period, wages decrease.

Table 6: Short-run effects:

First period after change in enforcement	
Variable	Value under complete enforcement relative to benchmark
Y	1.044
τ_y	1
K	1
TFP	1.044
L	1.120
μ	2.104
r	1.130
w	0.829

The gains in TFP are associated with the removal of the distortions present under incomplete tax enforcement. The effects of incomplete enforcement include distorting three margins: optimal occupational choices, allocation of resources across establishments, and the capital-labor ratios of informal establishments. The first two distortions occur across establishments, while the last occurs within establishments. Incomplete tax enforcement distorts occupational choices because it makes entrepreneurship more attractive; furthermore it distorts the allocation of resources directly because it makes it possible to have some establishments paying taxes and some others not; finally, it distorts the capital-labor ratio of a group of informal establishments, because this is the optimal response to a probability of detection that increases with capital.

The short-run gains in TFP respond to the elimination of the distortions mentioned above. Note first, that the employee/entrepreneur threshold z_1 increases because a group of low-ability individuals no longer find it attractive to be entrepreneurs, so the average ability in the economy (μ in Table 6) improves;¹³ second, note that since every establishment pays the same tax rate, marginal products are equalized and the allocation of resources improves; and finally, note that since the probability of detection plays no important role under full enforcement, capital-labor ratios become undistorted.

In Table 6, it can be confirmed that, as a result of the change in threshold z_1 , the proportion of employees in the economy increases by 12% and the average ability by 110%. Note that consistent with this, wages decline to a level that is 0.83 of the benchmark level. Also note that the rental rate of capital increases by 13% in this first period.

An interesting aspect of full enforcement is that occupational choices and the allocation of resources across establishments are the same as those in a version of the model with $\tau_y = 0$. In other words, once every establishment pays a uniform tax, the value of the tax rate does not affect either

¹³This also entails a reduction in the mass of firms which reduces output, see Section 7 for a decomposition of the gains.

occupational choices or the allocation of resources across establishments. To gain some intuition, consider the relative labor demands for any two establishments, z and z' , facing the same tax rate $\tau \geq 0$; this is given by:

$$\frac{l_F(z', w, r; \tau)}{l_F(z, w, r; \tau)} = \frac{((1 - \tau)z')^{\frac{1}{1-\gamma}} \phi(w, r, \theta_k, \theta_l)}{((1 - \tau)z)^{\frac{1}{1-\gamma}} \phi(w, r, \theta_k, \theta_l)} = \left(\frac{z'}{z}\right)^{\frac{1}{1-\gamma}}, \quad (15)$$

which is independent of the tax rate, and depends only on relative productivity and γ . Now consider occupational choices under complete tax enforcement and the same tax rate; these choices are fully described by the employee/entrepreneur threshold z_1 , which is given by:

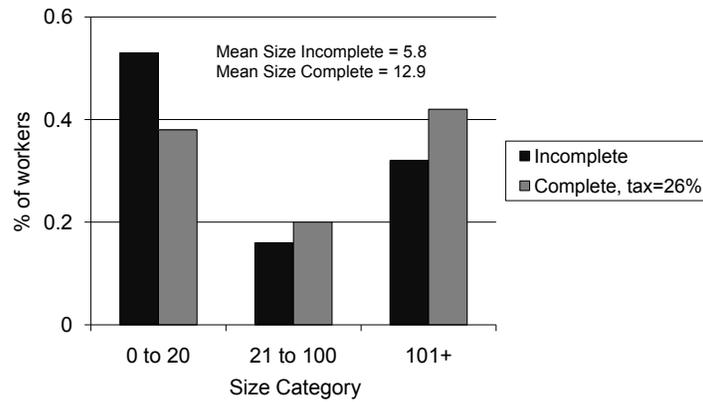
$$z_1 = \left[\frac{\theta_l \int_{z_1}^{\bar{z}} z^{\frac{1}{1-\gamma}} g(z) dz}{1 - \gamma G(z_1)} \right]^{1-\gamma} \quad (16)$$

This threshold is also independent of the tax rate. Now consider two economies under full enforcement that differ only in the tax rate $\tau_1 \neq \tau_2$. Since the total mass of employees (i.e., occupational choices) and relative labor demands are unaffected by the tax rate, it must be that: $l_F(z, w_1, r; \tau_1) = l_F(z, w_2, r; \tau_2), \forall z$, with w_1 and w_2 corresponding to the steady-state wages. Since this holds for any pair of tax rates, it also holds for the case of $\tau = 0$.

What this means is that economies with complete enforcement resemble the economy with $\tau_y = 0$ in two aspects: occupational choices, and allocation of resources across establishments. In this sense, occupational choices and resource allocation are “efficient” under full enforcement. The aspect in which these economies differ is in the amount of capital accumulation, and therefore, prices and output.

The effects on the efficient use of resources are also captured by the labor reallocation across establishment-size categories. This is presented in Figure 6. The improved enforcement policy reduces the allocation of resources to small establishments and increases the allocation to medium-sized and large establishments. As a consequence, the mean size is almost doubled.

Figure 6: Labor reallocation induced by complete enforcement



The next step is to look at long-run effects of the introduction of complete tax enforcement. This exercise is presented in Table 7. Note that there are no more TFP gains beyond those that occur in the short run, because occupational choices are not affected by the allocation of resources either. However, capital accumulation increases by 20% relative to the benchmark, because capital is more productive relative to the case with incomplete enforcement. Finally, note that wages increase to 0.88 (still below the benchmark level), because workers have more capital to produce with.

