CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.


# RESOURCE MISALLOCATION IN MEXICO

TESINA

QUE PARA OBTENER EL GRADO DE

MAESTRO EN ECONOMÍA

PRESENTA

JOAQUIN MAYORGA GARRIDO CORTÉS

DIRECTOR DE LA TESINA: DR. DAVID STRAUSS

CIUDAD DE MÉXICO JUNIO, 2017

## Acknowledgments

*All the data from the Mexican Economic Censuses in this thesis has been reviewed by INEGI to ensure that no confidential information is disclosed.*

*I am particularly grateful with my advisor Dr. David Strauss for his guidance, advise and encouragement. I thank Dr. Arturo Antón for his kind and precise remarks and advice, and Dr. Alexander Dentler for his suggestions and comments. I also thank Dr. Natalia Volkow, Liliana Martínez, Carmen Márquez and Andrea López at INEGI's Laboratorio de Microdatos for their work and kindness. I thank Dr. Ricardo Nogales for his encouragement before my enrollment in the MA in Economics at CIDE, my parents and sister for their continuing support, and my friends here and there, especially Rodrigo, Julio, and Christian, for all the good moments spent during the last two years.*

## Abstract

*I apply and adapt the framework by Hsieh and Klenow (2009a) to study misallocation in Mexico. I analyze the Mexican Economic Censuses of 1993, 1998, 2003, 2008 and 2013. I find that misallocation was stable until 2003 but increased in 2008 and 2013. The increase in misallocation is related to the growth over time of the shares of value added, capital stock net of depreciation and employment corresponding to informal establishments. It also stems from the increase in the dispersion of distortions on the ratio of remuneration over capital costs at the establishment level.*



*Keywords: misallocation, Mexico, informality*
*Clasificación JEL: O40, O47*

# Content

# List of figures

# List of tables

# Chapter 1

# Introduction

GDP per capita has grown slowly in Mexico over the last two decades. TFP sluggishness is responsible for this slow growth. The question that follows is, why has TFP stagnated?

TFP differs across countries mainly for two reasons: the use of different production technologies and the presence of misallocation. There is misallocation when firms within a country use inefficient amounts of resources. This thesis studies the effects and evolvement misallocation on the Mexican manufacturing, retailing and services sectors. It also explores how misallocation is related to firm size, age and formality status. I apply and adapt the framework by (Hsieh and Klenow, 2009a) to analyze data from the five Mexican Economic Censuses from 1993 to 2013.

Hsieh and Klenow (2009a) propose a static general equilibrium model with monopolistic competition, where each firm faces idiosyncratic distortions to its scale and capital costs. They estimate their model using data from the manufacturing sectors of China, India and the United States. Using the United States as the undistorted benchmark, they find that misallocation is more pervasive in India than in China than in the United States. That is, equalizing the idiosyncratic distortions within industries would increase potential output more in the Asian countries than in the United States.

The literature on misallocation argues that informal establishments present inefficiently large scales (Busso, Fazio, and Levy (2012) and Hsieh and Klenow (2009a)). This thesis finds that

according to data from the Mexican Economic Census, the shares of value added, net capital stock net of depreciation and employment corresponding to informal establishments increased over time, at both the aggregate and the sectoral levels. I explore whether the growth of those shares is related to an increase of misallocation in Mexico.

Using the US manufacturing, retailing and services sectors as undistorted benchmarks, I calculate that the gains from reallocation remained constant from 1993 to 2003 at 135%, but started increasing in 2008 to reach 183% in 2013. I find that formal establishments face increasingly larger restrictions to their overall scale over time. I find that due to distortions to the ratio of remuneration to capital costs, informal establishments tend to favor hiring labor over acquiring or renting capital. Distortions to the ratio of remuneration to capital costs became more dispersed in 2013, leading to a further increase in the gains from reallocation.

This thesis proceeds as follows. I review the literature on the misallocation in the remainder of this chapter. Chapter 2 overviews the data from the Mexican Economic Censuses from 1993 to 2013. In chapter 3, I provide a step by step derivation of the equations in the model by Hsieh and Klenow (2009a) and explain how this thesis implements it empirically. Chapter 4 shows the results of the computations and estimations. Chapter 5 concludes.

## 1.1 Literature review

Hanson (2010) reviews two arguments that could explain why the growth of GDP per capita in Mexico has been slower than in other developing countries in recent decades. One argument is related to an external factor: the competition between the Mexican and Chinese manufacturing sectors. The other reason is a combination internal factors which limit TFP growth and are in general related to misallocation. These factors are inefficient credit markets, distortions in the supply of non-tradable inputs, and incentives for informality.

As I mentioned before, TFP can vary across countries due to misallocation. Hsieh and Klenow (2009a), whose paper was reviewed in the previous section, offer a method to estimate

the effects of misallocation on aggregate sectoral TFP. Kim, Oh, and Shin (2016) assert that the approach by Hsieh and Klenow (2009a) may not result in accurate reflections of the actual degree of misallocation because it is based on strong simplifying assumptions, but that it is "a first-pass diagnostics of allocative efficiency."

Multiple authors have applied the framework by Hsieh and Klenow (2009a) to analyze the manufacturing sector in several countries. Busso, Madrigal, and Pagés (2012) use it to calculate the gains from reallocation in the manufacturing sectors of a list of Latin American which report manufacturing sector data at the firm-level and estimate that manufacturing TFP would increase between 45% and 140% depending on the country and year. Similar results are found by Machicado and Birbuet (2012) for Bolivia, and Oberfeld (2013) for Chile. Outside Latin America, important gains from reallocation are found by Dias, Marques, and Richmond (2015) in Portugal, Kim et al. (2016) in Korea, and Ha, Kiyota, and Yamanouchi (2016) in Vietnam. Remarkably, Bellon and Mallen-Pisano (2013), find that in the manufacturing sector of France, a developed country, increases in TFP after the equalization of distortions are lower than in the manufacturing sector of the United States.

It could be suspected that there is a negative correlation between TFP, which is a proxy for GDP per capita, and gains from reallocation. However, Inklaar, Lashitew, and Timmer (2016) analyze a list of 52 middle and low-income countries and find that TFP and gains from reallocation, as calculated applying the method by Hsieh and Klenow (2009a), are uncorrelated. That is, those authors conclude that it is not the poorest countries which would benefit the most from eliminating distortions. However, on a macroeconomic level, Machicado and Birbuet (2012), Oberfeld (2013) and Dias et al. (2015) observe that in the manufacturing sectors of the countries they studied, gains from reallocation are greater during recessions. This relationship could explain the slight increase in misallocation that this thesis estimates for Mexico in 2008, a year during a recession.

At the firm level, Busso, Fazio, and Levy (2012), using OLS regressions, deduce that in the efficient benchmark, resources would be reallocated from informal to formal establishments in

Mexico. At the firm level, Kim et al. (2016), measuring plant size by value added, estimate that younger and larger firms underproduce, while older and smaller establishments overproduce in the Korean manufacturing sector. The use of non-parametric regressions analysis is a salient feature of the research by Kim et al. (2016). The reason they use non-parametric regressions is that they focus on analyzing only the firm size and age dimensions, possibly due to the absence of informal establishments in the Korean manufacturing sector.

Dias, Marques, and Richmond (2016) modify the method by Hsieh and Klenow (2009a) slightly to analyze why misallocation is greater in services than in manufacturing in Portugal. The authors show that 50% of that gap is attributable to differences in the choice of how to measure misallocation. They use regression analysis to find that higher output-price rigidity, labor adjustment lags and informality in the service sector explain the remaining gap.

Hsieh and Klenow (2009b) and Busso, Fazio, and Levy (2012) also analyze the Mexican Economic Censuses applying the framework developed by Hsieh and Klenow (2009a). However, they analyze shorter periods and use a procedure to impute the wages of non-salaried workers, data absent from the Mexican Economic Censuses, which is different from mine. Additionally, a key difference between those papers and this thesis is that this work attempts to explain changes in the gains from reallocation over time.

In this thesis, the missing wages are predicted from an OLS regression on the establishment average wage of salaried workers on establishment characteristics. Hsieh and Klenow (2009b) impute establishment remuneration by multiplying the number of workers by the average wage per worker at the sector level. However, my results relating misallocation, size age and formality status to misallocation are roughly similar to those found by Hsieh and Klenow (2009b), who analyze data from 1998 and 2003. Busso, Fazio, and Levy (2012) do not disclose their method to impute wages. Regardless of that absence of information, the findings of this thesis regarding the relationship between formality and productivity and misallocation are akin to those found by Busso, Fazio, and Levy (2012). These authors study data from 1998, 2003 and and arrive at their results after separating their sample by size categories, where establishment size is measured as

the number of workers.

López-Calva and Levy Algazi (2016) argue that misallocation is responsible for the decrease in the wage returns to education in Mexico. They find that the wage premium to education has decreased in Mexico in the last 20 years. Using data from 4 out of the 5 Mexican Economic Censuses that this thesis analyzes, they propose a counterfactual scenario where misallocation is absent based on Hsieh and Klenow (2009a). They argue that in the absence of misallocation the wage premium gap would increase because more (less) productive establishments would expand, and hire more educated workers at higher wages.

Bento and Restuccia (2016) propose a tractable dynamic general equilibrium model to analyze a specific type of misallocation. They analyze the so-called correlated distortions, which disproportionately affect larger establishments. They calibrate their model using US data and find that their calibrated model capture to a large extent the gains from reallocation computed by Hsieh and Klenow (2009a).

Finally, other papers also study misallocation proposing dynamic, instead of static, models. Hsieh and Klenow (2014) focus on the distortions than plants in Mexico and India face according to their age and estimate that they decrease potential TFP in the manufacturing sectors of those countries by 25%. Restuccia and Rogerson (2008), calibrating their model with data from the United States, calculate that policies that increase the dispersion of factor prices can diminish TFP from 30% to 50%. Buera, Kaboski, and Shin (2011) observe that frictions in the credit market can explain a substantial part of the differences in TFP and relative sectoral productivities across countries.

# Chapter 2

# Background statistics

This thesis uses data from the Mexican Economic Censuses from 1993, 1998, 2003, 2008 and 2013. The Mexican Economic Census captures data from private and non-agricultural establishments with fixed locations. It consists of data from all the urban establishments of that type, and a random sample of the rural units of that type. This thesis focuses on establishments belonging to the manufacturing, retailing and services sectors.

Table 2.1 shows the proportion of total employment in each sector captured by the Mexican Economic Census in all years. The share of employment surveyed by theses censuses increases in the three sectors over time but fails to surpass 60% of aggregate employment in those sectors. These censuses capture a larger share of employment in manufacturing and retailing than in services in all years. The share of manufacturing employment surveyed by INEGI increases the least over time, while the share of services employment expands the most.

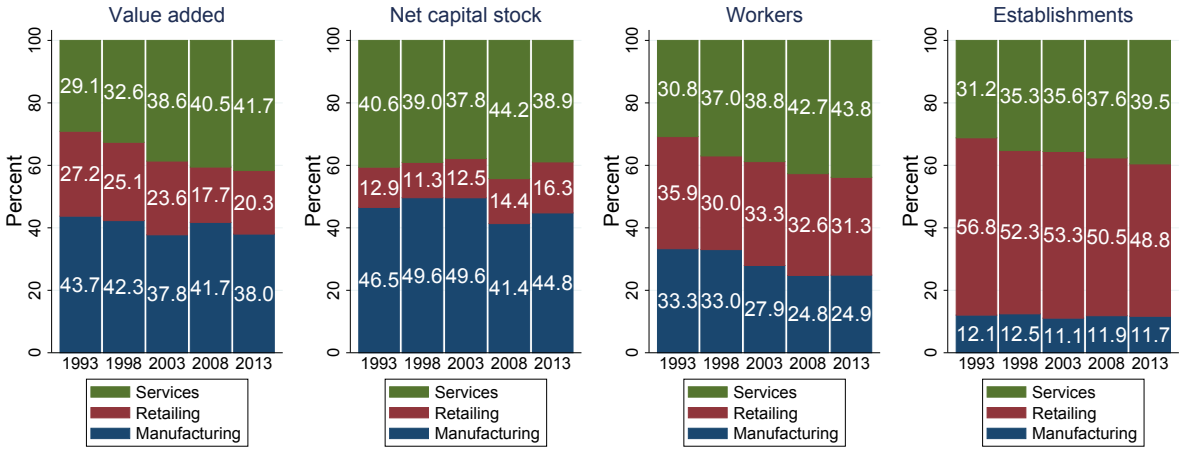
With this caveat in mind, let us provide some aggregate data from the Mexican Economic Censuses. Figure 2.1 shows the shares of aggregate value added, capital stock net of depreciation, employment, and establishments corresponding to the manufacturing, retailing and services sectors from 1993 to 2013. Regarding value added, the share corresponding manufacturing fluctuates, the proportion belonging to retailing shrinks and the share of services grows. Regarding employment, the share of services increases while the shares corresponding to retailing and

Table 2.1: Percentage of total employment captured by the Mexican Economic Census

| Sector | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| Manufacturing | 63.35 | 61.87 | 61.56 | 65.19 | 66.35 |
| Retailing | 56.94 | 58.54 | 67.12 | 71.30 | 66.11 |
| Services | 30.75 | 40.48 | 44.92 | 50.95 | 47.90 |
| Total | 46.23 | 50.97 | 55.14 | 59.72 | 56.70 |

*Source*: Author's calculations based on data from INEGI.
*Note*: For 1993 I used total employment data from 1995, because data from previous years is unavailable. The total employment data from 1995 to 2008 corresponds to the National Survey of Employment (ENE, by its initials in Spanish), and the data from 2013 to the National Survey of Occupation and Employment (ENOE, by its initials in Spanish).

Figure 2.1: *Shares of value added, capital stock net of depreciation, employment, and establishments by sector*



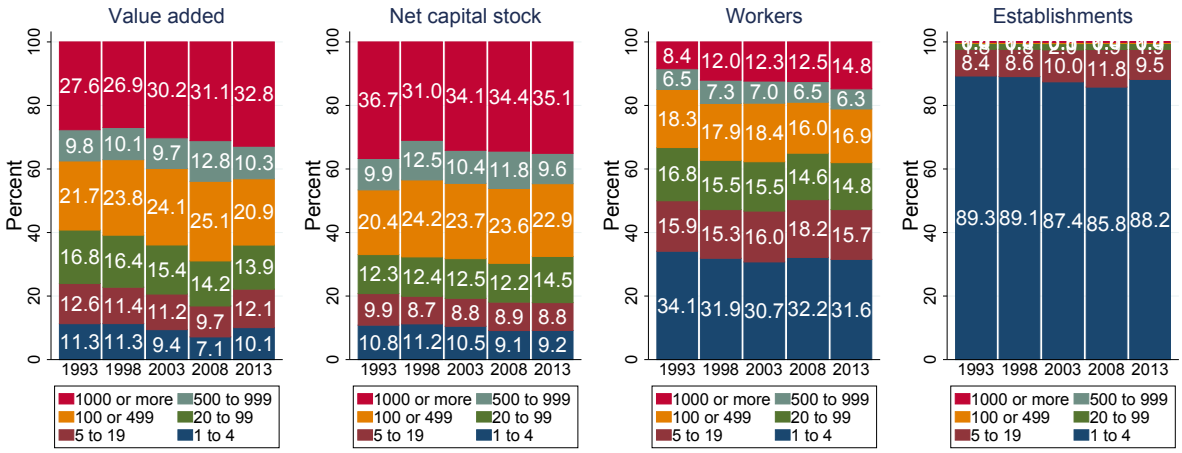

*Source*: Author's calculations based on data from INEGI.

manufacturing decrease. The proportion of establishments belonging to manufacturing remains approximately constant, while the share corresponding to services increases in detriment of the share of retailing.

Manufacturing has the largest share of value added in three out of five years, and the smallest proportion of the number of establishments in all years. Clearly, manufacturing displays the largest net capital stock and output both per worker and per establishment. In contrast, retailing has the smallest share of value added and the largest share of establishments in all years. Output and net capital both per worker and per establishment are also the lowest in retailing.

## 2.1 Size and age

Figure 2.2: *Shares of value added, capital stock net of depreciation, workers, and establishments by size category*

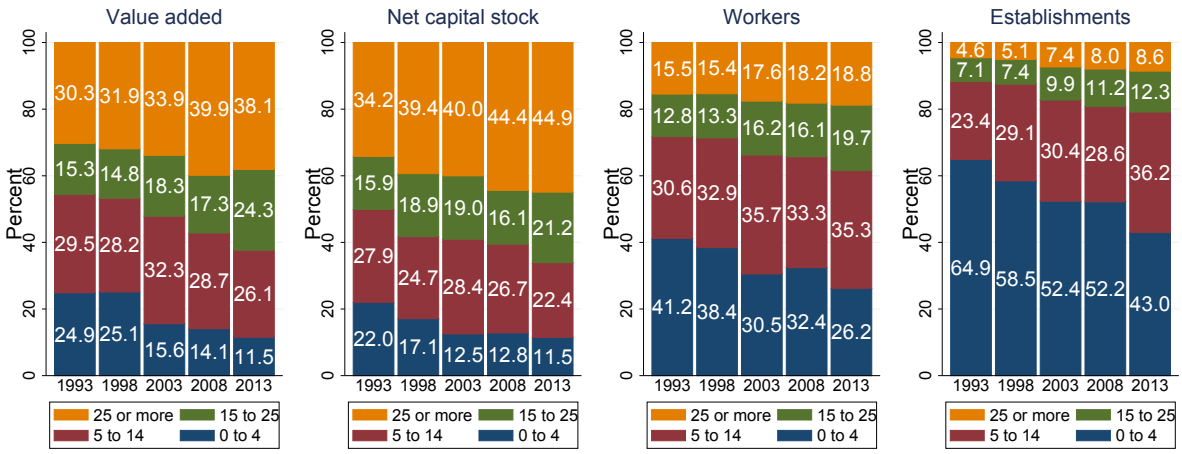



*Source*: Author's calculations based on data from INEGI.

The analysis of the shares regarding establishment size relies on figure 2.2. This figure indicates that the share of value added and employment corresponding to establishments employing 500 or more workers increases over time. Interestingly, the share of value added corresponding to the smallest establishments is the lowest in 2008, which is the only year in the sample during a recession. Establishments employing more workers display higher value added and net capital stock both per worker and per establishment.

Figure 2.3 shows the relevant shares by age category. It shows that the shares corresponding to establishments older than 15 years of age increase over time in the four areas. The greater shares of units of that age may simply reflect a higher survival rate of already established firms. Appendix B shows the survival rates by year of birth from one Mexican Economic Census to the next. However, the increase in the share of value added and employment of establishments of 15 or more years of age is possibly related to the increasing shares of value added, net capital stock and employment corresponding to larger establishments, which tend to be older.

Figure 2.3: *Shares of value added, capital stock net of depreciation, workers and establishments by age category*




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

## 2.2 Formality

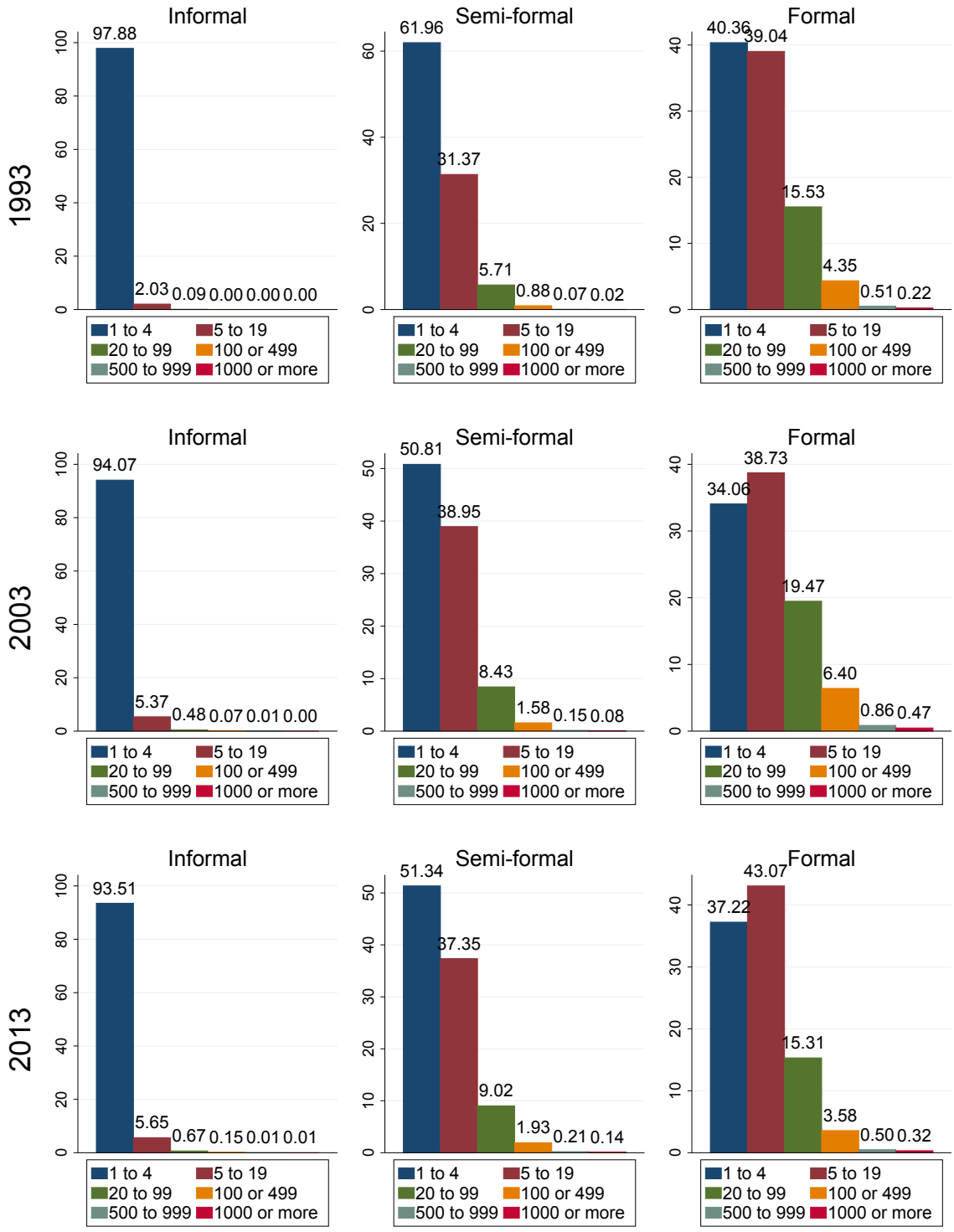
The chosen definition of formality status comes from Busso, Fazio, and Levy (2012) [1]. Formal establishments pay over 18% of total remuneration on social contributions, while informal units do not pay any social contributions. Semi-formal establishments pay social contributions that represent above 0% but less than or equal to 18% of their total remuneration.

An evident difference between informal and non-informal establishments is that the former are smaller than the latter. Figure 2.4 shows the number of establishments by formality status and size category in 1993, 2003 and 2013. The distribution of units by size category is more uniform for semi-formal and formal establishments than for informal units. Regarding informal and semi-formal establishments, the share of establishments employing 5 or more workers increases over time.

The shares of value added, capital stock net of depreciation, employment and number of establishments by formality status are displayed on figure 2.5 The increase over time of the shares corresponding to informal units is evident. The shares corresponding to formal establishments decrease from 1993 to 2008 but increase in 2013. From table 2.1, it could be suspected that the

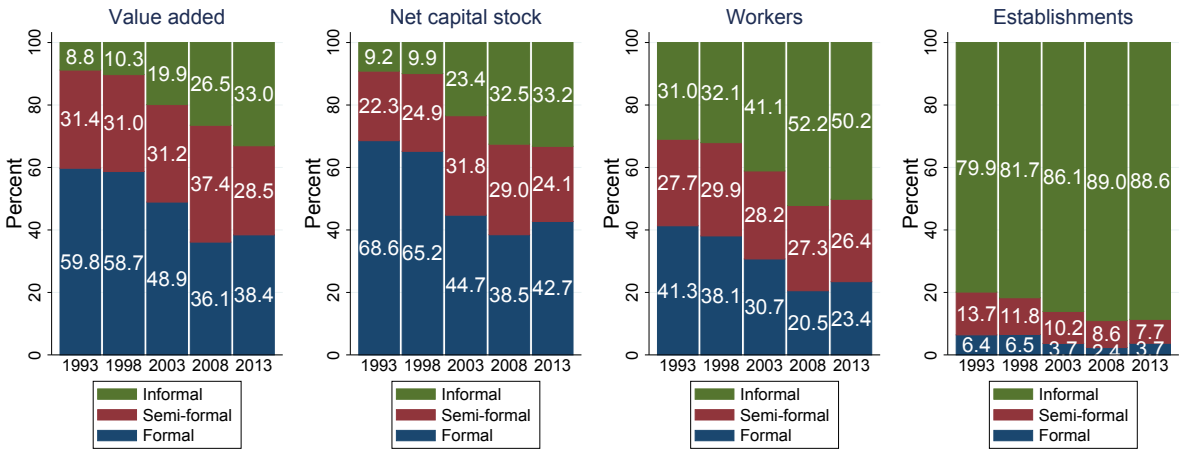
[1] I classified establishments by their formality status after imputing the wages of non-salaried workers using a method described in section 3.2.2.

Figure 2.4: *Number of establishments by size category and formality status*




*Source*: Author's calculations based on data from INEGI.
*Note*: Percentages of 0.0 indicate that at least one establishment belong to the corresponding category.

Figure 2.5: *Shares of value added, capital stock net of depreciation, employment and establishments by formality status*



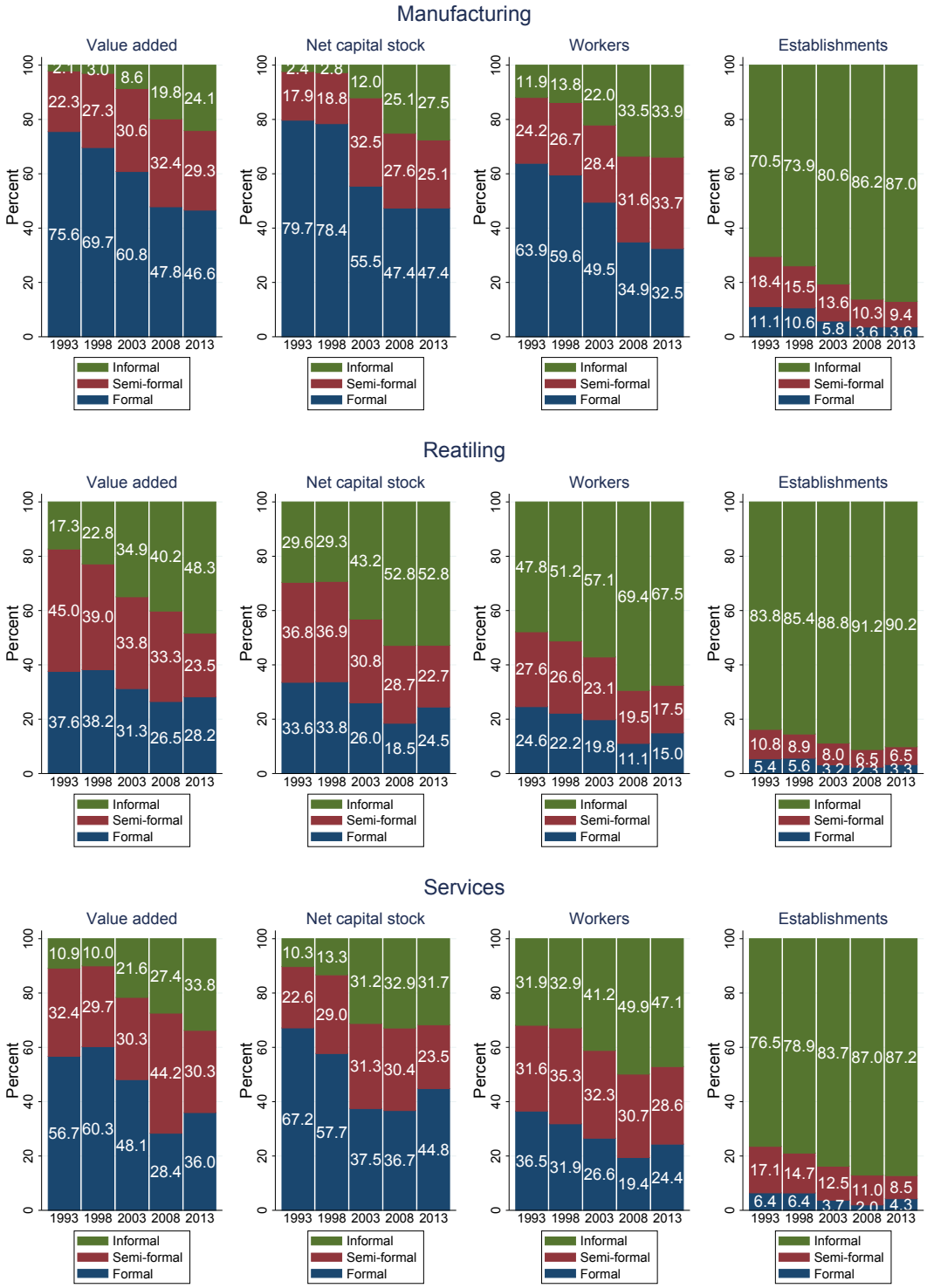

*Source*: Author's calculations based on data from INEGI.

increase in the shares of informality shown in figure 2.5 stems from an effort by INEGI to survey a greater proportion of small and informal establishments. This change in the data gathering approach by INEGI could imply that the Mexican Economic Censuses understate informality less over time.

I claim that even if the Mexican Economic Censuses capture data from a greater share of informal establishments and workers in more recent years, it is reasonable to assume that actual informality has increased for two reasons. First, from table 2.1, it is clear that the share of employment in manufacturing captured by the Mexican Economic Censuses remains approximately stable over time. However, the shares of value added, employment and establishments of informal units in manufacturing in the Mexican Economic Censuses increases every year since 1993, according to figure 2.6. Second, it is unlikely that the Mexican Economic Census had omitted surveying informal establishments that were large in any year. 2.7 and 2.8 show that the shares of value added, net capital stock, workers and number of establishments corresponding to informal units increase across all size categories at least until 2008.

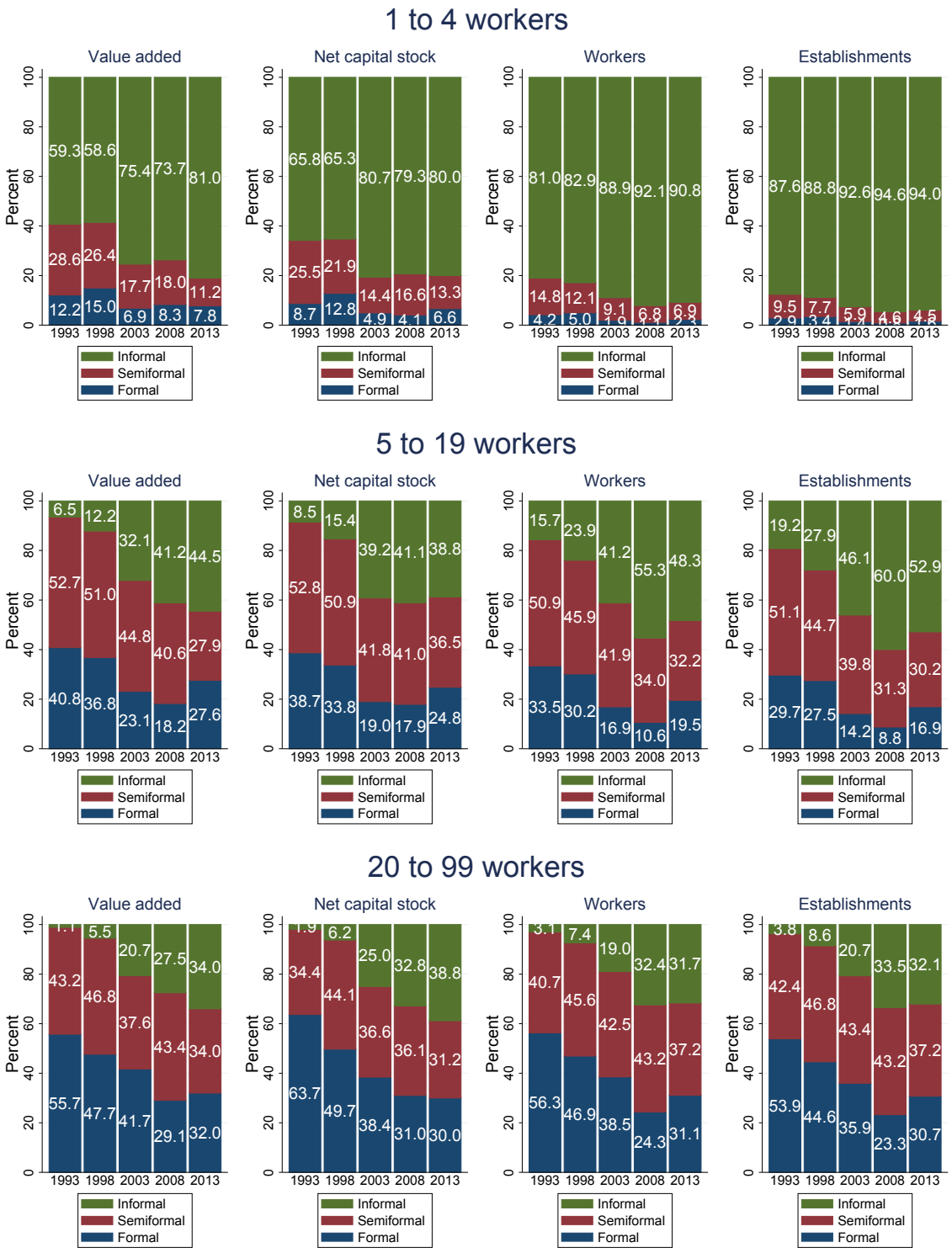
Summing up, there is an increase in the shares of informality even in the sector where the proportion of workers captured by the Mexican Economic Census remains constant over time. Additionally, the shares corresponding to informal establishments increase across all size cate-

Figure 2.6: *Shares of value added, workers, and establishments by formality status across sectors*




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

Figure 2.7: *Shares of value added, capital stock net of depreciation, workers, and establishments by formality status for units employing 1 to 99 workers*




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

Figure 2.8: *Shares of value added, capital stock net of depreciation, workers, and establishments by formality status for units employing 100 or more workers*


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

gories, and not only among establishments employing 1 to 4 workers. Therefore the increase of those shares in the database must reflect a similar increase in reality.

The interest of this thesis lies on estimating the extent and evolvement of misallocation in Mexico. From Hsieh and Klenow (2009b) and Busso, Fazio, and Levy (2012), I expect the analysis of thesis to establish that resources would flow from informal to formal units in an efficient environment. Therefore I expect an increase in misallocation over time, stemming from the growth in the shares of value added, net capital and employment corresponding to informal establishments.

# Chapter 3

# Methodology

This chapter firstly presents the one-period general equilibrium model with monopolistic competition developed by Hsieh and Klenow (2009a). The purpose of the model is to show how misallocation can lower aggregate $TFP$ at the sector level. The chapter continues outlining the procedure chosen to estimate that model, including the setting of parameters. It proceeds explaining how I imputed the missing data on the wages of non-salaried workers in the Mexican Economic Census. Finally, it summarizes the procedure I used to select the sample.

## 3.1 The model by Hsieh and Klenow (2009)

The agents of the model are the final good producer, industries, and firms. The final good producer aggregates the output of $S$ industries, where $S$ is the number of industries. Industry $s$ produces intermediate good $s \in \{1, 2, ..., S\}$ and agregates the output of $M_s$ firms, where $M_s$ is the number of firms selling their output to industry $s$. The final good producer and the industries operate under perfect competition: they attain zero profits. Each firm selling its output to industry $s$ produces a differentiated good $si$, where $i \in \{1, 2, ..., M_s\}$. Firms operate under monopolistic competition. Let us look at the model in detail.