The change in incentives to accumulate capital brought by complete enforcement is reflected in the 13% increase in the rental rate r in Table 6. This price change is merely a consequence of the rise in the marginal productivity of capital, which is explained by the removal of the distortions faced by informal entrepreneurs and the better allocation of resources in the economy. Remember that an important group of establishments that used to be informal remain in operation after the enforcement change; these were using capital $k(z) = b$ and exhibited distorted capital-labor ratios; now these establishments demand capital without restriction.

Variable	Value under complete enforcement relative to benchmark
Y	1.109
τ_y	1
K	1.200
TFP	1.044
L	1.120
μ	2.104
w	0.88

In the final exercise I perform in this section, I aim to address the effects of incomplete enforce-

ment, taking into account the argument put forth in Lewis (2004): specifically, that the combination of incomplete enforcement and big government leads to high taxes being levied on a small subset of firms. From this perspective, the relevant exercise would involve increasing enforcement levels, while decreasing the tax rate to leave revenue unchanged. By increasing enforcement levels, the tax base is broadened and, therefore, a lower tax rate will result in the same revenue as before. This will allow me to capture the costs of incomplete enforcement associated with the need for higher taxes.

In Table 8, I present the effects of such an exercise on the steady-state values. If Mexico's present enforcement policy were complete, it would be able to reduce taxes to 52% of the current levels. This tax reduction gives an extra boost to the economy. Overall, output would increase 19%. The table shows that this increase would be driven mainly by a 50% increase in capital accumulation, while TFP would play a smaller role, with an increase of 4.4% which occurs fully in the short run. In the long run, once accumulation of capital takes place, labor productivity is increased and the wage rate rises to a level that is 10% higher than the benchmark.

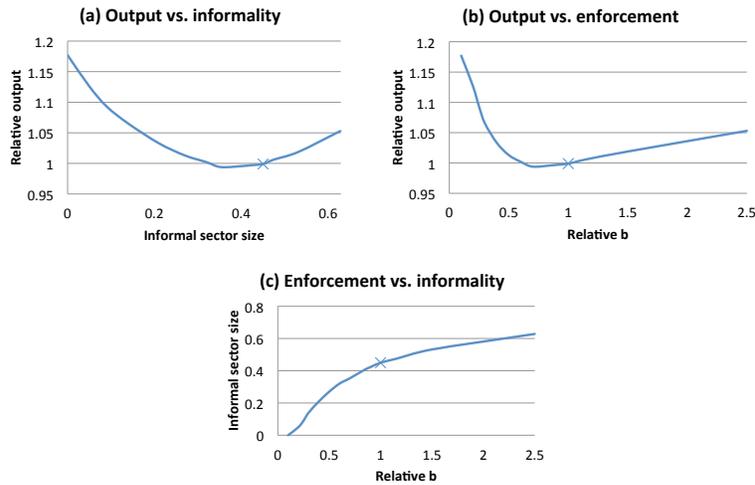
Table 8: Comparison across steady states

Variable	Value under complete enforcement relative to benchmark
Y	1.193
τ_y	0.52
K	1.498
TFP	1.044
L	1.120
μ	2.104
w	1.10

6.1. Partial enforcement

Full enforcement corresponds to an extreme policy change; one interesting question is, what are the gains from partial improvements in enforcement? After all, many developed countries do not enforce taxes fully.

Figure 7: Partial enforcement



In Figure 7, partial enforcement is achieved by moving the parameter b over a range that goes from 0.1 to 2.5 times its calibrated value. Panel (a) of Figure 7 plots output vs. the size of the informal sector; Panel (b) output vs. enforcement; and Panel (c) enforcement vs. the size of the informal sector. In the figure, output is measured relative to the benchmark and enforcement is measured by \tilde{b} , which represents the value of the enforcement parameter relative to the calibrated value. Thus, in the benchmark $\tilde{b} = 1$. Furthermore, the benchmark values in each panel are indicated by a cross.

A few important points arise when analyzing this picture. First, note in panels (a) and (b) that the gains from enforcement are nonlinear; in fact, starting at the benchmark, a marginal increase in enforcement (i.e., a reduction of b or, equivalently, a reduction in the informal sector) reduces output; second, it is not until \tilde{b} reaches 0.5 that we can see output gains; third, as we further reduce \tilde{b} (from 0.5 to 0.25), output increases rapidly; finally, note that achieving the enforcement levels of a developed country (i.e., 10% of informality) generates around 50% of the gains obtained under full enforcement. One reason behind this nonlinear result is the relationship between \tilde{b} and the size of the informal sector; note that the slope in panel (c) around $\tilde{b} = 1$ is flatter than the slope around $\tilde{b} = 0.5$; thus, improvements in enforcement around the benchmark create relatively small reductions in the informal sector.

One important lesson from Figure 7 is that making enforcement worse also increases output. This occurs as marginal firms (near z_1) enjoy a lower tax burden, while the burden on formal firms stays constant. This is consistent with the common idea that the informal sector allows firms to operate more efficiently. Figure 7 helps us to see the effects of increasing informality as a tradeoff: on the one hand, there is a positive force that originates from “lower” taxes; on the other, there is negative force that stems from resource misallocation, and from distorted occupational and capital choices. When informality is high, the positive force dominates, because as informality increases, taxes are lower

for a set of large and very productive establishments; in contrast, when informality is low, increasing informality reduces taxes for a set of relatively small, and low-productive establishments. For middle levels of informality, these two forces tend to offset each other and a U-shape relationship between informality and output emerges. In fact, Mexico, with an informal sector size equal to 45%, is close to the worst possible output level, which is associated with 35% informality.