The producer of the final good $Y$ combines the output $Y_s$ of $S$ industries using a Cobb-

Douglass production technology:

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

where $\theta_s$ is the output share of the intermediate good produced by industry $s$. The final good producer solves the following cost minimization problem:

$$\begin{aligned} \underset{Y_1,...,Y_S}{\text{minimize}} \quad & \sum_{s=1}^{S} P_s Y_s \\ \text{subject to} \quad & Y = \bar{Y}, \end{aligned}$$

where $P_s$ is the price of the intermediate good produced by industry $s$.

Industry $s$ output $Y_s$ is a $CES$ aggregate of $M_s$ differentiated goods:

$$Y_s = \left( \sum_{i=1}^{M_s} Y_{si}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \text{where } \sigma > 1. \tag{3.2}$$

The industry producing intermediate good $s$ solves the following profit maximization problem:

$$\underset{Y_{s1},...,Y_{sM_s}}{\text{maximize}} \quad \pi_s = P_s Y_s - \sum_{si=1}^{M_s} P_{si} Y_{si},$$

where $P_{si}$ is the price of differentiated good $Y_{si}$. From this problem's first order condition with respect to $Y_{si}$, demand for differentiated good $Y_{si}$ is written as:

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

The production function of each differentiated good $si$ producing firm is given by a Cobb-Douglass function of capital and labor:

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

The capital share $\alpha_s$ is common to all firms selling their output to industry $s$. Firm $si$ profits are

given by:

$$\pi_{si} = (1 - \tau_{Ysi})P_{si}Y_{si} - (1 + \tau_{Ksi})RK_{si} - wL_{si}, \tag{3.5}$$

where the rental price of capital, $R$, and the cost of a unit of labor, $w$, are common to all firms. $\tau_{Ysi}$ is a distortion that increases the marginal products of both factors, and $\tau_{Ksi}$ is the distortion that increases the marginal product of capital relative to the marginal product of labor.

Distortions are firm-specific. We call $\tau_{Ysi}$ the output distortion, and $\tau_{Ksi}$ the capital distortion. For example, a firm receiving subsidies to output will display a beneficial output distortion ($\tau_{Ysi} < 0$), while an establishment facing high transportation costs will show a hurtful output distortion ($\tau_{Ysi} > 0$). However, in a context where numerous firms avoid paying taxes, such as the Mexican, output distortions could simply represent the taxes paid by formal firms. The capital distortion can be advantageous ($\tau_{Ksi} < 0$) for a firm enjoying preferential access to credit. We could observe a harmful capital distortion ($\tau_{Ksi} > 0$) in firms hiring labor indirectly subsidized by social programs (see Azuara and Marinescu (2013)). Policies and market inefficiencies can be the source of distortions.

Firm $si$ solves a two-stage problem involving cost minimization and profit maximization. The cost minimization problem is:

$$\begin{aligned} \underset{K_{si}, L_{si}}{\text{minimize}} \quad & (1 + \tau_{Ksi})R + wL_{si} \\ \text{subject to} \quad & Y_{si} = \bar{Y_{si}}. \end{aligned}$$

Manipulating its first order conditions yields the optimal ratio of capital to labor, and the optimal cost per unit of output:

$$\frac{K_{si}}{L_{si}} = \frac{\alpha_s}{1 - \alpha_s}\frac{w}{R}\frac{1}{1 + \tau_{Ksi}} \quad \text{and} \quad c_{si} = A_{si}^{-1}\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1 - \alpha_s}\right)^{1-\alpha_s}(1 + \tau_{Ksi})^{\alpha_s}. \tag{3.6}$$

Using $c_{si}$ and the industry $s$ demand for differentiated good $si$, we express the profit maxi-

mization problem as:

$$\underset{P_{si}}{\text{maximize}} \quad \pi_{si} = (1-\tau_{Ysi})P_{si}Y_{si}^{d} - c_{si}Y_{si}^{d} \quad .$$

The solution to the profit maximization problem of firm $si$ is

$$P_{si} = \frac{\sigma}{\sigma-1}A_{si}^{-1}\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\frac{(1+\tau_{Ksi})^{\alpha_s}}{1-\tau_{Ysi}}, \tag{3.7}$$

which is a fixed marked-up over unitary cost $c_{si}$.

### 3.1.1 Total factor productivity of revenue

The distinction between physical and revenue productivity is key to the model. At the firm level, physical productivity is simply $A_{si}$, and it is often referred to as $TFPQ$. Revenue productivity or total factor productivity of revenue, $TFPR$, is defined as price times physical productivity:

$$TFPR_{si} \equiv P_{si}A_{si} = \frac{\sigma}{\sigma-1}\left(\frac{R}{\alpha_s}\right)^{\alpha_s}\left(\frac{w}{1-\alpha_s}\right)^{1-\alpha_s}\frac{(1+\tau_{Ksi})^{\alpha_s}}{1-\tau_{Ysi}}. \tag{3.8}$$

### 3.1.2 Marginal revenue products of capital and labor

It is also useful to express firm $TFPR$ as

$$TFPR_{si} = \left(\frac{MRPK_{si}}{\alpha_s}\right)^{\alpha_s}\left(\frac{MRPL_{si}}{1-\alpha_s}\right)^{1-\alpha_s} = \frac{P_{si}Y_{si}}{K_{si}^{\alpha_s}L_{si}^{1-\alpha_s}}, \tag{3.9}$$

where $MRPK$ and $MRPL$ refer to the marginal revenue products of capital and labor:

$$MRPK_{si} \equiv \frac{\sigma}{\sigma-1}R\frac{1+\tau_{Ksi}}{1-\tau_{Ysi}} = \alpha_s\frac{P_{si}Y_{si}}{K_{si}} \quad \text{and}$$

$$MRPL_{si} \equiv \frac{\sigma}{\sigma-1}w\frac{1}{1-\tau_{Ysi}} = (1-\alpha_s)\frac{P_{si}Y_{si}}{L_{si}}.$$

Let us interpret the expressions of firm total factor productivity of revenue. For example, a firm facing high capital rental costs and high transportation costs displays an inefficiently higher $TFPR$. This firm's marginal revenue product of capital is too high, as it rents too few units of capital and hires too few workers. The model aims at comparing the actual with the efficient allocation of resources. In the efficient benchmark, it is assumed that marginal revenue products equalize across firms within an industry, leading to an efficient reallocation of resources.

### 3.1.3 Industry TFP

At the industry level, the allocations of capital and labor are defined as $K_s = \sum_{i=1}^{M_s} K_{si}$ and $L_s = \sum_{i=1}^{M_s} L_{si}$. Industry $TFPR$ is given by:

$$\overline{TFPR_s} \equiv \left(\frac{\overline{MRPK_s}}{\alpha_s}\right)^{\alpha_s} \left(\frac{\overline{MRPL_s}}{1-\alpha_s}\right)^{1-\alpha_s} = \frac{P_s Y_s}{K_s^{\alpha_s} L_s^{1-\alpha_s}}, \tag{3.10}$$

where $\overline{MRPK_s}$ and $\overline{MRPL_s}$ refer to the marginal revenue products of capital and labor at the industry level:

$$\overline{MRPK_s} \equiv \frac{\sigma}{\sigma-1} R \left(\sum_{i=1}^{Ms} \frac{1-\tau_{Ysi}}{1+\tau_{Ksi}} \frac{P_{si}Y_{si}}{P_s Y_s}\right)^{-1} = \alpha_s \frac{P_s Y_s}{K_s} \quad \text{and} \tag{3.11}$$

$$\overline{MRPL_s} \equiv \frac{\sigma}{\sigma-1} w \left(\sum_{i=1}^{Ms} (1-\tau_{Ysi)} \frac{P_{si}Y_{si}}{P_s Y_s}\right)^{-1} = (1-\alpha_s) \frac{P_s Y_s}{L_s}. \tag{3.12}$$

Then industry $TFP$ can be defined as [1]:

$$TFP_s \equiv \left[\left(A_{si} \frac{\overline{TFPR_s}}{TFPR_{si}}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}} = \frac{Y_s}{K_s^{\alpha_s} L_s^{1-\alpha_s}} \tag{3.13}$$

Finally, note that if $log(A_{si}), log(1+\tau_{Ksi}), log(1-\tau_{Ysi})$ is multivariate normal, then:

[1] See Appendix A for a derivation of this expression for TFP.

$$logTFP_s = \frac{1}{\sigma - 1}\left(logM_s + E[A_{si}^{\sigma-1}]\right) - \frac{\sigma}{2}var(logTFPR_{si}) - \frac{\alpha_s(1-\alpha_s)}{2}var(log(1+\tau_{Ksi})) \quad (3.14)$$

### 3.1.4 Gains from reallocation

Now we are ready to compare the actual output with the production in the efficient benchmark. But first, we must lay out the assumptions behind the efficient benchmark. First, distortions and thus marginal revenue products equalize within industries, leading to a reallocation of factors across firms. And second, the allocation of capital and labor across industries remains unchanged [2]. Therefore potential gains of output stem from the efficient reallocation of capital and labor within industries [3].

The equalization of distortions within industries implies $\overline{TFPR_s} = TFPR_{si}$ for all $i$. This leads to the efficient industry $TFP$:

$$\overline{As} \equiv \left[\sum_{i=1}^{M_s} A_{si}^{\sigma-1}\right]^{\frac{1}{\sigma-1}}. \quad (3.15)$$

$\overline{TFPR_s}/\overline{A_s}$ equals the ratio of actual to efficient industry output, given that the allocation of capital and labor across industries remains unchanged. Plug the ratios of actual to efficient industry output into the aggregate production function to obtain:

$$\frac{Y}{Y_{eff}} = \prod_{s=1}^{S}\left[\frac{\overline{TFPR_s}}{\overline{A_s}}\right]^{\theta_s} = \prod_{s=1}^{S}\left[\sum_{i=1}^{M_s}\left(\frac{A_{si}}{\overline{A_s}}\frac{\overline{TFPR_s}}{TFPR_{si}}\right)^{\sigma-1}\right]^{\frac{\theta_s}{\sigma-1}}. \quad (3.16)$$

From this equation, it is straightforward to calculate the potential gains from reallocation.

Finally, we want to compare actual versus efficient output at the firm level. Firm output is given by[4]:

[2] See Appendix A for a derivation of the allocation of resources across industries.

[3] Limiting reallocation across firms within industries yields estimates of the gains from reallocation that are conservative as compared to the case in which there is also reallocation across industries. Moreover, limiting reallocation to occur within industries makes the estimates independent of the value of the rental rate of capital.

[4] See Appendix A for a derivation of this expression.

$$P_{si}Y_{si} = \theta_s Y \left(\frac{A_{si}}{TFPR_{si}}\right)^{\sigma-1} \left[\sum_{j=1}^{M_s} \left(\frac{A_{sj}}{TFPR_{sj}}\right)^{\sigma-1}\right]^{-1}. \tag{3.17}$$

In the efficient benchmark, $Y$ equals $Y_{eff}$, and $TFPR_{si} = TFPR_{sj}$ is satisfied for all $i$ and $j$. Therefore efficient firm output is:

$$P^*_{si}Y^*_{si} = \theta_s Y_{eff} \frac{A_{si}^{\sigma-1}}{\sum_{j=1}^{M_s} A_{sj}^{\sigma-1}}. \tag{3.18}$$

## 3.2 Emprirical implementation

As told before, I use data from the five Mexican Economic Censuses from 1993 to 2013.

### 3.2.1 Selection of parameters and estimation procedure

Recovering some measures of productivity and distortions is necessary to calculate the gains from reallocation and analyze misallocation. This requires defining some parameters: the rental rate of capital, the elasticity of substitution, and the capital shares. Following Hsieh and Klenow (2009a), I set the rental rate of capital at $R = 0.10$, assuming a $5\%$ real interest rate and a $5\%$ inflation rate. This choice is made to facilitate comparisons with results from the literature. Additionally, it is worth noting that the interest rate does not affect the estimates of the gains from reallocation and that setting different interest rates does not affect further findings in an economically significant manner.

Hsieh and Klenow (2009a) fix the elasticity of substitution at $\sigma = 3$, and define industries at the ISIC 4-digit level. Instead, I set the elasticity of substitution at $\sigma = 2$, and define industries more broadly, at the NAICS 3-digit level. Intuitively, setting a lower elasticity of substitution stems from the recognition that the demand for differentiated goods is less elastic when industries are defined more broadly. For example, the elasticity of substitution between types of bread is arguably higher than between bread and canned foods. Due to data availability, this thesis includes the bread and canned food industries within the food manufacturing industry, while other

authors are able to define those two industries and thus set a higher $\sigma$ than this thesis.

Defining industries at that level allows me to use publicly available data from the Bureau of Economic Analysis to construct the capital shares of the United States manufacturing, retailing and services industries. For each year of the sample, I calculate industry $s$ capital shares as follows:

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

Let us now look at the recovery of distortions and productivities. In this case, $K_{si}$, $P_{si}Y_{si}$ and $wL_{si}$ refer to the capital net of depreciation, value added and remuneration of establishment $si$. Similarly, $K_s$, $P_sY_s$ and $wL_s$ stand for the aggregate capital net of depreciation, value added and remuneration of industry $s$.

I estimate distortions as functions of observable variables as follows:

$$1 + \tau_{Ksi} = \frac{\alpha_s}{1-\alpha_s}\frac{wL_{si}}{RK_{si}} \quad \text{and} \tag{3.19}$$

$$1 + \tau_{Ysi} = \frac{\sigma}{\sigma - 1}\frac{wL_{si}}{(1-\alpha_s)P_{si}Y_{si}}. \tag{3.20}$$

These expressions are derived from equations 3.8 and 3.9. I calculate average distortions at the industry level similarly, but replacing establishment-level variables with industry-level observables:

$$1 + \overline{\tau_{Ks}} = \frac{\alpha_s}{1-\alpha_s}\frac{wL_s}{RK_s} \quad \text{and} \tag{3.21}$$

$$1 - \overline{\tau_{Ys}} = \frac{\sigma}{\sigma - 1}\frac{wL_s}{(1-\alpha_s)P_sY_s}. \tag{3.22}$$

If all establishments within an industry faced these average distortions, $MRPK_s$ and $MRPL_s$ would be the same as those observed in the data.

I obtain the ratio of an establishment to industry revenue productivities as follows:

$$\frac{TFPR_{si}}{\overline{TFPR_s}} = \frac{\frac{P_{si}Y_{si}}{K_{si}^{\alpha_s} w L_{si}^{1-\alpha_s}}}{\frac{P_s Y_s}{K_s^{\alpha_s} w L_s^{1-\alpha_s}}}, \tag{3.23}$$

where the wage $w$, which is unobservable, cancels out. This equation is derived from equations 3.9 and 3.10. From the production function of $Y_{si}$ in equation 3.4, and the demand for $Y_{si}$ in equation 3.3, physical productivity times $(w/w)^{1-\alpha_s}$ is written as:

$$A_{si} = \frac{w^{1-\alpha_s}}{P_s (P_s Y_s)^{\frac{1}{\sigma-1}}} \frac{(P_{si}Y_{si})^{\frac{\sigma}{\sigma-1}}}{K_{si}^{\alpha_s}(wL_{si})^{1-\alpha_s}} \quad . \tag{3.24}$$

Since both $P_s$ and $w$ are not observable, $A_{si}$ cannot be recovered. However, it suffices to calculate the ratio $A_{si}/\overline{A_s}$ given the purposes of this thesis. Using equation 3.15 for $\overline{A_s}$, defining $\kappa_s = w^{1-\alpha_s} P_s^{-1} (P_s Y_s)^{-\frac{1}{\sigma-1}}$, and setting $\kappa_s = 1$ allows the calculation of the ratio $A_{si}/\overline{A_s}$ from observables.

Finally, the value added share of intermediate good $s$ is obtained as follows:

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

With the ratios $TFPR_{si}/\overline{TFPR_s}$ and $A_{si}/\overline{A_s}$, and the shares $\theta_s$, the estimation of the gains from reallocation using equation 3.16 is straightforward.

### 3.2.2 Imputation of missing wages

Recall that the variables required at the establishment-level are the capital stock net of depreciation, value added and remuneration. The Mexican Economic Censuses report the first two of those variables for all establishments. However, there is an issue with the remuneration data in the Mexican Economic Censuses, where remuneration is defined as the sum of wages, social contributions and shared profits.

The total number of workers of an establishment consists of the quantity of salaried and non-salaried workers it employs. However, establishments do not report the wages of non-salaried

workers to the Mexican Economic Census. Therefore, first, virtually all establishments employing only non-salaried workers report zero wages and thus null remuneration. And, second, establishments employing both types of workers underreport remuneration.

I provide some data on non-salaried workers, from a pooling of observations from all years. 37% of the workers are non-salaried workers, of which 79% work in establishments that employ only non-salaried workers. 68% of establishments employ only unpaid workers. 18% of establishments employ both salaried and non-salaried workers. Therefore the proportion of firms that could be excluded from the calculations due to missing wage data is significant.

To cope with this partial lack of wage data, I impute wage data for non-salaried workers. I predict establishment wages per salaried worker using the regression results shown in table 3.1. The sample of this regression consists of all establishment reporting strictly positive capital stock net of depreciation, value added and wages. The specification of this regression aims at maximizing its adjusted R-squared, and it satisfies three additional criteria. First, it displays only variables that are highly statistically significant except for some federal entity dummies. Second, all the average wages it predicts are positive. And third, the average wages it predicts are not unreasonably high for the largest and oldest establishments for which wage data was imputed.

I will explain in detail how I constructed the semi-legal and illegal dummies that are independent variables in that regression. I constructed those dummies applying the definitions of establishment legality provided by Busso, Fazio, and Levy (2012). These definitions state that within establishments employing at least one salaried worker, an establishment is defined as illegal if it does not pay any social contributions, as semilegal if its social contributions represent less than or equal to 18% of total remunerations, and as legal if its social contributions represent above 18% of remuneration. When imputing the wages of non-salaried workers, I assume that all establishments remunerate non-salaried workers as if these workers were being paid by illegal establishments.