7. Decomposing the gains from full enforcement

To decompose the gains from full tax enforcement, I perform a number of quantitative exercises asking what would happen to the gains of eliminating distortions if I shut down or add specific features to the model. I start by analyzing a static endowment economy version of the model and then take out or add features, following the guidelines of five leading papers in the literature on resource misallocation across plants. I classify these papers into three groups: a) Restuccia & Rogerson (2008), which assumes free entry to determine the mass of firms in equilibrium; b) Gollin (1995) and Guner *et al.* (2008), which emphasize occupational choices; and c) Hsieh & Klenow (2007) and Jones (2011), which focus on linkages and complementarities using models with monopolistic competition. The contribution of this exercise is twofold: first, it provides a way to decompose the gains from full enforcement; second, it clarifies the differences among the aforementioned papers by comparing the results that alternative models provide for the same change in policy.

7.1. Marginal product equalization

As a first step, I look at the effects of full enforcement in a static endowment economy version of the model where the supply of labor, the supply of capital, the mass of firms, and the average ability of entrepreneurs are all fixed. The results are presented in Table 9.

Table 9: Pure misallocation effects

Var.	Bench.	Full enforcement.
Y	1	1.023
TFP	1	1.023
w	1	0.910
r	1	1.107

The economy in Table 9 operates with exactly the same resources as the benchmark economy. Since all resources are fixed, the only way to increase output is through a better allocation of resources. With the introduction of full enforcement, marginal products are equalized across establishments; therefore, this exercise isolates the pure effect of resource misallocation across plants.

Notice also that the introduction of full enforcement has an asymmetric impact on factor returns: wages go down, but the rental rate of capital increases. This is intuitive: on the one hand, higher taxes

affect both factor prices; on the other, the elimination of capital constraints increases both the demand for capital and r .

Since marginal products are equalized under full enforcement, this leads to an important issue regarding the decomposition of output gains. When there is no informal sector and marginal products are equalized, we can express aggregate output as a function of four main aggregates:¹⁴

$$Y = N^{1-\gamma} \mu^{1-\gamma} K^{\theta_k} L^{\theta_l}, \quad (17)$$

where K is aggregate capital, L is aggregate labor (employees), N is the mass of firms, and μ is average entrepreneurial ability. Given the 2.3% gain from a better allocation of resources, the remaining gains associated with full enforcement must be generated by changes in the amount of the productive factors in equation 17.

7.2. The mass of firms (adding free entry)

To investigate the importance of distortions in the mass of firms, I add a free-entry condition but keep the assumptions of a fixed endowment of labor, a fixed endowment of capital, and a fixed average ability (i.e., I shut down occupational choices and investments). This modification brings the model closer that used by Restuccia & Rogerson (2008) (group “a”).

Consider a potential entrepreneur that is evaluating whether to enter or not. This agent takes into account the expected profits of incumbent entrepreneurs. Thus, the value of entry is described by:

$$W_e = \int_{z_1}^{\bar{z}} \pi(z) \hat{g}(z) dz - c_e,$$

where $\hat{g}(z) = g(z|z > z_1)/(1 - G(z_1))$, c_e is an entry cost, $\pi(z)$ are profits, and z_1 is given by $w = \pi_M(z_1; w, r)$ (from the benchmark, in the baseline model). Now let N be the mass of firms and E the endowment of labor; the labor market clearing condition is as follows:

$$E = N \int_{z_1}^{\bar{z}} l(z) \hat{g}(z) dz,$$

where $E = 1 - G(z_1)$.

Notice that in this economy, average entrepreneurial ability is constant, independent of N , and is equal to $\mu = \int_{z_1}^{\bar{z}} z^{\frac{1}{1-\gamma}} \hat{g}(z) dz$. Given c_e , I can find equilibrium allocations by solving the free-entry and the labor market clearing conditions for w and N . Column A in Table 10 shows the results.¹⁵

¹⁴See the appendix

¹⁵To obtain the value of c_e , I use the fact that W_e must be zero in equilibrium (the free-entry condition) and set it equal to the value of $\int_{z_1}^{\bar{z}} \pi(z) \hat{g}(z) dz$ in the benchmark.

Table 10: Effects on factor levels

Var.	Bench.	A. Free entry	Occupational choices	
			B. Short-run	C. Long-run
Y	1	0.925	1.044	1.109
K	1	1	1	1.200
L	1	1	1.120	1.120
N	1	0.662	0.424	0.424
μ	1	1	2.104	2.104
TFP	1	0.925	1.044	1.044

The introduction of complete enforcement reduces W_e because the expected value of incumbent entrepreneurs is hit by higher taxes; therefore, the mass of firms is negatively affected. In equation 17, the mass of firms is an important component of aggregate output, thus output decreases. We can perform the accounting of output losses using equation 17. Letting Y_1 be the value of output in Table 9 and Y the corresponding value in Table 10, the effect of full enforcement under this exercise is: $\frac{Y}{Y_1} = 0.662^{0.24} \approx \frac{0.925}{1.023} \approx 0.904$. Thus, adding firm entry reduces output by 10% (see also Table 11).

7.3. Occupational choices

Now consider an economy with an occupational-choice rather than a free-entry condition, but keep the assumption of a fixed capital stock. This brings the model closer to Gollin (1995) and Guner *et al.* (2008) (group “b”), and is also our baseline model in the short run. Adding occupational choices will bring good news, because it will allow the economy to increase average ability and the supply of labor, while reducing entry.

Notice first that the introduction of full enforcement pushes z_1 to the right. As a result, the mass of firms $N = 1 - G(z_1)$ is reduced; however, the average ability μ and the supply of labor $F(z_1)$ both increase. This exercise is identical to that presented in Table 6 of Section 6, and the results are repeated in Columns B and C of Table 10. Notice that the negative effect on the mass of firms, which falls to 42% of the value in the benchmark, is reversed with a substantial increase in average ability, which more than doubles. This can be seen with the help of equation 17. The contribution of these two factors to output is: $(.424 * 2.104)^{(1-\gamma)} \approx 0.972$, which is quite close to the value in the benchmark (of 1).

A further advantage of occupational choices is that aggregate labor increases with z_1 . Since labor increases by 12%, its contribution to output is now: $1.12^{\theta_l} \approx 1.049$. In summary, adding occupational choices increases output by 4% relative to the benchmark. Furthermore, the pure effect of occupational choices is to increase output by 12% with respect to the free-entry scenario (see also Table 11).¹⁶

¹⁶The accounting for this number is $1.12 \approx (0.972 * 1.049)/0.904$

7.4. Capital accumulation

Finally, when capital is allowed to accumulate, it increases by 20% relative to the benchmark and the contribution to output is now $1.20^{\theta_k} \approx 1.062$. In summary, when all factors are allowed to change, output increases by a factor of $0.972 * 1.049 * 1.062 \approx 1.083$ with respect to output in Table 9, and by $1.109 \approx 1.082 + 0.023$ with respect to the output in the benchmark.

Table 11 presents a summary of the decomposition of the effects of full enforcement. There are four main results: a) the pure effect of misallocation is non-negligible (+2); b) with a free-entry condition, the negative effect of higher taxes on the value of firms dominates and output decreases (-10); c) occupational choices reverse the effect on the reduction in the mass of firms by increasing average ability and the labor supply (+12); and d) a sizable proportion of the gains is provided by capital accumulation, which suggests that capital constraints on informal firms constitute an important distortion (+7).

Table 11: Decomposing gains of perfect enforcement

Var.	A = MP equalization	B = A + free entry	C = A + occupational choice	D = C + capital accumulation	Total
Y	1.02	0.92	1.04	1.11	
(ΔY)	+2	-10	+12	+7	+11
TFP	1.02	0.91	1.04	1.04	
ΔTFP	+2	-10	+12	0	+4

8. Monopolistic competition

As explained in Jones (2011), introducing monopolistic competition amplifies the effects of resource misallocation through complementarity across varieties. In the problem considered in this paper, adding monopolistic competition will increase the gains from complete enforcement because distortions in one establishment affect relative prices across the full range of operating producers. In a traditional model of this type, a multiplier is generated as distortions in one producer affect the price index and aggregate demand. In turn, changes in aggregate demand affect the efficiency of individual producers, because they take into account this demand when they make their pricing decisions.

To clarify this idea, I outline the model with monopolistic competition and then establish a set of conditions on the parameters under which the equilibrium allocations in this model are the same as those in the model with perfect competition. This brings the model closer to Hsieh & Klenow (2007) and Jones (2011) (group “c”). I outline the model for a distortion-free economy, without loss of generality.

8.1. Final-good producer

There exists a representative firm producing a final good using a CRS technology. There is a continuum of intermediate input varieties (ω). The technology of the final-good producer is:

$$Y = \left(\int y(\omega)^\rho d\omega \right)^{1/\rho},$$

where $y(\omega)$ represents the quantity of variety ω used to produce Y . Profits are given by $\Pi_Y = P \left(\int y(\omega)^\rho d\omega \right)^{1/\rho} - \int p(\omega)y(\omega)d\omega$ and standard optimization arguments lead to the demand for each variety:

$$y(\omega) = \left(\frac{p(\omega)}{P} \right)^{-\sigma} \left(\frac{I}{P} \right), \quad (18)$$

where $\sigma = 1/(1 - \rho)$, $P = \left(\int p(\omega)^{1-\sigma} d\omega \right)^{1/(1-\sigma)}$ is the final-good price index, which I choose as the numeraire and normalize this to 1, and $I = \int p(\omega)y(\omega)d\omega$ is total expenditure on intermediate inputs (or aggregate demand).

8.2. Intermediate producers

Intermediate producers are heterogeneous in their productivity level, x , and have access to a Cobb-Douglas technology:

$$y(x) = xk^\alpha l^{1-\alpha}. \quad (19)$$

As in Melitz (2003), I assume that regardless of its productivity, each intermediate producer faces an individual demand curve. Profits are given by

$$p(x)y(x) - wl(x) - rk(x),$$

therefore, the problem for a producer x is to maximize profits subject to equations 18 and 19.

8.3. Household problem

I keep the feature of the model that includes a representative household populated by a continuum of members with abilities x and each member faces an occupational choice, as in the benchmark model. The household problem with monopolistic competition is only slightly different from that in the model with perfect competition, with one difference being in the budget constraint. In the case of monopolistic competition, we have a firm that produces the final good—which is owned by the household—and its profits must be part of the resources available to the household. It is found that, due to CRS, the profits for the final producer are zero in equilibrium.

A further difference is in the profits of the intermediate producers $x \in (\underline{x}, \bar{x})$, which are, in general, different from the profits of entrepreneurs $z \in (\underline{z}, \bar{z})$. However, by choosing the right value of the parameters, an equivalence between these two models can be established.