Table 3.1: Regressions on establishment wage per salaried worker

| Variable | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| Number of workers | 0.01 | 0.01 | 0.02 | 0.03 | 0.02 |
| Squared number of workers | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Age | 0.23 | 0.41 | 0.42 | 0.89 | 1.21 |
| Squared age | 0.00 | 0.00 | 0.00 | 0.00 | -0.01 |
| Semilegal | 0.97 | 1.71 | 3.03 | 2.54 | 4.31 |
| Illegal | -2.99 | -9.60 | -13.07 | -16.20 | -13.69 |
| Social contributions / Remuneration | -5.79 | -30.31 | -44.77 | -26.49 | -16.03 |
| Size bins | YES | YES | YES | YES | YES |
| Age bins | YES | YES | YES | YES | YES |
| Sector F.E. | YES | YES | YES | YES | YES |
| Federal entity F.E. | YES | YES | YES | YES | YES |
| N | 607234 | 838203 | 873458 | 1034412 | 1067198 |
| R-squared | 0.10 | 0.03 | 0.04 | 0.13 | 0.10 |
| Adjusted R-squared | 0.10 | 0.03 | 0.04 | 0.13 | 0.10 |

*Source*: Author's calculations based on data from INEGI.
*Note*: Except for some federal entity dummies, all variables are significant at the 0.01% level. In the "Social contributions / Remuneration" coefficient, remuneration refers to remuneration as observed in the Mexican Economic Censuses.

### 3.2.3 Sample

I do not use all the observations available in the Mexican Economic Census to perform the calculations. First, I exclude all establishments reporting zero or negative values for their capital stock net of depreciation and/or value added. This exclusion is required by the estimation procedure. Second, I exclude the industries with ten or fewer establishments in one or more years. These industries are the rail and pipeline transportation industries. And third, following Hsieh and Klenow (2009a), I trim the top 1% tails of two variables, $ln(A_{si}M_s^{\frac{1}{\sigma-1}}/\overline{A_s})$ and $ln(TFPR_{si}/\overline{TFPR_s})$, in each year [5]. Trimming extreme outliers has the purpose of moderating the effects of measurement error on the results.

Table 3.2 shows some descriptive statistics about the sample. The final sample excludes establishments reporting negative value added or stock of capital net of depreciation. The second

[5]The main conclusion do not change when the top 0.5% of $ln(A_{si}M_s^{\frac{1}{\sigma-1}}/\overline{A_s})$ and $ln(TFPR_{si}/\overline{TFPR_s})$ are trimmed

Table 3.2: Descriptive statistics of the sample

| Year | Sector | VA | Net capital | Workers | Establishments | % Excluded | |
|---|---|---|---|---|---|---|---|
| | | % | % | % | % | Est. | Wkrs. |
| 1993 | Manufacturing | 38 | 43 | 32 | 12 | 1 | 3 |
| | Retail | 33 | 18 | 28 | 57 | 4 | 6 |
| | Services | 29 | 39 | 30 | 31 | 4 | 8 |
| | Total | | | | | 4 | 6 |
| 1998 | Manufacturing | 37 | 48 | 32 | 13 | 8 | 6 |
| | Retail | 35 | 17 | 33 | 52 | 8 | 10 |
| | Services | 28 | 35 | 35 | 35 | 9 | 14 |
| | Total | | | | | 8 | 10 |
| 2003 | Manufacturing | 35 | 44 | 27 | 11 | 10 | 6 |
| | Retail | 32 | 18 | 36 | 53 | 11 | 10 |
| | Services | 33 | 38 | 37 | 36 | 8 | 11 |
| | Total | | | | | 9 | 9 |
| 2008 | Manufacturing | 35 | 35 | 25 | 12 | 31 | 10 |
| | Retail | 26 | 16 | 33 | 48 | 33 | 28 |
| | Services | 39 | 49 | 42 | 40 | 24 | 22 |
| | Total | | | | | 29 | 21 |
| 2013 | Manufacturing | 40 | 47 | 24 | 11 | 19 | 9 |
| | Retail | 29 | 19 | 34 | 48 | 19 | 16 |
| | Services | 31 | 34 | 42 | 41 | 14 | 11 |
| | Total | | | | | 17 | 12 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The value added column excludes establishments reporting negative value added and/or capital stock net of depreciation. The last two columns show the percetange of establishments and workers excluded from a sector or the census in the final sample.

to the last column of table 3.2 indicates that 1993 is the year in which the smallest proportion establishments was removed from the sample. In contrast, 2008, the only year of the database occurring during a recession, is the year in which the largest share of establishments was excluded from the sample.

# Chapter 4

# Results

From here on, I refer to $ln(A_{si}M_s^{\frac{1}{\sigma-1}}/\overline{A_s})$ as $TFPQ$ and to $ln(TFPR_{si}/\overline{TFPR_s})$ as $TFPR$. I also denote $ln((1+\tau_{Ksi})/(1+\overline{\tau_{Ks}}))$ as the scaled capital distortion and to $ln((1-\tau_{Ysi})/(1-\overline{\tau_{Ys}}))$ as the scaled output distortion. Before proceeding, I provide heuristic interpretations of those variables using simple examples.

An establishment $si$ displaying a $TFPQ$ of 1.10 (0.90) is approximately 10% more (less) productive than the average establishment in industry $s$. From equation 3.16, an establishment $si$ facing a $TFPR$ of 1.10 (0.90) comprises a share of the industry $s$ value added that is 10% greater (smaller) in the efficient benchmark than in the actual data. An establishment exhibiting a scaled capital distortion of 1.10 (0.90) faces a cost of capital 10% greater (smaller) than its industry's average cost of capital. Similarly, an establishment confronting a scaled output distortion of 1.10 (0.90) faces restrictions to output 10% more beneficial (harmful) than the average output distortion in its industry.

Table 4.1 shows the distributions of $TFPQ$ and $TFPR$. $TFPQ$ is more dispersed in 2013 than in 1993. On average, formal establishments are more productive than semi-formal units, and semi-formal establishments are more productive than informal units. $TFPQ$ is more dispersed in the sample consisting of informal establishments.

$TFPR$ displays a higher average among formal establishments than among informal units.

Table 4.1: Distribution of $TFPQ$ and $TFPR$

| Variable | Sample | Statistic | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|---|---|
| $TFPQ$ | All | Mean | -2.55 | -2.47 | -2.48 | -2.87 | -2.83 |
| | | SD | 2.70 | 2.47 | 2.55 | 2.78 | 2.79 |
| | | 75-25 | 3.62 | 3.43 | 3.47 | 3.79 | 3.88 |
| | | 90-10 | 7.01 | 6.50 | 6.67 | 7.31 | 7.36 |
| | Formal | Mean | -0.23 | -0.42 | -0.15 | -0.37 | -0.49 |
| | | SD | 2.09 | 2.08 | 2.11 | 2.40 | 2.23 |
| | | 75-25 | 2.78 | 2.82 | 2.89 | 3.23 | 3.02 |
| | | 90-10 | 5.29 | 5.35 | 5.26 | 6.17 | 5.78 |
| | Semi-formal | Mean | -0.83 | -0.78 | -0.75 | -1.22 | -1.01 |
| | | SD | 2.07 | 2.06 | 2.04 | 2.39 | 2.26 |
| | | 75-25 | 2.70 | 2.77 | 2.72 | 3.17 | 2.99 |
| | | 90-10 | 5.23 | 5.33 | 5.26 | 6.13 | 5.77 |
| | Informal | Mean | -3.01 | -2.87 | -2.78 | -3.15 | -3.11 |
| | | SD | 2.63 | 2.37 | 2.48 | 2.73 | 2.75 |
| | | 75-25 | 3.56 | 3.30 | 3.41 | 3.73 | 3.83 |
| | | 90-10 | 6.83 | 6.23 | 6.51 | 7.17 | 7.24 |
| $TFPR$ | All | Mean | -0.25 | -0.22 | -0.18 | -0.27 | -0.19 |
| | | SD | 1.25 | 1.12 | 1.19 | 1.31 | 1.30 |
| | | 75-25 | 1.66 | 1.52 | 1.59 | 1.73 | 1.75 |
| | | 90-10 | 3.23 | 2.94 | 3.12 | 3.42 | 3.41 |
| | Formal | Mean | 0.40 | 0.39 | 0.41 | 0.33 | 0.39 |
| | | SD | 0.96 | 0.90 | 0.89 | 1.00 | 1.01 |
| | | 75-25 | 1.17 | 1.14 | 1.11 | 1.21 | 1.22 |
| | | 90-10 | 2.38 | 2.25 | 2.28 | 2.44 | 2.47 |
| | Semi-formal | Mean | 0.27 | 0.26 | 0.28 | 0.11 | 0.28 |
| | | SD | 0.97 | 0.91 | 0.91 | 1.06 | 1.02 |
| | | 75-25 | 1.20 | 1.16 | 1.16 | 1.30 | 1.27 |
| | | 90-10 | 2.43 | 2.28 | 2.28 | 2.64 | 2.54 |
| | Informal | Mean | -0.38 | -0.34 | -0.26 | -0.33 | -0.26 |
| | | SD | 1.27 | 1.13 | 1.22 | 1.33 | 1.32 |
| | | 75-25 | 1.72 | 1.56 | 1.65 | 1.79 | 1.81 |
| | | 90-10 | 3.31 | 2.97 | 3.19 | 3.49 | 3.47 |

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

Table 4.2: Distribution of the scaled capital and output distortions

| Variable | Sample | Statistic | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|---|---|
| *Scaled capital distortion* | All | Mean | 0.86 | 0.83 | 1.10 | 1.26 | 1.46 |
| | | SD | 1.54 | 1.42 | 1.62 | 1.69 | 1.78 |
| | | 75-25 | 1.97 | 1.97 | 2.32 | 2.48 | 2.60 |
| | | 90-10 | 3.92 | 3.71 | 4.26 | 4.42 | 4.66 |
| | Formal | Mean | 0.65 | 0.64 | 0.74 | 0.94 | 0.99 |
| | | SD | 1.55 | 1.49 | 1.51 | 1.61 | 1.67 |
| | | 75-25 | 2.02 | 1.98 | 1.96 | 2.21 | 2.14 |
| | | 90-10 | 3.93 | 3.78 | 3.92 | 4.08 | 4.22 |
| | Semi-formal | Mean | 0.55 | 0.66 | 0.82 | 0.99 | 1.10 |
| | | SD | 1.51 | 1.45 | 1.52 | 1.65 | 1.70 |
| | | 75-25 | 2.02 | 1.96 | 2.07 | 2.36 | 2.35 |
| | | 90-10 | 3.86 | 3.71 | 3.92 | 4.29 | 4.36 |
| | Informal | Mean | 0.93 | 0.87 | 1.15 | 1.30 | 1.52 |
| | | SD | 1.53 | 1.41 | 1.63 | 1.69 | 1.78 |
| | | 75-25 | 1.97 | 1.96 | 2.37 | 2.49 | 2.62 |
| | | 90-10 | 3.92 | 3.70 | 4.30 | 4.43 | 4.67 |
| *Scaled output distortion* | All | Mean | 0.59 | 0.59 | 0.56 | 0.62 | 0.73 |
| | | SD | 1.24 | 1.12 | 1.18 | 1.27 | 1.29 |
| | | 75 25 | 1.57 | 1.57 | 1.51 | 1.63 | 1.70 |
| | | 90 10 | 3.15 | 2.90 | 3.02 | 3.28 | 3.32 |
| | Formal | Mean | -0.13 | -0.11 | -0.11 | 0.04 | 0.02 |
| | | SD | 0.85 | 0.82 | 0.80 | 0.89 | 0.88 |
| | | 75-25 | 0.97 | 0.97 | 0.91 | 0.99 | 0.98 |
| | | 90-10 | 1.99 | 1.95 | 1.94 | 2.08 | 2.09 |
| | Semi-formal | Mean | -0.04 | 0.01 | 0.06 | 0.26 | 0.16 |
| | | SD | 0.84 | 0.82 | 0.80 | 0.94 | 0.89 |
| | | 75-25 | 0.99 | 0.97 | 0.95 | 1.09 | 1.05 |
| | | 90-10 | 2.02 | 1.96 | 1.94 | 2.29 | 2.17 |
| | Informal | Mean | 0.75 | 0.69 | 0.72 | 0.81 | 0.85 |
| | | SD | 1.26 | 1.13 | 1.20 | 1.30 | 1.31 |
| | | 75-25 | 1.66 | 1.54 | 1.59 | 1.70 | 1.76 |
| | | 90-10 | 3.24 | 2.94 | 3.09 | 3.36 | 3.39 |

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

However, the gap between those averages declines over time. Assuming that larger informal establishments display a higher $TFPR$, that increase is possibly due to the increase in the number of informal establishment accross all size categories in more recent years.

The distribution of the scaled capital and output distortions are shown in table 4.2. The scaled capital distortion, which increases the remuneration over capital costs ratio, is more advantageous for formal than for informal establishments. This was expected under the assumption that formal firms have easier access to credit markets and pay labor taxes, unlike informal firms.

The dispersion of the scaled capital distortion increases over time across all formality categories in all years except 1998. However, the dispersion is similar across formality categories at the beginning of the period. By 2013, in contrast, the scaled capital distortions is more dispersed among informal units than semi-formal or formal establishments.

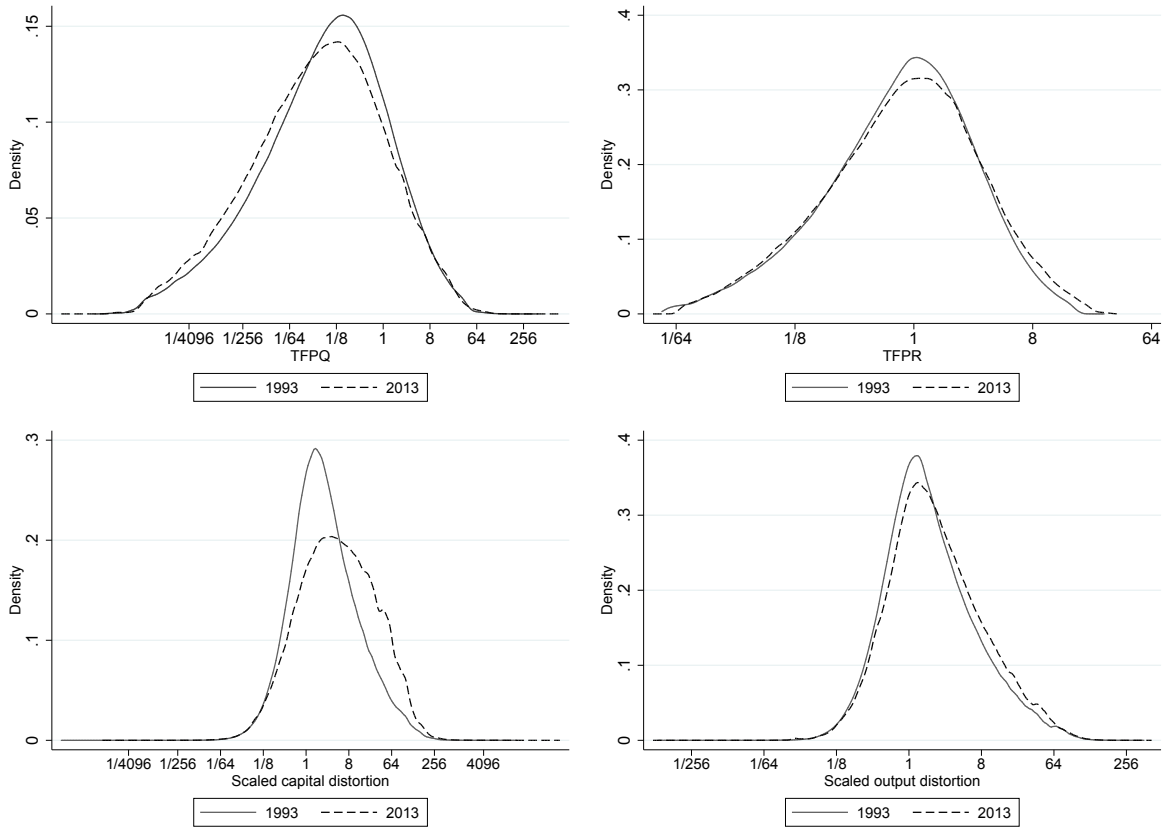
The scaled output distortion is harmfull for formal units and beneficial for informal establishments. This was expected under the assumption that formal establishments pay taxes while informal units do not. The average output distortion increases across all formality categories and in general. However, the gap between the average scaled output distortions of formal and informal establishments does not vary significantly over time. Finally, the dispersion of the scaled capital distortion increases only slightly over time.

Figure 4.1 shows the kernel densities of $TFPQ$ and $TFPR$ in 1993 and 2013. Both densities display lower peaks in the latter year, reflecting their increased dispersion. Moreover, the tails of their densities are longer in 2013 than in 1993.

On the one hand, the left-hand side of the distribution of $TFPQ$ is more massive in 2013 than in 1993. This suggests that policies favoring the survival of inefficient, possibly informal establishments are more prevalent in the latter year. Establishment survival rates by year of birth are shown in B. On the other, $TFPR$ displayed a more massive right-hand side in 2013. This indicates that a proportionally larger number of firms are inefficiently small in 2013 as compared to 1993.

Figure 4.1 also displays the kernel densities of the scaled capital and output distortions.

Figure 4.1: *Densities of TFPQ, TFPR, and the scaled capital and output distortions*




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

How to interpret the graphs in this figure? Both densities display a lower peak and a more massive right-hand side in 2013 than in 1993. This means that a proportionally larger number of establishments faced above-average capital distortions in the latter year. It also implies that a larger proportion of establishments was implicitly subsidized in 2013 than in 1993.

Table 4.3 shows the correlations between $TFPR$, $TFPQ$, and the scaled capital and output distortions in 1993, 2003 and 2013. The correlations between most pairs of variables are similar across time. There is a strong positive correlation between $TFPQ$ and $TFPR$, indicating that more (less) productive establishments use too few (many) resources. More productive establishments also face higher capital costs and are implicitly or explicitly taxed.

Establishments that would expand in the efficient benchmark do not seem to face much higher capital costs due to distortions. In particular, the correlation between $TFPR$ and the scaled capital distortion is close to zero in 2013. The correlation between $TFPR$ and the scaled

Table 4.3: Correlations between measures of productivity and distortions

| | | *TFPQ* | *TFPR* | *S. capital distortion* | *S. output distortion* |
|---|---|---|---|---|---|
| 1993 | *TFPQ* | 1 | | | |
| | *TFPR* | 0.89 | 1 | | |
| | *S. capital distortion* | 0.25 | 0.04 | 1 | |
| | *S. output distortion* | -0.87 | -0.88 | 0.25 | 1 |
| 2003 | *TFPQ* | 1 | | | |
| | *TFPR* | 0.89 | 1 | | |
| | *S. capital distortion* | 0.29 | 0.05 | 1 | |
| | *S. output distortion* | -0.84 | -0.88 | 0.25 | 1 |
| 2013 | *TFPQ* | 1 | | | |
| | *TFPR* | 0.89 | 1 | | |
| | *S. capital distortion* | 0.27 | 0.01 | 1 | |
| | *S. output distortion* | -0.84 | -0.90 | 0.26 | 1 |

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

capital distortion is strongly negative, ranging from -0.87 to -0.84. This means that inefficiently small establishments tend to be explicitly or implicitly taxed.

The correlation between the output and capital distortions is just above 0.20. This means that establishments facing higher capital costs tend to be implicitly subsidized, possibly due to their status as informal establishments. Conversely, that correlation means that establishments facing lower capital costs tend to be implicitly or explicitly taxed, possibly because they are formal.

## 4.1 Gains from reallocation

Equation 3.14 associates higher variances of $TFPR_{si}$ and $(1 + \tau_{Ksi})$ with lower $TFP$ at the industry level. The previous subsection shows that $TFPR$ and the scaled capital distortion become more dispersed over time. Therefore, from equation 3.16 it is expected for gains from reallocation to grow from 1993 to 2013. Table 4.4 summarizes the gains from reallocation at the aggregate and sector levels in all years.

Panel A from figure 4.2 displays the gains from reallocation at the aggregate level. Gains from reallocation remain virtually constant from 1993 to 2003 and increase in both 2008 and

Table 4.4: Gains from reallocation

| | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| Actual value added shares | | | | | |
| Aggregate | 135% | 135% | 135% | 149% | 183% |
| Manufacturing | 94% | 89% | 86% | 98% | 143% |
| Retailing | 195% | 182% | 194% | 213% | 245% |
| Services | 121% | 139% | 131% | 153% | 176% |
| 1998 value added shares | | | | | |
| Aggregate | 141% | 135% | 138% | 161% | 187% |
| Manufacturing | 92% | 89% | 86% | 97% | 146% |
| Retailing | 195% | 182% | 194% | 213% | 244% |
| Services | 140% | 139% | 138% | 181% | 173% |

*Source*: Author's calculations based on data from INEGI.
*Note*: I chose 1998 instead of 1993 value added shares because in 1998 there as many industries as in the following years, while in 1993 there 5 less industries than in the following years.

2013. Gains from reallocation increased from 135% in 2003 to 183% in 2013.

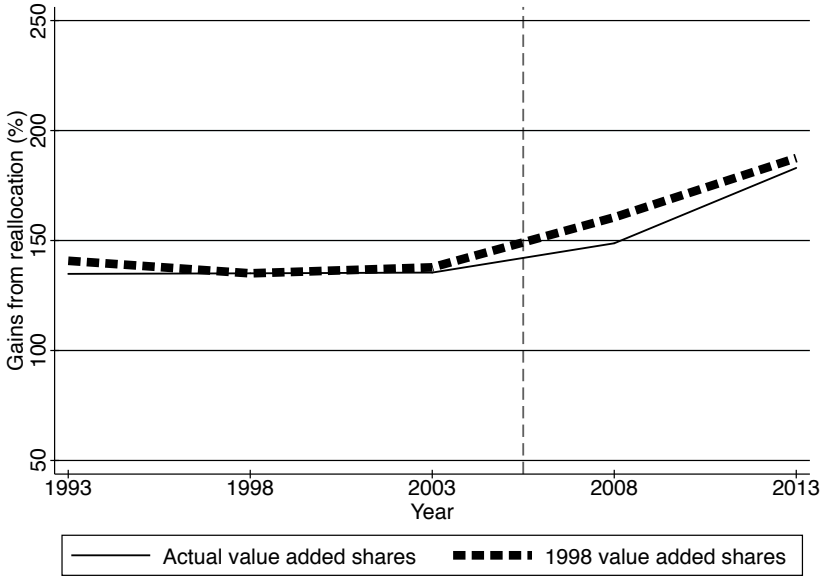
Panel A from figure 4.2 also shows counterfactual gains from reallocation, calculated under the assumption that the industry and sector value added shares remain constant at their 1998 levels. The purpose of this counterfactual exercise is to explore the following question. Is an increase in the share of value added of the industries more affected by misallocation responsible for the increase in the aggregate gains from reallocation? Counterfactual gains from reallocation are slightly larger than those with the actual value added shares. However, the increase in the gains from reallocation in the last two years of the sample does not seem to stem from changes in the distribution of value added across industries or sectors.

Panel B from figure 4.2 shows the gains from reallocation at the sector level. Two results are clear. First, misallocation is greater in retailing than in services and greater in services than in manufacturing. This conclusion resembles the findings of Dias et al. (2016) for Portugal and Busso, Fazio, and Levy (2012) for Mexico. Second, gains from reallocation in the three sectors vary slightly from 1993 to 2003 but increase in 2008 and 2013.

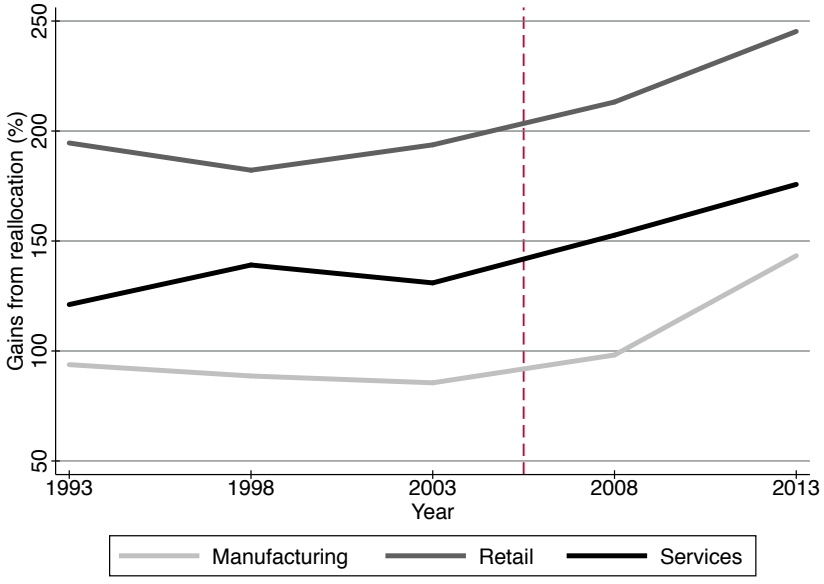
The counterfactual gains from reallocation are the same as the actual gains from reallocation in the manufacturing and retailing sectors. However, they are slightly larger in the services

Figure 4.2: *Gains from reallocation*

Panel A: Aggregate gains from reallocation




Panel B: Gains from reallocation by sector




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

sector in 1994, 2003 and 2008. This means that the industries in that sector showing the largest gains from reallocation represent a lower share of value added in those years than in 1998.

Tables showing the proportion of establishments than grow or shrink after reallocation are shown in appendix C. This tables classify establishments by formality status and size categories and show that a larger proportion of informal than formal establishments would shrink in the efficient benchmark. The table also shows that within the largest establishments, those employing 1000 or more workers, the very largest units, which on average employ over 2000 workers, are more likely to shrink after reallocation.

## 4.2 Misallocation and establishment characteristics

This subsection explores the relationship between misallocation and three establishment characteristics: number of workers, age and formality status. The reasons why I focus on this categories are the following. First, larger firms tend to be more productive, as documented by Busso, Madrigal, and Pagés (2012), among others. Following Bento and Restuccia (2016), who argue that more productive firms may face larger distortions, we expect larger establishments to grow relatively more than smaller establishments after the efficient reallocation of resources.

Second, perhaps some firms have invested in increasing their physical productivity but have not increased their output yet. These firms would display a higher $TFPR$ than similar firms that did not undergo investments in productivity. There may be a relation between the timing of these investments and age.

Third, notice that formal firms pay several taxes in addition to contributions to social security. Meanwhile, informal firms do not pay any contributions social security, but their non-salaried workers are entitled to a bundle of social security services provided by government agencies. Thus informal firms, and to a lesser extent semi-formal firms, are implicitly subsidized. Therefore we would expect semi-formal and informal establishments to be inefficiently large in comparison with formal units even after taking taxation into account.

Table 4.5: Regression of $TFPR$ on establishments size measured as the number of workers

| Number of workers | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| 2 | 0.10 | 0.10 | -0.02 | -0.05 | 0.05 |
| 3 | 0.25 | 0.28 | 0.13 | 0.01 | 0.25 |
| 4 | 0.35 | 0.38 | 0.23 | 0.10 | 0.37 |
| 5 to 9 | 0.50 | 0.54 | 0.41 | 0.29 | 0.56 |
| 10 to 19 | 0.64 | 0.65 | 0.56 | 0.53 | 0.73 |
| 20 to 49 | 0.65 | 0.64 | 0.56 | 0.65 | 0.74 |
| 50 to 99 | 0.53 | 0.55 | 0.47 | 0.57 | 0.52 |
| 100 to 249 | 0.43 | 0.47 | 0.41 | 0.47 | 0.47 |
| 250 to 499 | 0.38 | 0.43 | 0.38 | 0.40 | 0.36 |
| 500 to 999 | 0.33 | 0.38 | 0.39 | 0.35 | 0.35 |
| 1000 or more | 0.14 | 0.34 | 0.29 | 0.25 | 0.28 |
| Industry F.E. | Yes | Yes | Yes | Yes | Yes |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-squared | 0.051 | 0.043 | 0.030 | 0.027 | 0.052 |
| Ajusted R-squared | 0.051 | 0.043 | 0.030 | 0.027 | 0.052 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The excluded category is establishments with 1 worker. All coefficients are significant at the 0.01% level, except for the coefficient of the 1000 or more workers dummy in the 1993 regression. This coefficient is significant at the 5% level.

Table 4.5 shows the results of regressions of $TFPR$ on bins of the number of workers. In general, establishments with only one worker use too many resources. Larger establishments tend to use too few resources, but this relationship may mask the fact that formal firms tend to be larger regarding capital stock and employment. We see an inverted U-shaped relation between $TFPR$ and size. In all years, the establishments with the most inefficiently small scale are those employing 10 to 19 or 20 to 49 workers.

Table 4.6 displays the output of regressions on age bins. Establishments with 5 to 9 years display the highest $TFPR$. While the oldest establishments face the most favorable distortions from 1993 to 2003, the youngest establishments face the most beneficial distortions in 2008 and 2013.

The regressions of $TFPR$ on age are related to an observation central to Hsieh and Klenow (2014). These authors use data from the Mexican Economic Censuses from 1998 to 2008. They document that Mexican manufacturing establishments stop growing once they reach 25 years of

Table 4.6: Regression of $TFPR$ on age

| Age in years | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| 5 to 9 | 0.16 | 0.04 | 0.09 | 0.20 | 0.12 |
| 10 to 14 | 0.07 | -0.03 | 0.03 | 0.15 | 0.07 |
| 15 to 19 | 0.05 | -0.08 | 0.03 | 0.15 | 0.07 |
| 20 to 24 | 0.05 | -0.13 | -0.01 | 0.14 | 0.06 |
| 25 or more | -0.07 | -0.17 | -0.02 | 0.11 | 0.10 |
| Industry F.E. | Yes | Yes | Yes | Yes | Yes |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-squared | 0.037 | 0.022 | 0.016 | 0.019 | 0.033 |
| Ajusted R-squared | 0.037 | 0.022 | 0.016 | 0.019 | 0.033 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The excluded category is establishments 0 to 4 years old. All coefficients are significant at the 0.01% level, except for the coefficient of the 20 to 24 years old dummy in the 2003 regression. This coefficient is significant at the 5% level.

age. Table 4.6 indicates that in 1998 firms with 25 or more years of age were inefficiently large. However, starting in 2008, their efficient size is greater rather than smaller than their actual size.

The results of the regressions of $TFPR$ on formality status occupy table 4.7. These regressions include size and age controls because formal firms tend to be both larger and older. However, despite the inclusion of those controls, the results are as expected. Informal firms are inefficiently large, while semi-formal and formal firms are inefficiently small. Also, formal firms face greater distortions than semi-formal firms. However, we must look at the relationship between formality status and the scaled capital and output distortions before discarding that the relationship between $TFPR$ and formality status stems from tax payments.

It is worth noting that the coefficients of the dummies of formality status shown in table 4.7 may not be directly comparable across years. This is because of the characteristics of informal firms in the sample change from the initial to the latter periods. Medium-sized and large informal establishments are almost inexistent in 1993 but become more numerous in more recent years.

The regressions on the scaled capital and output distortions on establishment size, age, and formality status are shown in appendix D. The scaled capital distortion does not have a clear relationship with size, possibly because formality is the main driver of the capital distortion. As

Table 4.7: Regression of $TFPR$ on formality status

| Formality status | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| Semi-formal | 0.58 | 0.60 | 0.56 | 0.41 | 0.52 |
| Formal | 0.75 | 0.73 | 0.71 | 0.64 | 0.63 |
| Industry F.E. | Yes | Yes | Yes | Yes | Yes |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-squared | 0.086 | 0.068 | 0.045 | 0.042 | 0.056 |
| Ajusted R-squared | 0.086 | 0.068 | 0.045 | 0.042 | 0.056 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The excluded category is informal. Quintic terms for number of workers and age are included as controls. All coefficients are significant at the 0.01% level.

expected, the older and formal establishments face more beneficial capital distortions.

The scaled output distortion is the most harmful for medium-sized establishments, those employing 10 to 49 workers. This distortion also tends to negatively affect older establishment more, possibly because formal establishments tend to be older. Formal establishments display a punishing output distortion as compared to informal establishments, but probably simply due to tax payments.

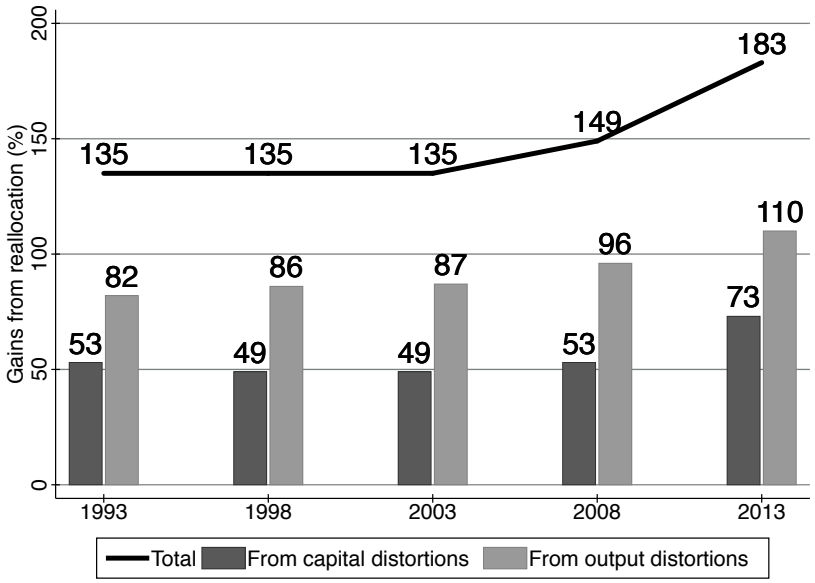
## 4.3 Why has misallocation increased?

I found that the gains from reallocation in Mexico remained stable from 1993 to 2003, and started increasing in 2008. In this section, I attempt to explain why misallocation increased.

Figure 4.3 shows the decomposition of the gains from reallocation into the isolated effects of the capital and output distortions [1]. The gains from correcting the capital distortion remain approximately constant until 2008 and increase in 2013. Interestingly, the gains from reallocation stemming from equalizing the output distortion increase year by year. This is consistent with the observed increase, also from year to year, of the shares of value added, net capital stock and employment corresponding to informal establishments. Assuming that the output distortion stems mainly from taxation to formal firms, it could be of interest to analyze the gains from

[1] I propose a straighforward method to calculate the effect of equalizing the capital distorion within industries on the gains from reallocation using, in the following order, equations 3.18, 3.21, 3.22, 3.19, 3.20, 3.10 and 3.16.

Figure 4.3: *Gains from reallocation*




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

reallocation after equalizing the capital distortion across firm within industries but keeping the observed output distortions at the firm level.

The exploration continues with an analysis of regressions of $TFPQ$, $TFPR$, and the scaled capital and output distortions on interaction terms between formality status dummies and polynomials of size and age. Table 4.8 summarizes information about those regressions. The graphs of the predicted values and confidence intervals from those regressions for formal establishments in 1993 and 2013 are shown in this section. The graphs of the predicted values for formal units in all years are shown in appendix E.