8.4. Equivalency between perfect- and monopolistic-competition models

Equilibrium allocations in the model with perfect competition can be delivered by the model with monopolistic competition, provided certain conditions are placed on the parameters and the productivity levels. To demonstrate this, first let us note the following equivalency between establishment z 's problem in the model with perfect competition and that for producer x in the model with monopolistic competition. Producer x 's sales are:

$$p(x)y(x) = I^{(1/\sigma)}y(x)^{\frac{\sigma-1}{\sigma}} = I^{(1/\sigma)}x^{\frac{\sigma-1}{\sigma}}k^{\alpha\frac{\sigma-1}{\sigma}}l^{(1-\alpha)\frac{\sigma-1}{\sigma}}, \quad (20)$$

where the first equality arrives when using equation 18 to substitute for $p(x)$ and the second when we use equation 19 to substitute for $y(x)$. Conversely, remember that sales for a typical establishment z in the perfect competition model are simply given by $y(z) = zk^{\theta_k}l^{\theta_l}$. Furthermore, notice that the total cost in both problems is $(wl + rk)$ and, therefore, the optimal choices of entrepreneur z in the perfect competition model will be the same as the optimal choices of the intermediate producer x in the model with imperfect competition, provided the following conditions are met:¹⁷

1.

$$z = I^{(1/\sigma)}x^{\frac{\sigma-1}{\sigma}}, \forall z \in (\underline{z}, \bar{z})$$

2.

$$\theta_k = \alpha \frac{\sigma - 1}{\sigma} \text{ and}$$

3.

$$\theta_l = (1 - \alpha) \frac{\sigma - 1}{\sigma}$$

If the above three conditions hold, then the equilibrium allocations are exactly the same in both models. Moreover, given equilibrium allocations for the perfect competition model, we can always choose values of α , σ , and $x \in (\underline{x}, \bar{x})$, such that the above conditions hold. This equivalency is satisfied regardless of whether we are dealing with the incomplete or the complete enforcement versions. I use the aforementioned conditions to obtain α , σ , and $x \in (\underline{x}, \bar{x})$, such that the two problems referred to above are equivalent. I compute this equivalency for the incomplete enforcement versions.

8.5. Full enforcement with monopolistic competition

Now, I introduce complete enforcement in the model with monopolistic competition and look at the new steady state. Table 12 shows the results. Adding monopolistic competition triples the gains in measured TFP, increases capital accumulation by a factor of 1.33, and output by a factor of 1.8 with respect to perfect competition. The gains in TFP and output are not negligible. Restuccia (2008) studies the Latin American development problem and finds that in a model with human-capital accumulation,

¹⁷This point is also found in footnote 6 in Restuccia and Rogerson (2008).

TFP would have to increase 60% in order to eliminate the gap in income per capita between these countries and the US. Thus, introducing full enforcement could make a 20% (12/60) contribution to achieving this goal.

The mechanics of the multiplier are evident in condition 1 above. Notice that the productivity of each firm is now affected by aggregate demand (I). Since a distortion in one firm affects aggregate demand, it also affects the productivity of the remaining firms, which in turn feeds back into I , and so on.

Table 12: Complete enforcement with monopolistic competition

	Perfect competition		Monopolistic competition	
	Fixed tax rate	Revenue neutral	Fixed tax rate	Revenue neutral
Y	1.11	1.19	1.18	1.34
K	1.20	1.50	1.27	1.70
TFP	1.04	1.04	1.09	1.12

9. Summary of the output-loss decomposition

Table 13 presents a summary of the decomposition of output losses. It presents this decomposition for both the discontinuous distortion analyzed here and a continuous distortion analyzed in the Appendix (see also Section 9.1). For the case of the step function (discontinuous distortion), the total loss is 18% when the tax rate is fixed; this, in turn, can be decomposed into five sources: factor misallocation, firm entry, distortions of occupational choices, capital accumulation, and complementarities. Pure factor misallocation contributes 13% to the output loss, firm entry -56%, occupational choices 68%, capital accumulation 37%, and complementarities 38%. Notice that informality provides an incentive for firm entry, and therefore the increase in the mass of firms contributes -56% to the output loss. However, this is reversed by the contribution of occupational choices, which increase average ability and labor supply, whilst reducing entry. Thus, the net contribution of occupational choices and free entry is 12% (68%-56%).

Table 13: Summary of output loss decomposition (%)

		Discontinuous dist.	Continuous dist.	Relative	Related paper
Total loss with fixed revenue		-0.34	-0.24	0.70	
Total loss with fixed tax rate		-0.18	-0.12	0.66	
Contribution of	factor misallocation	0.13	0.11	0.57	
	firm entry	-0.56	-0.39	0.46	RR
	occupational choices	0.68	0.47	0.45	Gollin,
	capital accumulation	0.37	0.43	0.77	GVX
	complementarities	0.38	0.37	0.65	HK,Jones

9.1. Discussion

A natural question that arises from this analysis is, what is the cost of increasing enforcement? These costs have to be subtracted from the output gains reported in the previous section, in order to provide a sense of the net gain associated with the policy change in question.

Since the informal sector consists of many small businesses, better enforcement increases the number of taxpayers over which the authorities need to provide surveillance. In the benchmark, there are 22 times more informal establishments than formal establishments; in other words, informal establishments represent 95% of all entrepreneurs. According to official Mexican tax authority data, the office currently spends around 1 peso for every 100 pesos collected.¹⁸ We can use this information to obtain the enforcement cost per establishment in the model and then calculate the cost of full enforcement, assuming that cost per establishment is a constant and taking into account the increase in the number of formal establishments. I calculate that introducing full enforcement will increase total enforcement costs by a factor of 9.6, a huge increase; however, it represents a small fraction of gross revenue. Gross revenue increases to 25% of the model's GDP and enforcement costs to 1.3% of GDP. This means that most of the gains in enforcement remain after subtracting the extra monitoring costs.