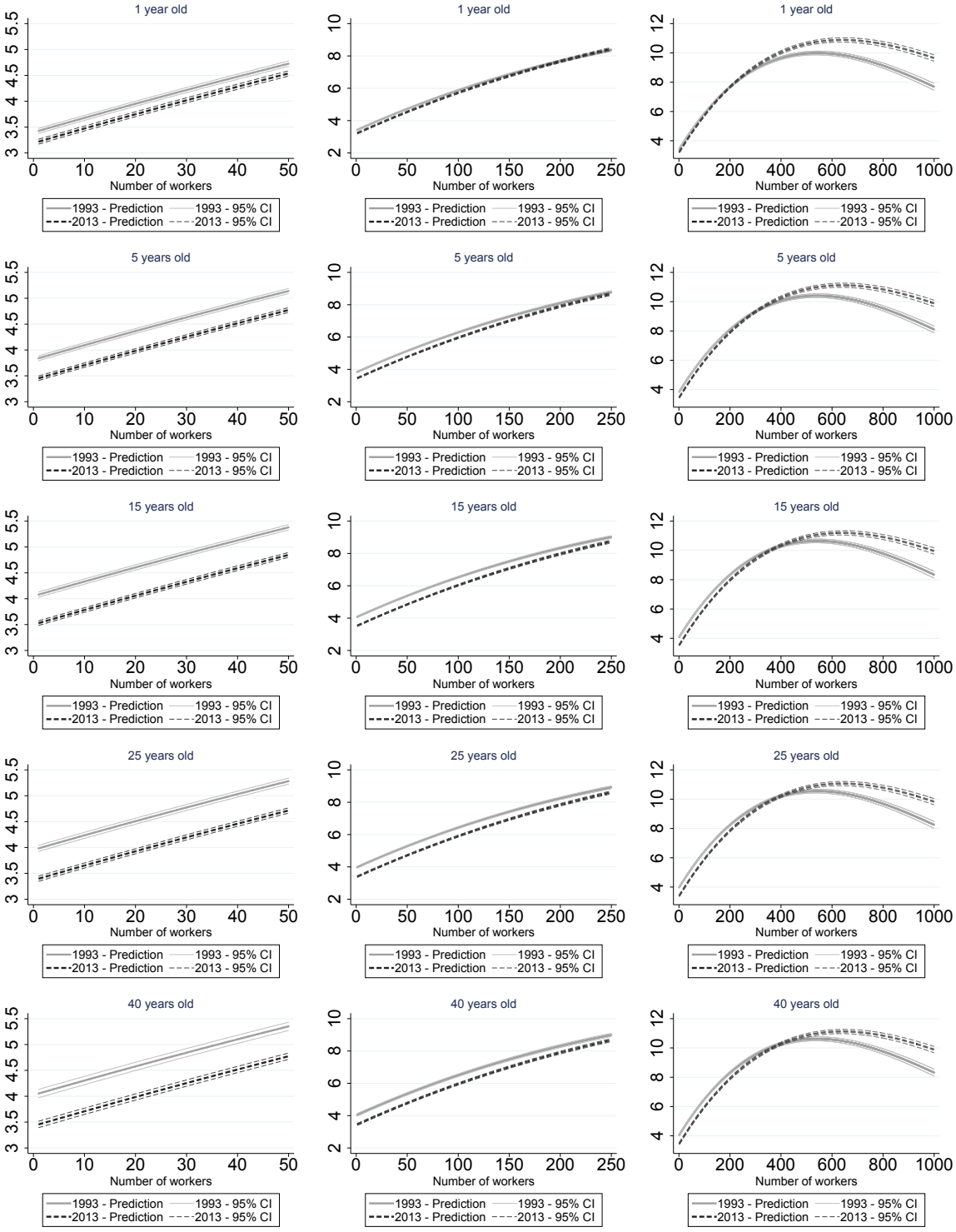
Figure 4.4 displays the predicted values of $TFPQ$ for formal establishments in 1993 and 2013. Formal establishments with 50 or fewer workers appear less productive in comparison with the smallest informal units in 2013 than in 1993. However, formal units with 500 or more workers are on average more productive relative to the smallest informal establishments in the more recent year. Figure 4.5 shows that informal establishments employing 2 or more workers are less productive in 2003 than in 1993 in comparison with the smallest informal establishments.

Table 4.8: Regressions on $TFPQ$, $TFPR$, and the scaled capital and output distortions

| | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| *TFPQ* | | | | | |
| Semi-formal intercept | 3.206 | 3.236 | 2.688 | 2.454 | 2.730 |
| Formal intercept | 3.989 | 3.570 | 3.242 | 3.374 | 3.378 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.260 | 0.259 | 0.190 | 0.177 | 0.191 |
| Adjusted R-2 | 0.260 | 0.259 | 0.190 | 0.177 | 0.191 |
| *TFPR* | | | | | |
| Semi-formal intercept | 0.869 | 0.857 | 0.777 | 0.678 | 0.782 |
| Formal intercept | 1.077 | 1.005 | 0.992 | 0.942 | 0.966 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.089 | 0.073 | 0.046 | 0.043 | 0.058 |
| Adjusted R-2 | 0.089 | 0.073 | 0.046 | 0.043 | 0.058 |
| *Scaled capital distortion* | | | | | |
| Semi-formal intercept | -0.347 | -0.069 | -0.140 | -0.165 | -0.371 |
| Formal intercept | -0.328 | -0.165 | -0.330 | -0.315 | -0.560 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.059 | 0.067 | 0.059 | 0.053 | 0.076 |
| Adjusted R-2 | 0.059 | 0.067 | 0.059 | 0.053 | 0.076 |
| *Scaled output distortion* | | | | | |
| Semi-formal intercept | -1.014 | -0.883 | -0.830 | -0.734 | -0.915 |
| Formal intercept | -1.217 | -1.069 | -1.128 | -1.067 | -1.168 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.109 | 0.088 | 0.069 | 0.057 | 0.073 |
| Adjusted R-2 | 0.109 | 0.088 | 0.069 | 0.057 | 0.073 |

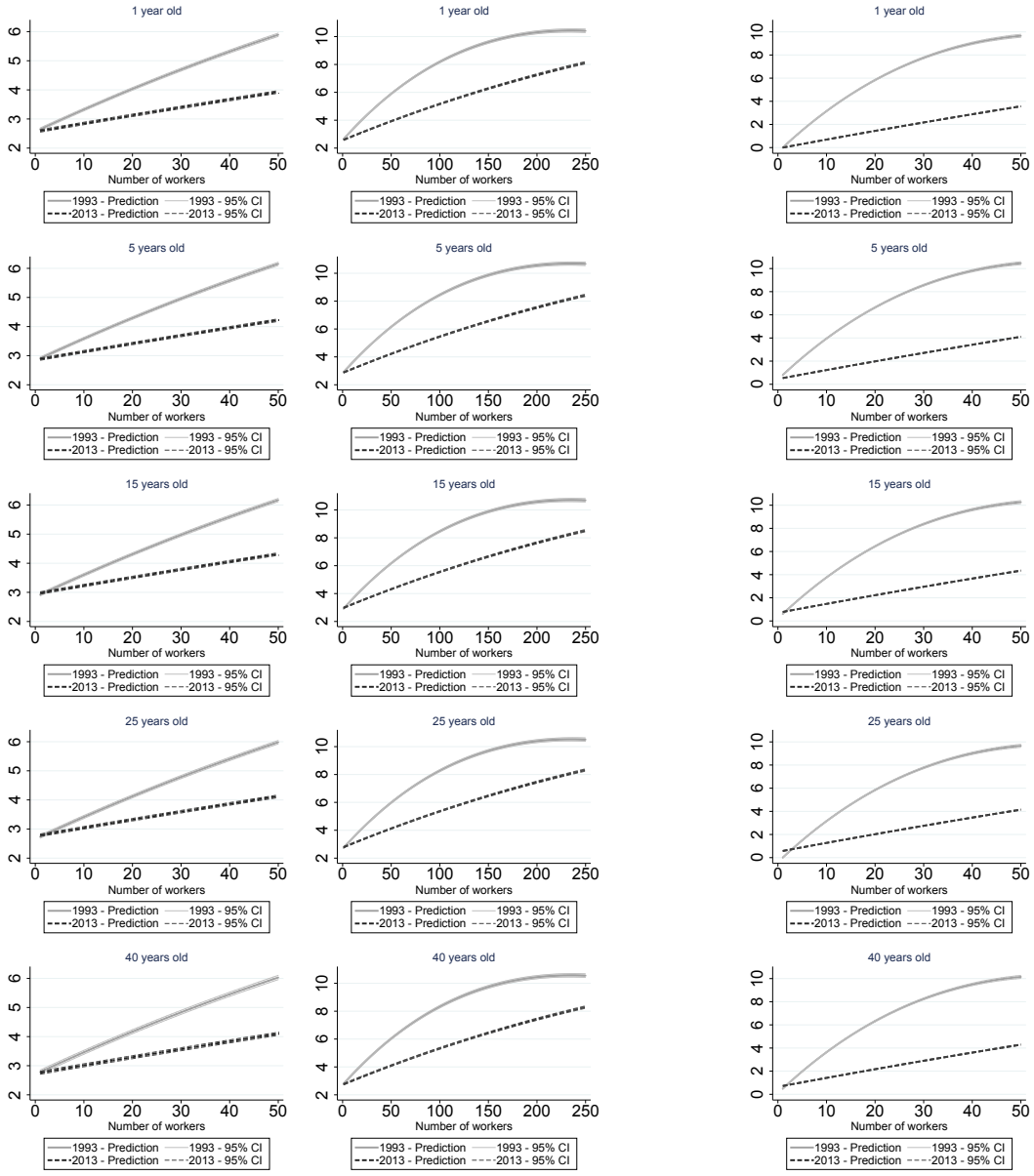
*Source*: Author's calculations based on data from INEGI.
*Note*: The excluded intercept is the dummy for informal establishments. Quintic terms for the number of workers and age are included interacted with formality status dummies as controls. All regressions include industry fixed effects. Except for some industry dummies in some regressions, all coefficients are significant at the 0.01% level. It is worth noticing that informal establishments with 20 or more workers are less than 10% of informal establishments in 1993 and 1998 so that the fit of the regressions for large informal establishments in those years is probably poor.

Figure 4.4: *Predicted TFPQ for formal establishments in 1993 and 2013*




*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure 4.5: *Predicted TFPQ for semi-formal and informal establishments in 1993 and 2013*




(a) Semi-formal (b) Informal

*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

The predicted values of $TFPR$ for formal establishments in 1993 and 2013 are shown in figure 4.6. Formal establishments 5 or fewer years old and with less than 50 workers are just as constrained by distortions in 2013 than in 1993. However, within that size category, formal units 15 or more years old display lower $TFPR$ the more recent year. Interestingly, formal establishments employing 100 to around 800 workers, especially if their 15 years old or less, would grow more after reallocation in 2013 than in 1993.

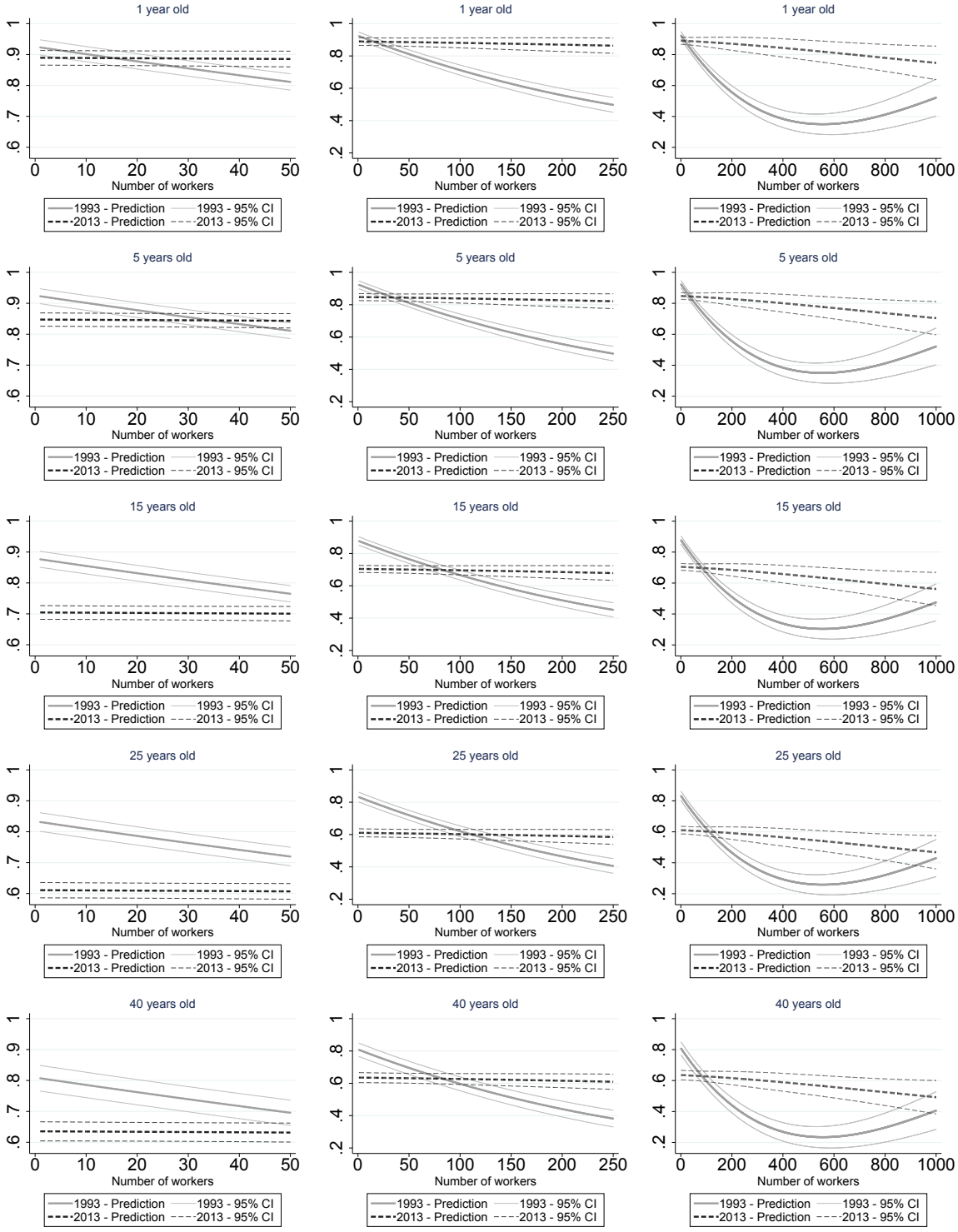
Figure 4.7 shows that informal establishments face approximately the same $TFPR$ in 2013 than in 1993. We expect gains from reallocation to increase because the shares of value added, capital and employment belonging to informal establishments grow over time.

Why is $TFPR$ higher for large formal establishments in 2013 than in 1993? It is necessary to observe the evolvement of the scaled capital and output distortions to answer that question. Figure 4.8 displays the predicted values and confidence intervals of the scaled capital distortion in 1993 and 2013. All formal establishments except the largest and oldest face capital distortions that are more beneficial ($\tau_{Ksi} < 0$) in comparison with the capital distortions of the smallest informal establishments in 2013 than in 1993. This figure is consistent with figure 4.3 the regression results from table D.3, which shows that the gap between the average capital distortions faced by informal and formal establishments increased in 2013.

The regressions on the scaled capital distortion may be capturing one or two of the following possible effects. First, the improvement in the access to credit by formal firms, and, second, an increase in the subsidies to labor that increase the marginal revenue product of capital of informal establishments. Disentangling those effects is beyond the scope of this thesis. However, it is worth noting that the implementation of social programs in the middle of the previous decade has slightly increased the probability to obtain informal employment for workers with 9 or fewer years of education (see Azuara and Marinescu (2013)).

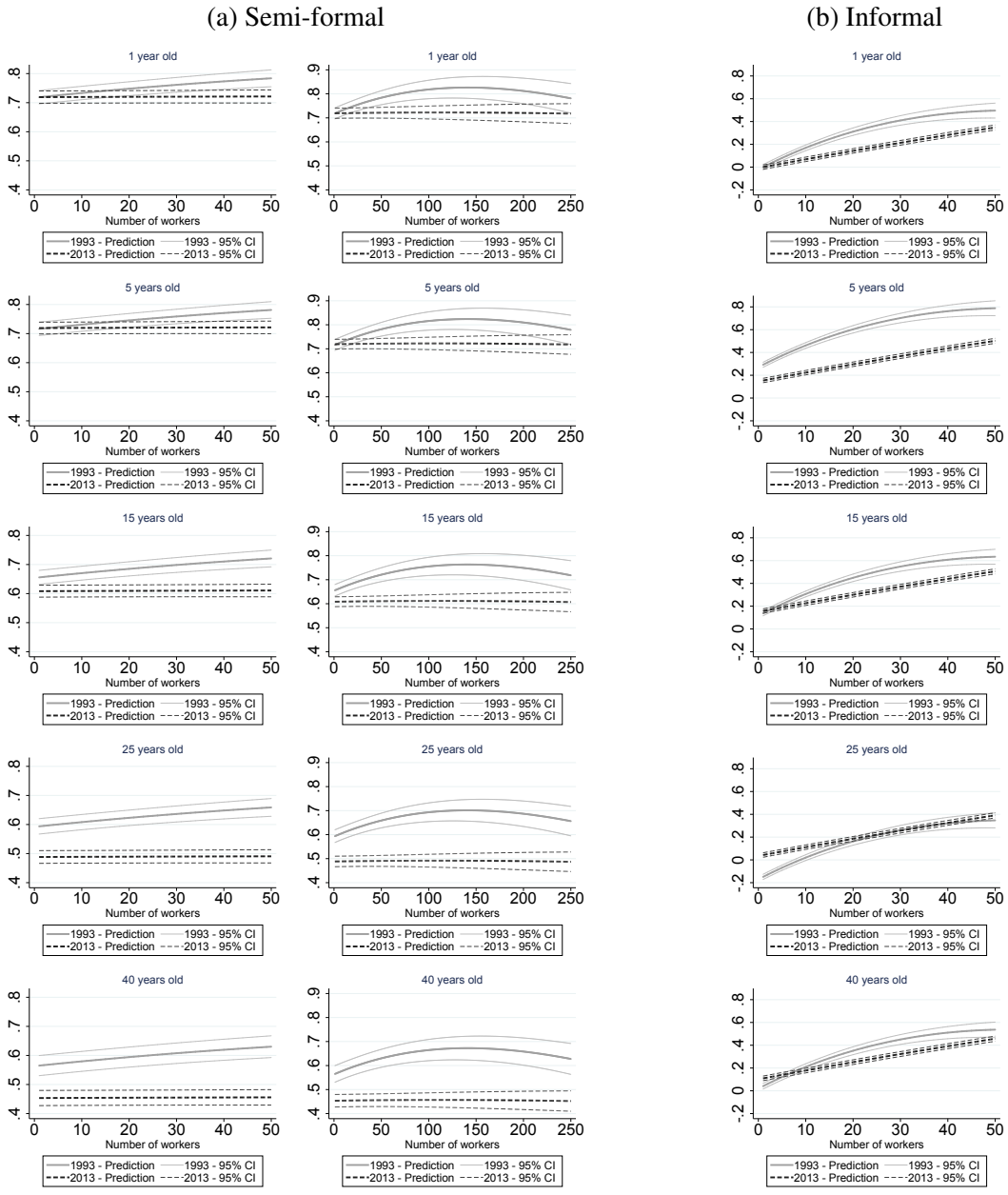
The predicted values and confidence intervals of the scaled output distortion for formal units are shown in Figure figure 4.8. Clearly, formal establishments of all sizes and ages face more harmful output distortions ($\tau_{Ysi} > 0$) in 2013 than in 1993. Recall, however, that this increase