Another source of concern is the importance of the assumption regarding the shape of the probability of detection. So far I have assumed that this probability has a discontinuity. In the Appendix, I investigate the sensitivity of my results to this assumption; I find that the bulk of the output losses driven by incomplete enforcement remain when a continuous distortion is considered. In particular, I use a linear function and find that output losses would be 75% of the losses in the baseline case.

In Table 13, the decomposition of output losses is also presented for the continuous distortion case. Notice that the bulk of the effects remain when a continuous distortion is considered. However, not all sources of distortions remain with the same intensity. For example, capital accumulation and complementarities remain as important as in the discontinuous distortion case, whilst factor misallocation and distortions of occupational choices, as well as firm entry seem less important.

10. Conclusion

I have analyzed the distortions associated with the presence of the informal sector and, in particular, of incomplete tax enforcement. The results indicate that these distortions can be quite considerable.

There are three main lessons in the paper. First, the distortions induced by the informal sector are as follows: 1) a misallocation of resources towards small and unproductive plants, as they engage in tax evasion; 2) a distortion in occupational choices as unproductive entrepreneurs are attracted to the market; and 3) a distortion in the capital use of informal establishments, as they reduce their scale to remain undetected. Additionally, when complementarities are considered, these distortions are amplified.

¹⁸This information is available at http://www.sat.gob.mx/sitio_internet/transparencia/51_8833.html

The second lesson is that for the change in policy analyzed here, models such as Restuccia & Rogerson (2008), in which the mass of firms is endogenous and the labor supply is exogenous, tend to obscure the effects of resource misallocation. In contrast, models with occupational choices, such as Gollin (1995) and Guner *et al.* (2008), allow the average ability and the supply of labor to move in the opposite direction to the mass of firms. This feature tends to reverse any change in the mass of firms, thus clarifying the effect of resource misallocation. Finally, models with monopolistic competition, such as those used by Hsieh & Klenow (2007) and Jones (2011), amplify the effects of misallocation. For the case of the policy change presented in this paper, TFP effects are multiplied by 3.

Finally, improving enforcement entails a tradeoff between more taxes and fewer distortions. When informality is relatively small, the burden of more taxes is low and the burden of distortions is significant; as a result, better enforcement increases output. The opposite happens when informality is high. Thus, *ceteris paribus*, there is an inverted-U relationship between output and the informal sector.

References

- Amaral, P. S., & Quintin, E.: A competitive model of the informal sector, *Journal of Monetary Economics*, 53(7), (2006), 1541–1553.
- Atkeson, A., & Kehoe, P. J.: Modeling and measuring organization capital, *Journal of Political Economy*, 113(5), (2005), 1026–1053.
- Axtell, R.: Zipf distribution of us firm sizes, *Science*, 293(5536), (2001), 1818–1820.
- Bergoeing, R., Kehoe, P., Kehoe, T., & Soto, R.: A decade lost and found: Mexico and Chile in the 1980s, *NBER Working Paper*.
- Bigio, S., & Zilberman, E.: Optimal self-employment income tax enforcement, *Journal of Public Economics*, 95(9), (2011), 1021–1035.
- Buera, F., Kaboski, J., & Shin, Y.: Finance and development: A tale of two sectors, *The American Economic Review*, 101(5), (2011), 1964–2002.
- Dabla-Norris, E., Gradstein, M., & Inchauste, G.: What causes firms to hide output? the determinants of informality, *Journal of Development Economics*, 85(1-2), (2008), 1–27.
- de Paula, A., & Scheinkman, J.: The Informal Sector, Third Version, *PIER Working Paper Archive*.
- de Soto, H.: *The Other Path: The Invisible Revolution in the Third World*, I B Tauris & Co Ltd, 1989.
- de V. Cavalcanti, T. V., & Antunes, A. R.: Start up costs, limited enforcement, and the hidden economy, *European Economic Review*, 51, (2007), 203–224.

- Desmet, K., & Parente, S.: The evolution of markets and the revolution of industry: A quantitative model of England's development, 1300-2000, CEPR Discussion Papers 7290, C.E.P.R. Discussion Papers, 2009.
- Di-Giannatale, G. R. . R. S., S.: Estimating the effects of formality on Mexican informal microfirms. a joint multivariate approach.
- Dixit, A., & Stiglitz, J.: Monopolistic competition and optimum product diversity, *The American Economic Review*, (pp. 297–308).
- Farrell, D.: The hidden dangers of the informal economy, *McKinsey Quarterly*, (pp. 26–37).
- Fortin, B., Marceau, N., & Savard, L.: Taxation, wage controls and the informal sector, *Journal of Public Economics*, 66(2), (1997), 293–312.
- Fuentes-Castro, H., Zamudio, A., Barajas, S., & Brown, A.: Estudio de evasión fiscal en el régimen de pequeños contribuyentes, Tech. rep., Centro de Estudios Estratégicos (ITESM), 2011.
- García-Verdu, R.: Factor shares from household survey data, Working Papers 2005-05, Banco de México, 2005.
- Gollin, D.: Do taxes on large firms impede growth? evidence from Ghana, Bulletin 7488, University of Minnesota, Economic Development Center, 1995.
- Gollin, D.: Getting income shares right, *Journal of Political Economy*, 110(2), (2002), 458–474.
- Granovetter, M.: Small is beautiful: labor markets and establishment size, *American Sociological Review*, 49(3), (1984), 323–334.
- Guner, N., Ventura, G., & Xu, Y.: Macroeconomic implications of size-dependent policies, *Review of Economic Dynamics*, 11(4), (2008), 721 – 744.
- Herrendorf, B., & Teixeira, A.: Barriers to entry and development, *Forthcoming: International Economic Review*.
- Hofman, A.: Standardised capital stock estimates in Latin America: a 1950-94 update, *Cambridge Journal of Economics*, 24(1), (2000), 45–86.
- Hsieh, C.-T., & Klenow, P. J.: Misallocation and manufacturing TFP in China and India, Working Paper 13290, National Bureau of Economic Research, 2007.
- Ihrig, J., & Moe, K.: Tax policies and informal employment: the Asian experience, *Asian Economic Journal*, 15(4), (2001), 369–383.