Figure 4.6: *Predicted TFPR for formal establishments in 1993 and 2013*



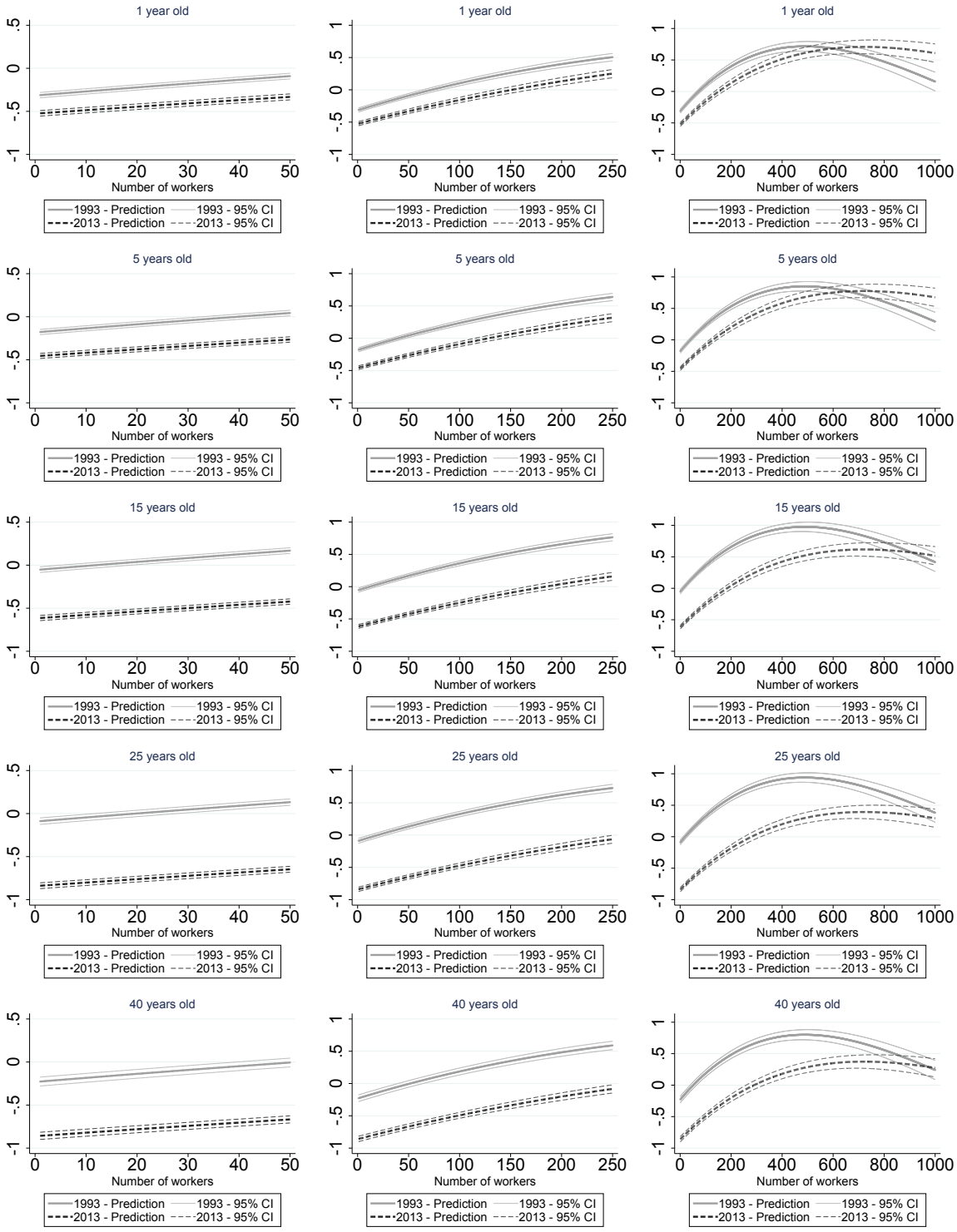

*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure 4.7: *Predicted TFPR for semi-formal and informal establishments in 1993 and 2013*



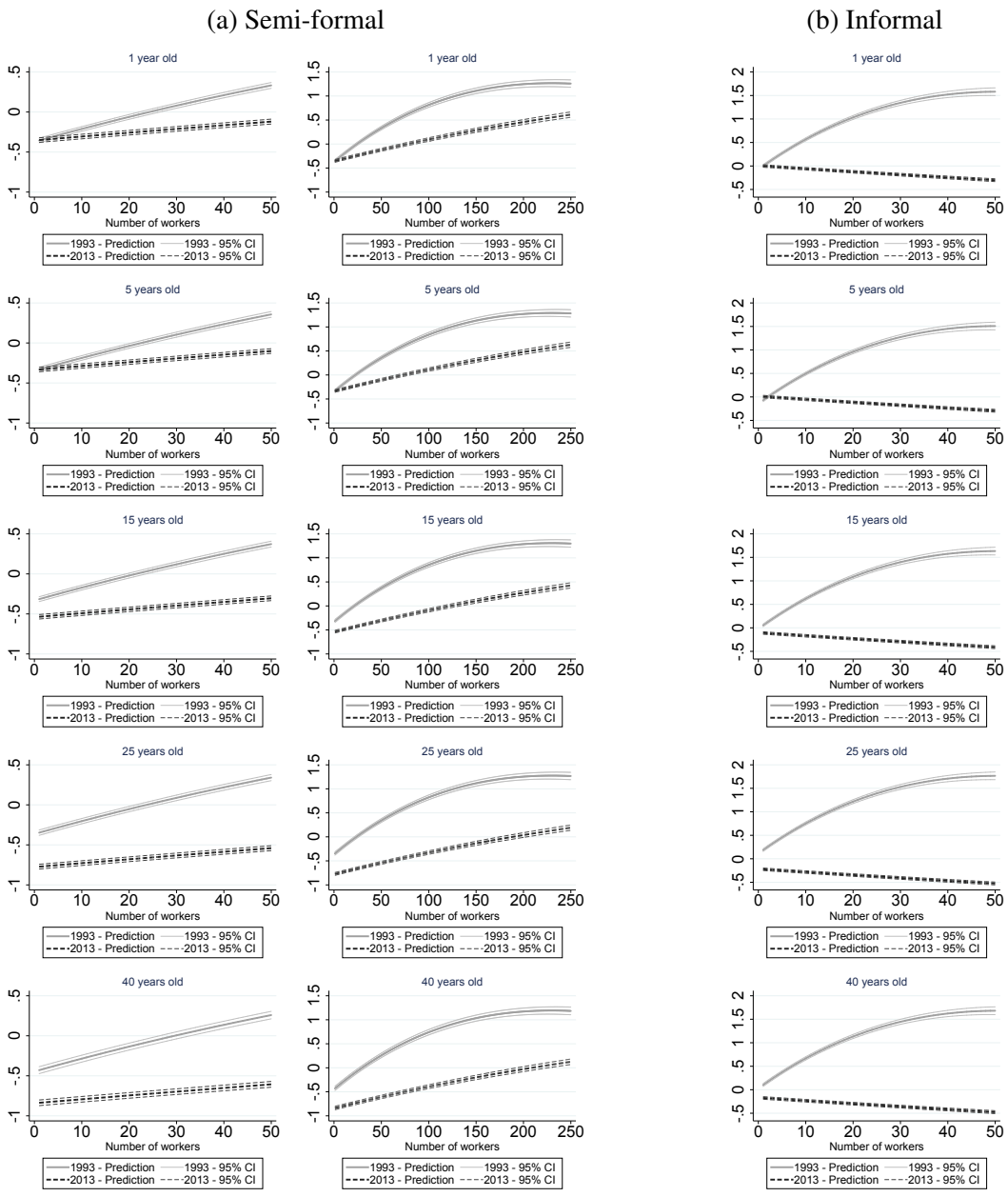

*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure 4.8: *Predicted scaled capital distortion for formal establishments in 1993 and 2013*




*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure 4.9: *Predicted scaled capital distortion for semi-formal and informal establishments in 1993 and 2013*

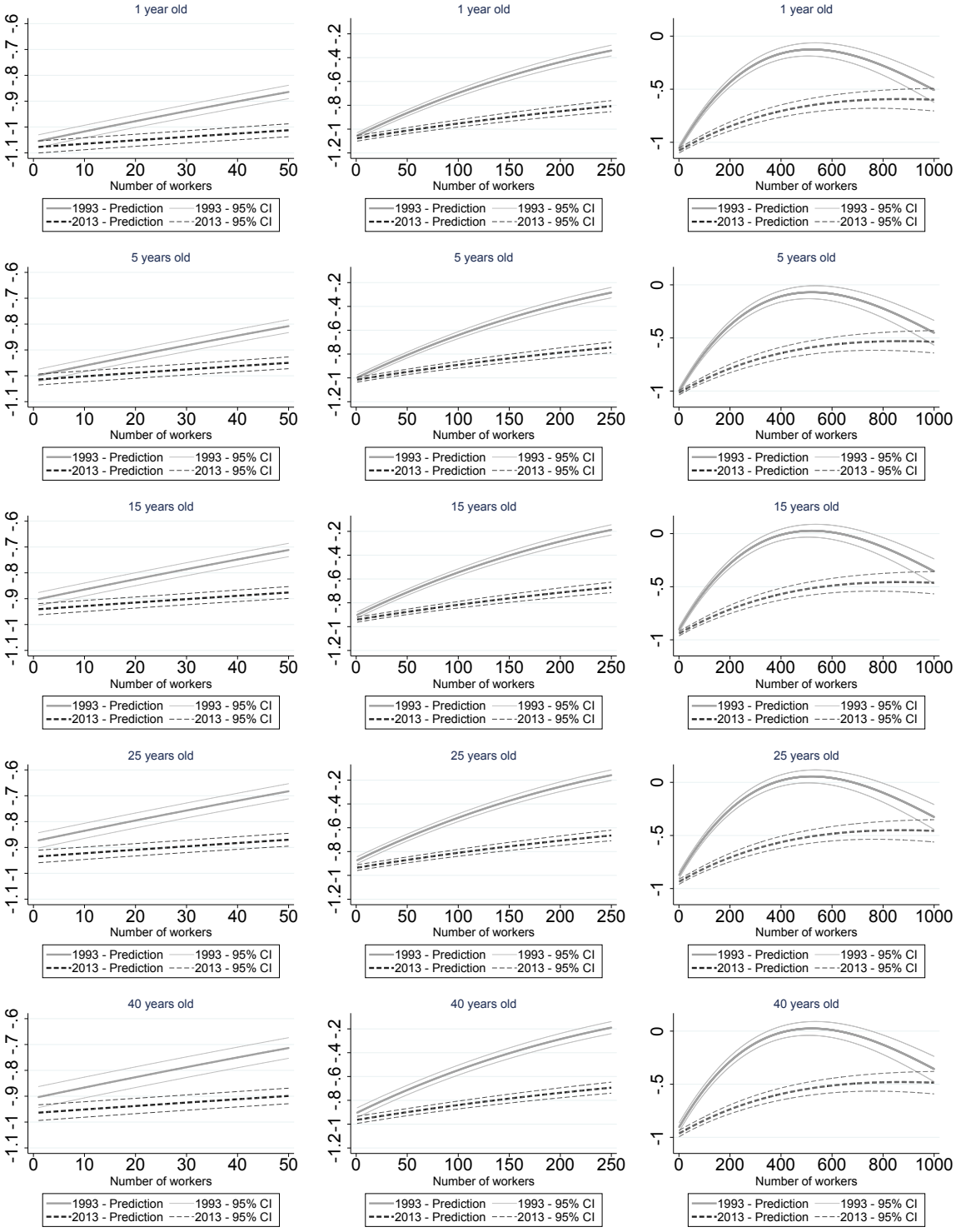



*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

in the output distortion faced by formal establishments may simply reflect higher taxation to revenue or profits instead of idiosyncratic distortions such as higher transportation costs. Why has the average difference between the output distortions of formal and informal establishments shown by the regression results in table D.4 decreased slightly instead of increasing? The reason is that informal establishments employing 2 or more workers face output distortions which are less beneficial than before, as shown by figure 4.11.

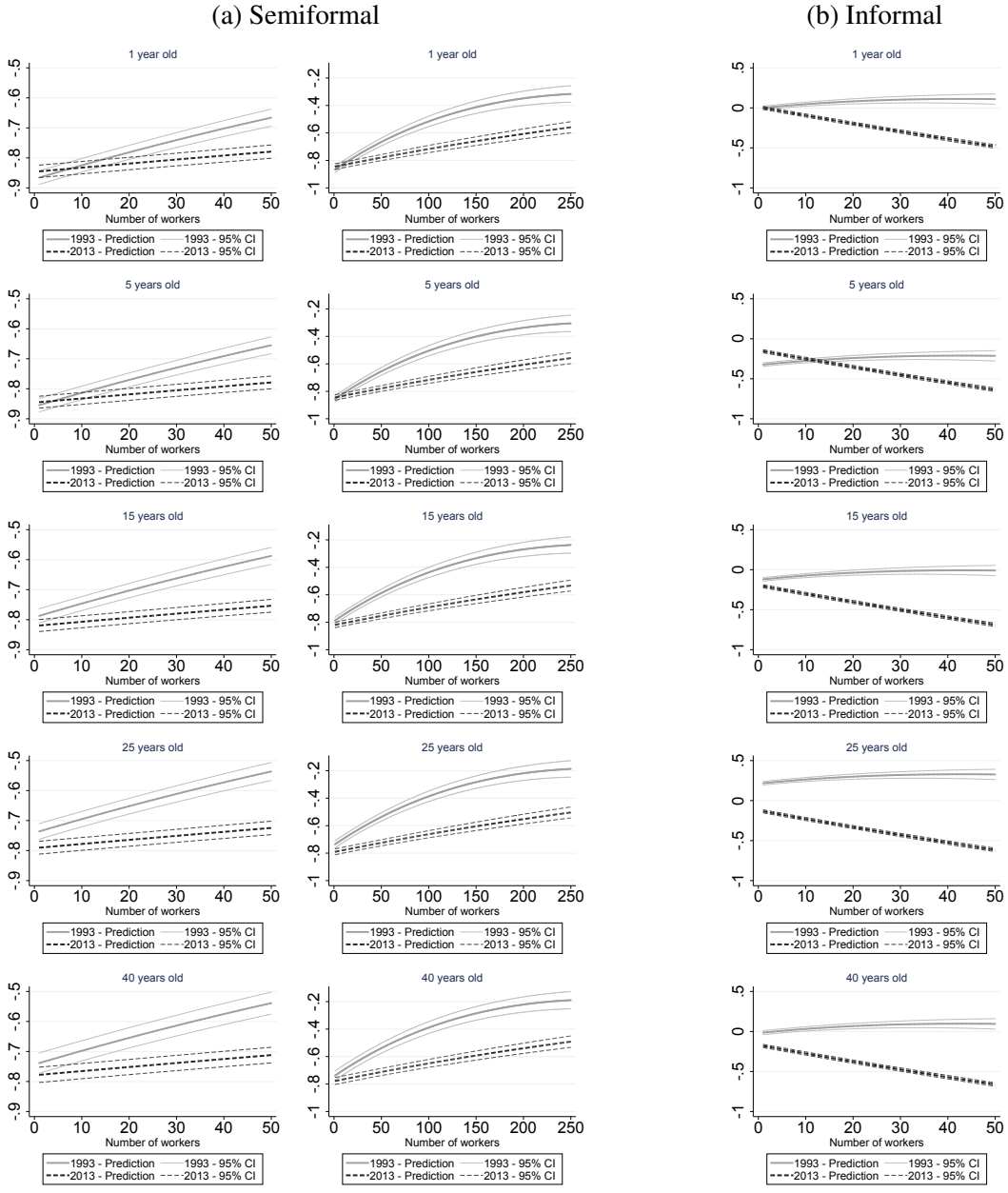
Misallocation seems to have increased for three main reasons. First the shares of value added, net capital and employment corresponding to informal establishments have increased over time, leading to a greater availability of resources to be reallocated from informal units to formal establishments. Second, the output distortions that formal establishments face in comparison with informal establishments has increased, though this increase may simply reflect higher taxation instead of implicit of implicit subsidies to informal establishments. Third, the effect of capital distortions on the gains from reallocation remained approximately constant from 1993 to 2008 but increased in 2013. The increase of that effect of that distortion could be stemming from easier access to credit by formal firms, higher labor taxes for formal establishments, greater implicit subsidies for informal units to hire labor or some combination of these factors.

Figure 4.10: *Predicted scaled output distortion for formal establishments in 1993 and 2013*




*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure 4.11: *Predicted scaled output distortion for semi-formal and informal establishments in 1993 and 2013*




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

# Chapter 5

# Conclusion

This thesis documents that the share of value added, net capital stock net of depreciation and workers corresponding to informal establishments increased from 1993 to 2013. It also finds that the potential gains from reallocation remained stable from 1993 to 2003, and increased in 2008 and 2013. The effects of output distortions on the gains from reallocation grow in every analyzed year since 1993, possibly due to changes in the taxation to formal establishments or incentives for informality. The effects of capital distortions varied little from 1993 to 2008 but increased in 2013.

Further research is required to link the increase in the gains from reallocation with specific policies and establishment characteristics which this thesis did not analyze, such as location and exporter status. It would be interesting to study if the largest formal establishments that have become more productive and constrained by misallocation are also exporters. Additionally, further analyzing the effects of capital distortions on aggregate $TFP$ and disentangling idiosyncratic output distortions from the effects of taxation could yield interesting results, given that in Mexico a majority of the establishments are informal.