- Indicators, W. D.: World Development Indicators 2005 CD-ROM.
- Jones, C.: Intermediate goods and weak links in the theory of economic development, *American Economic Journal: Macroeconomics*, 3(2), (2011), 1–28.
- Kaplan, D., Piedra, E., & Seira, E.: Entry regulation and business start-ups: evidence from Mexico, *World Bank Policy Research Working Paper*, (4322).
- La Porta, R., & Shleifer, A.: The Unofficial Economy and Economic Development, *SSRN eLibrary*.
- Levy, S.: *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*, Brookings Institution Press, 2008.
- Lewis, W. W.: *The Power of Productivity: Wealth, Poverty, and the Threat to Global Stability*, University Of Chicago Press, 2004.
- Lucas, R. E.: On the size distribution of business firms, *The Bell Journal of Economics*, 9(2), (1978), 508–523.
- Midrigan, V., & Xu, D.: Finance and misallocation: Evidence from plant-level data, Tech. rep., National Bureau of Economic Research, 2010.
- Moll, B.: Productivity losses from financial frictions: Can self-financing undo capital misallocation?, *University of Chicago mimeo*.
- Moscoso-Boedo, H. J., & D’Erasmus, P. N.: Financial structure, informality and development, Virginia Economics Online Papers 374, University of Virginia, Department of Economics, 2009.
- Prado, M.: Government policy in the formal and informal sectors, *European Economic Review*, 55(8), (2011), 1120–1136.
- Rauch, J.: Modelling the informal sector formally, *Journal of Development Economics*, 35(1), (1991), 33–47.
- Restuccia, D.: The Latin American development problem, Working Papers tecipa-318, University of Toronto, Department of Economics, 2008.
- Restuccia, D., & Rogerson, R.: Policy distortions and aggregate productivity with heterogeneous establishments, *Review of Economic Dynamics*, 11(4), (2008), 707 – 720.
- Rubini, L.: Innovation and the elasticity of trade volumes to tariff reductions, *unpublished*.
- Schneider, F., & Enste, D.: *The shadow economy*, Springer, 2002.

Solimano, A., Soto, R., & Nations, U.: *Economic growth in Latin America in the late 20th century: evidence and interpretation*, United Nations Pubns, 2005.

Thomas, J. J.: *Informal Economic Activity*, Harvester Wheatsheaf, New York, N.Y., 1992.

Tybout, J.: Manufacturing firms in developing countries: How well do they do, and why?, *Journal of Economic literature*, 38(1), (2000), 11–44.

AppendixA. Data

Here I describe the details of the data used in the text. Using the census+enamin data that was explained in Section 2, I am able to construct the establishment-size distribution and the employment-distribution by size.

Table A.14: Distribution of establishments and employment by size

census+enamin >1	Establishments	Employment	Mean size
2-5 workers	3,139,715	8,179,911	2.61
6-10 workers	223,277	1,619,209	7.25
11-15 workers	57,336	725,298	12.65
16-20 workers	27,270	485,864	17.82
21-30 workers	25,612	635,431	24.81
31-50 workers	20,079	781,118	38.90
51-100 workers	15,253	1,074,199	70.43
101-250 workers	9,970	1,573,921	157.87
251-500 workers	3,458	1,187,982	343.55
500-1000 workers	1,609	1,119,442	695.74
1001 workers	874	1,846,136	2112.28
Total	3,524,453	19,228,511	5.46

Similarly, the ENOE can be used to obtain size distributions for the formal and informal sectors. The distribution in the ENOE does not match the census+enamin data exactly.

Table A.15: Employment distribution by formality status based on the ENOE

Category	Formal workers	%	Informal workers	%
2-5 workers	710,489	0.08	4,998,485	0.67
6-10 workers	804,494	0.09	1,011,426	0.14
11-15 workers	548,767	0.06	339,537	0.05
16-20 workers	618,057	0.07	272,536	0.04
21-30 workers	672,148	0.07	197,875	0.03
31-50 workers	941,960	0.10	202,336	0.03
51-100 workers	1,212,449	0.13	184,303	0.02
101-250 workers	1,126,611	0.12	101,663	0.01
251-500 workers	756,780	0.08	41,763	0.01
500 +	1,793,019	0.20	82,879	0.01
Total	9,184,774	1.00	7,432,803	1.00

The data for Figure 1 is taken from the ENAMIN (2008). In this Appendix, I present the raw data used to construct Figure 1. I use bold font in order to highlight the range of establishments that have similar capital levels.

Table A.16: Raw data for Figure 1

Workers per establishment	Average capital (pesos)			
	Manufactures	Construction	Trade	Services
1	8,982.50	6,175.64	13,059.73	27,552.09
2	23,961.63	7,998.15	25,849.38	36,893.21
3	60,495.60	27,110.87	51,963.69	42,745.99
4	94,574.80	27,253.23	60,552.82	44,602.05
5	96,564.05	18,269.12	61,749.88	279,235.44
6	110,964.14	26,798.51	1,123,608.13	93,147.28
7	83,739.66			

Appendix B. Aggregate output with marginal product equalization

In this section, I provide a proof of equation 17. Without loss of generality, I do this in the context of a static endowment economy where the distribution of firms, the mass of firms, the labor supply and the stock of capital are all given. I assume the same functional form for the production function of individual firms as in the baseline model.

Let $\mu = \frac{\int z^{1-\gamma} f(z) dz}{N}$, where N is the mass of firms. Using first order conditions, aggregate output can be written as:

$$Y = \int y(z) f(z) dz = \left(\frac{\theta_k/r}{\theta_l/w} \right)^{\frac{\theta_k}{1-\gamma}} \left(\frac{\theta_l}{w} \right)^{\frac{\gamma}{1-\gamma}} N \mu$$

Notice that $\left(\frac{\theta_k/r}{\theta_l/w} \right) = k/l = K/L$, and $\frac{\theta_l}{w} = l/y = L/Y$. Thus,

$$\begin{aligned}
Y &= \left(\frac{K}{L}\right)^{\frac{\theta_k}{1-\gamma}} \left(\frac{L}{Y}\right)^{\frac{\gamma}{1-\gamma}} N\mu \Leftrightarrow \\
Y &= K^{\theta_k} L^{\theta_l} (N\mu)^{1-\gamma}.
\end{aligned}$$

Appendix C. Continuous probability of detection

In Section 2, I made an argument in favor of a step function for the probability of detection; however, it is true that compared to a continuous probability, the former tends to increase the negative effects of resource misallocation and the negative effects of capital constraints. A priori, the analysis in the last section is informative as to what we could expect by changing the probability of detection to a less distortionary shape. In particular, we have shown that an important proportion of the gains come from the improvement in occupational choices; if occupational choices are still sufficiently distorted in the case of a continuous probability of detection, then we should expect still sizable gains from full enforcement, as in the previous section. We have also shown that capital constraints are important for long-run gains; if capital demand is still sufficiently distorted with a continuous function, a similar conclusion follows.

To perform this analysis, I start by replacing the step function of the probability of detection with a linear one.¹⁹ In order to keep all model economies comparable, I use the same parameter values as in Section 5. The question I ask is, how close can we get to the benchmark output (step function) by replacing the probability of detection with a linear function?

Table C.17 presents the results of this exercise when we use perfectly competitive markets. When the probability of detection is a continuous function, we get a substantial amount of the loss we get when the function is stepwise. In particular, output under the linear function is only 3.5% higher than the benchmark. From a different angle, now let the benchmark be the case of full enforcement; now the step function generates an output loss of 9.8%, while the linear function generates 6.7% (that is, 68% of the former). Another interesting aspect in Table C.17 is that the TFP gains from full enforcement are reduced to half of their value in the step-function case; this means that a linear function greatly reduces the negative effects of pure resource misallocation. Nonetheless, the general picture is that the shape of the probability of detection is not crucial to generate the bulk of the effects.

What seems to be crucial is the distortions on capital accumulation. I perform an exercise that helps to clarify the importance of the distortions on capital suffered by informal firms. I replace the

¹⁹Specifically, I assume that the probability of detection is as follows:

$$p(k) = \begin{cases} ak, & 0 < ak \leq 1 \\ 1, & ak > 1 \end{cases}, \tag{C.1}$$

where $a > 0$.

probability of detection with a step function that depends on z , rather than on capital. Under these conditions, informal firms face no capital constraints yet still evade taxes, so we should still expect negative effects from resource misallocation. I depart from full enforcement and ask how close I can get to the benchmark when I introduce informality. Table C.17 presents the results of this exercise.

When the tax authority can observe z , the economy generates 9.2% more output than the benchmark, which is almost the same output as is generated under full enforcement (1.11 of the benchmark). More importantly, notice that in this case the capital stock is not affected by informality; in fact, K ends up being greater than under full enforcement (1.30 vs. 1.20).

Table C.17: Screening of z

	Incomplete enforcement			Full enforcement
	Screening of k		Screening of z (Step function)	
	Step function	Linear function		
Y	1	1.035	1.092	1.109
K	1	1.039	1.297	1.200
TFP	1	1.022	1.002	1.044

In Table C.18 I repeat the analysis of the previous section for the case of a continuous probability of detection. In this table, I have redefined the benchmark to be the distorted economy with a continuous function. Again, what is crucial for the bulk of the effects is the distortion on capital suffered by informal firms. With full enforcement under monopolistic competition, capital accumulation increases by 20%. The increase in output associated with capital accumulation corresponds to roughly 60% ((12-5)/12) of the total gain in output. Another way to see the same point is by looking at the short-run effects on capital price (not shown in the table). Keeping resources constant, when distortions are eliminated, the increase in the capital price is of almost the same magnitude as in the step function case.

Table C.18: Effects with a continuous distortion

Var	Y	K	TFP
Bench (continuous)	1	1	1
Pure misallocation	1.01	1	1.01
Free entry	0.97	1	0.97
Occupational choice SR	1.02	1	1.02
Occupational choice LR	1.07	1.15	1.02
Complementarities SR	1.05	1	1.05
Complementarities LR	1.12	1.20	1.05