# Appendix A

# Derivation of additional equations

## A.1 TFP at the industry level

From the demand for $Y_{si}$ in equation 3.3, the zero profit condition $P_s Y_s = \sum_{i=1}^{M_s} P_{si} Y_{si}$, and the definition of industry $TFPR$ in equation 3.10, the inverse of $P_s$ is written as:

$$\begin{aligned}
\frac{1}{P_s} &= \left( \sum_{i=1}^{M_s} \left( \frac{1}{P_{si}} \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \\
&= \left( \left( \sum_{i=1}^{M_s} \left( \frac{1}{P_{si}} \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \cdot \frac{\overline{TFPR_s}}{\overline{TFPR_s}} \right) \\
&= \left( \sum_{i=1}^{M_s} \left( A_{si} \frac{\overline{TFPR_s}}{TFPR_{si}} \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \frac{1}{\overline{TFPR_s}} \\
&= \left( \sum_{i=1}^{M_s} \left( A_{si} \frac{\overline{TFPR_s}}{TFPR_{si}} \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \frac{K_s^{\alpha_s} L_s^{1-\alpha_s}}{P_s Y_s}.
\end{aligned}$$

Then isolate $Y_s$ to obtain industry output:

$$Y_s = \left( \sum_{i=1}^{M_s} \left( A_{si} \frac{\overline{TFPR_s}}{TFPR_{si}} \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}} K_s^{\alpha_s} L_s^{1-\alpha_s},$$

from where the definitions of industry $TFP$ in equation 3.13 stems.

## A.2 Allocation of capital and labor across industries

Consider the first order conditions of the cost minimization problem of the final good producing firm, and the zero profit condition $PY = \sum_{s=1}^{S} P_s Y_s$. From them, obtain the optimal expenditure on intermediate good $s$:

$$P_s Y_s = \theta_s Y,$$

where $P = 1$ because the final good is the numeraire. Then, from the equation above and the marginal revenue products at the industry level, the industry expenditures on capital and labor are given by:

$$RK_s = R\frac{\alpha_s \theta_s PY}{\overline{MRPK_s}} \quad \text{and} \quad wL_s = w\frac{(1-\alpha_s)\theta_s PY}{\overline{MRPL_s}}.$$

The aggregate supplies of capital and labor are inelastically supplied at $K$ and $L$. Divide the industry expenditure on a factor by the sector expenditure on that factor. Then the allocations of capital and labor in an industry can be written as functions of $K$ and $L$:

$$K_s = K\frac{\theta_s \alpha_s / \overline{MRPK_s}}{\sum_{s'=1}^{S} \theta_{s'} \alpha_{s'} / \overline{MRPK_{s'}}} \quad \text{and} \quad L_s = L\frac{\theta_s (1-\alpha_s) / \overline{MRPL_s}}{\sum_{s'=1}^{S} \theta_{s'} (1-\alpha_{s'}) / \overline{MRPL_{s'}}}$$

Suppose that the marginal revenue products change at the firm level, but remain constant at the industry level. Then the allocation of capital and labor across industries stays the same. The exercise we perform in our empirical analysis involves reallocation within industries but not across industries.

## A.3 Firm output

From the demand for $Y_{si}$ in equation 3.4, the optimal expenditure on $Y_s$, the inverse of $P_s$, and the definition of $TFP_s$ in equation 3.13, firm output is given by:

$$\begin{aligned} P_{si}Y_{si} &= P_sY_s\left(\frac{P_{si}}{P_s}\right)^{(1-\sigma)} \\ &= \theta_sY\left(\frac{P_{si}}{P_s}\right)^{(1-\sigma)} \\ &= \theta_sY\left(P_{si}\left(\sum_{i=1}^{M_s}\left(\frac{1}{P_{si}}\right)^{\sigma-1}\right)^{\frac{1}{\sigma-1}}\right)^{1-\sigma} \\ &= \theta_sY\left(\frac{A_{si}}{TFPR_{si}}\right)^{\sigma-1}\left[\sum_{j=1}^{M_s}\left(\frac{A_{sj}}{TFPR_{sj}}\right)^{\sigma-1}\right]^{-1}. \end{aligned}$$

# Appendix B

# Survival rates by year of birth

Table B.1: Survival rates by year of birth

| Year of birth | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|
| 1985 | 0.78 | 0.86 | | 0.81 |
| 1986 | 0.60 | 0.79 | 0.72 | 0.86 |
| 1987 | 0.66 | 0.60 | 0.80 | 0.82 |
| 1988 | 0.55 | 0.68 | 0.82 | 0.80 |
| 1989 | 0.67 | 0.59 | | 0.70 |
| 1990 | 0.53 | 0.78 | 0.91 | 0.85 |
| 1991 | 0.42 | 0.48 | 0.73 | 0.83 |
| 1992 | 0.37 | 0.58 | 0.62 | 0.83 |
| 1993 | 0.28 | 0.47 | 0.70 | 0.80 |
| 1994 | | 0.71 | 0.55 | 0.97 |
| 1995 | | 0.39 | 0.96 | 0.76 |
| 1996 | | 0.37 | 0.52 | 0.83 |
| 1997 | | 0.23 | 0.65 | 0.78 |
| 1998 | | 0.28 | 0.60 | 0.76 |
| 1999 | | | 0.88 | 0.71 |
| 2000 | | | 0.81 | 0.82 |
| 2001 | | | 0.52 | 0.69 |
| 2002 | | | 0.45 | 0.73 |
| 2003 | | | 0.20 | 0.65 |
| 2004 | | | | 0.76 |
| 2005 | | | | 0.61 |
| 2006 | | | | 0.59 |
| 2007 | | | | 0.59 |
| 2008 | | | | 0.37 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The table shows the proportion of establishments born in a given year remaining from the previous Mexican Economic Census. For example, the 0.78 survival rate in 1998 for units born in 1985 means that 78% of the establishments born in 1985 present in the 1993 Census were also present in the 1998 Census. The missing values represent inconsistencies in the data, i.e., survival rates above 1.

# Appendix C

# Actual versus efficient establishment size

Table C.1: Actual and efficient value added and employment by size category for formal establishments

| Number of workers | 1993 0%-50% | 1993 50%-100% | 1993 100%-200% | 1993 200%+ | 2003 0%-50% | 2003 50%-100% | 2003 100%-200% | 2003 200%+ | 2013 0%-50% | 2013 50%-100% | 2013 100%-200% | 2013 200%+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 12% | 14% | 23% | 51% | 8% | 17% | 27% | 48% | 12% | 20% | 30% | 37% |
| | 0.1 | 0.4 | 1 | 7.7 | 0.1 | 0.3 | 0.8 | 7.1 | 0.1 | 0.2 | 0.6 | 8 |
| 2 | 12% | 16% | 26% | 46% | 7% | 16% | 29% | 48% | 12% | 20% | 28% | 40% |
| | 0.1 | 0.6 | 2 | 12 | 0.1 | 0.5 | 2 | 13 | 0.1 | 0.4 | 1 | 10 |
| 3 | 12% | 19% | 28% | 42% | 8% | 19% | 31% | 42% | 11% | 21% | 31% | 38% |
| | 0.2 | 0.9 | 3 | 16 | 0.2 | 0.8 | 2 | 16 | 0.1 | 0.6 | 2 | 14 |
| 4 | 11% | 20% | 29% | 40% | 8% | 21% | 31% | 40% | 10% | 20% | 30% | 40% |
| | 0.3 | 1 | 3 | 21 | 0.3 | 1.0 | 3 | 22 | 0.2 | 0.8 | 2 | 18 |
| 5 to 10 | 11% | 20% | 32% | 37% | 10% | 22% | 31% | 37% | 15% | 19% | 30% | 36% |
| | 0.4 | 2 | 5 | 30 | 0.4 | 1 | 5 | 32 | 0.2 | 1 | 4 | 29 |
| 11 to 19 | 10% | 20% | 34% | 35% | 10% | 23% | 33% | 34% | 10% | 18% | 31% | 40% |
| | 0.7 | 4 | 10 | 52 | 0.8 | 4 | 10 | 58 | 0.6 | 3 | 9 | 54 |
| 20 to 49 | 11% | 20% | 38% | 31% | 10% | 24% | 37% | 29% | 11% | 20% | 33% | 37% |
| | 1 | 8 | 23 | 99 | 2 | 8 | 25 | 113 | 1 | 7 | 21 | 107 |
| 50 to 99 | 14% | 22% | 39% | 25% | 10% | 26% | 42% | 22% | 12% | 23% | 33% | 32% |
| | 3 | 18 | 51 | 189 | 4 | 21 | 56 | 201 | 3 | 16 | 47 | 172 |
| 100 to 249 | 14% | 26% | 38% | 21% | 10% | 29% | 41% | 21% | 11% | 24% | 33% | 32% |
| | 8 | 43 | 114 | 336 | 10 | 46 | 116 | 380 | 7 | 38 | 103 | 293 |
| 250 to 499 | 15% | 26% | 39% | 21% | 9% | 29% | 44% | 18% | 10% | 24% | 36% | 30% |
| | 19 | 108 | 245 | 673 | 24 | 103 | 254 | 786 | 15 | 81 | 230 | 512 |
| 500 to 999 | 14% | 31% | 38% | 17% | 9% | 29% | 46% | 17% | 9% | 28% | 40% | 23% |
| | 45 | 201 | 514 | 1555 | 46 | 248 | 528 | 1230 | 36 | 185 | 418 | 1423 |
| 1000+ | 15% | 27% | 44% | 14% | 5% | 24% | 55% | 16% | 7% | 25% | 45% | 23% |
| Actual | 2519 | 1636 | 1652 | 1395 | 1673 | 1787 | 1775 | 1726 | 2231 | 2267 | 1851 | 1758 |
| Efficient | 152 | 542 | 1135 | 1984 | 112 | 592 | 1378 | 2981 | 85 | 627 | 1109 | 4124 |

*Source*: Author's calculations based on data from INEGI. *Note*: The columns show the percentage of establishments within each size category that would shrink or expand in the efficient benchmark. It also shows the average number of workers that establishments would employ in the efficient environment, assuming that the average wage per establishment is the same before and after reallocation.

Table C.2: Actual and efficient value added and employment by size category for semi-formal establishments

| Number of workers | 1993 0%-50% | 1993 50%-100% | 1993 100%-200% | 1993 200%+ | 2003 0%-50% | 2003 50%-100% | 2003 100%-200% | 2003 200%+ | 2013 0%-50% | 2013 50%-100% | 2013 100%-200% | 2013 200%+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 15% | 16% | 25% | 44% | 7% | 15% | 26% | 52% | 13% | 19% | 25% | 43% |
| | 0.1 | 0.3 | 0.9 | 7.1 | 0.1 | 0.3 | 0.8 | 6.4 | 0.0 | 0.2 | 0.5 | 6 |
| 2 | 16% | 22% | 28% | 34% | 14% | 24% | 30% | 33% | 18% | 22% | 28% | 32% |
| | 0.1 | 0.5 | 2 | 10 | 0.1 | 0.4 | 1 | 9 | 0.1 | 0.4 | 1 | 9 |
| 3 | 15% | 23% | 29% | 32% | 15% | 25% | 30% | 31% | 18% | 22% | 28% | 32% |
| | 0.2 | 0.8 | 2 | 14 | 0.2 | 0.7 | 2 | 13 | 0.1 | 0.5 | 2 | 11 |
| 4 | 15% | 24% | 29% | 32% | 14% | 25% | 30% | 30% | 17% | 23% | 29% | 32% |
| | 0.2 | 1.0 | 3 | 18 | 0.2 | 0.9 | 3 | 16 | 0.2 | 0.8 | 2 | 15 |
| 5 to 10 | 14% | 23% | 32% | 31% | 13% | 25% | 31% | 31% | 14% | 22% | 29% | 34% |
| | 0.4 | 1.7 | 5.0 | 28.0 | 0.4 | 2 | 4 | 27 | 0 | 1 | 4 | 24 |
| 11 to 19 | 11% | 22% | 34% | 33% | 11% | 25% | 34% | 30% | 12% | 21% | 31% | 35% |
| | 0.7 | 4 | 10 | 52 | 0.8 | 3 | 10 | 54 | 0.6 | 3 | 9 | 49 |
| 20 to 49 | 10% | 20% | 38% | 32% | 11% | 25% | 38% | 27% | 12% | 22% | 32% | 34% |
| | 1 | 8 | 23 | 95 | 2 | 8 | 23 | 109 | 1 | 7 | 20 | 105 |
| 50 to 99 | 12% | 22% | 38% | 28% | 10% | 27% | 40% | 22% | 16% | 25% | 33% | 26% |
| | 3 | 18 | 49 | 175 | 4 | 20 | 54 | 215 | 3 | 16 | 45 | 198 |
| 100 to 249 | 16% | 24% | 39% | 22% | 11% | 31% | 39% | 19% | 15% | 26% | 32% | 26% |
| | 7 | 39 | 98 | 342 | 7 | 48 | 113 | 393 | 6 | 35 | 96 | 338 |
| 250 to 499 | 16% | 24% | 36% | 23% | 14% | 25% | 42% | 19% | 19% | 26% | 33% | 22% |
| | 19 | 88 | 265 | 746 | 20 | 105 | 264 | 835 | 13 | 79 | 214 | 628 |
| 500 to 999 | 16% | 21% | 35% | 28% | 13% | 28% | 37% | 23% | 14% | 31% | 33% | 23% |
| | 48 | 167 | 513 | 965 | 41 | 179 | 649 | 1503 | 31 | 159 | 372 | 1171 |
| 1000+ | 35% | 19% | 31% | 15% | 13% | 24% | 43% | 20% | 14% | 27% | 34% | 26% |
| Actual | 40 | 406 | 827 | 4593 | 171 | 491 | 2067 | 2432 | 115 | 503 | 1288 | 4285 |
| Efficient | 1594 | 1731 | 1367 | 1409 | 2839 | 1790 | 2386 | 1681 | 2265 | 2002 | 2316 | 2190 |

*Source*: Author's calculations based on data from INEGI. *Note*: The columns show the percentage of establishments within each size category that would shrink or expand in the efficient benchmark. It also shows the average number of workers that establishments would employ in the efficient environment, assuming that the average wage per establishment is the same before and after reallocation.

Table C.3: Actual and efficient value added and employment by size category for informal establishments

| Number of workers | 1993 0%-50% | 1993 50%-100% | 1993 100%-200% | 1993 200%+ | 2003 0%-50% | 2003 50%-100% | 2003 100%-200% | 2003 200%+ | 2013 0%-50% | 2013 50%-100% | 2013 100%-200% | 2013 200%+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 39% | 20% | 20% | 21% | 36% | 21% | 21% | 22% | 39% | 20% | 19% | 22% |
| | 0.0 | 0.2 | 0.7 | 4.8 | 0.0 | 0.2 | 0.6 | 4.7 | 0.0 | 0.1 | 0.5 | 5 |
| 2 | 37% | 22% | 21% | 20% | 36% | 22% | 21% | 21% | 37% | 20% | 20% | 23% |
| | 0.0 | 0.3 | 1 | 7 | 0.0 | 0.3 | 1 | 7 | 0.0 | 0.2 | 1 | 7 |
| 3 | 33% | 24% | 23% | 20% | 32% | 23% | 23% | 23% | 31% | 20% | 22% | 28% |
| | 0.1 | 0.5 | 2 | 11 | 0.1 | 0.5 | 2 | 10 | 0.1 | 0.4 | 1 | 11 |
| 4 | 32% | 24% | 24% | 20% | 30% | 22% | 24% | 24% | 27% | 20% | 23% | 30% |
| | 0.1 | 0.8 | 3 | 15 | 0.1 | 0.7 | 2 | 15 | 0.1 | 0.6 | 2 | 15 |
| 5 to 10 | 31% | 24% | 25% | 20% | 24% | 23% | 26% | 27% | 22% | 18% | 26% | 34% |
| | 0.2 | 1.3 | 4.0 | 24.0 | 0.2 | 1 | 4 | 27 | 0.2 | 1 | 4 | 24 |
| 11 to 19 | 27% | 24% | 27% | 21% | 17% | 22% | 30% | 31% | 16% | 19% | 26% | 39% |
| | 0.4 | 3 | 10 | 84 | 0.5 | 4 | 10 | 68 | 0.4 | 3 | 9 | 71 |
| 20 to 49 | 26% | 23% | 28% | 23% | 17% | 21% | 33% | 30% | 15% | 21% | 29% | 35% |
| | 1 | 7 | 22 | 130 | 1 | 9 | 28 | 160 | 1 | 8 | 22 | 153 |
| 50 to 99 | 23% | 19% | 37% | 20% | 25% | 24% | 28% | 24% | 26% | 24% | 23% | 27% |
| | 1 | 18 | 46 | 316 | 3 | 18 | 58 | 331 | 2 | 14 | 48 | 223 |
| 100 to 249 | 41% | 20% | 20% | 20% | 31% | 25% | 26% | 18% | 28% | 28% | 26% | 18% |
| | 4 | 36 | 102 | 332 | 7 | 41 | 118 | 513 | 5 | 30 | 106 | 450 |
| 250 to 499 | 27% | 18% | 45% | 9% | 34% | 27% | 25% | 15% | 38% | 22% | 24% | 15% |
| | 2 | 40 | 364 | 1586 | 17 | 115 | 240 | 1259 | 10 | 72 | 334 | 1645 |
| 500 to 999 | 25% | 25% | 25% | 25% | 34% | 37% | 20% | 9% | 40% | 27% | 21% | 12% |
| | 0 | 25 | 289 | 920 | 34 | 189 | 521 | 2173 | 19 | 143 | 506 | 3436 |
| 1000+ | 0% | 0% | 0% | 100% | 53% | 20% | 14% | 14% | 44% | 24% | 20% | 13% |
| Actual | 0 | 0 | 0 | 1741 | 69 | 335 | 1330 | 4397 | 40 | 432 | 1023 | 5702 |
| Efficient | 0 | 0 | 0 | 1038 | 2414 | 1845 | 1572 | 1777 | 1964 | 2013 | 1590 | 1661 |

*Source*: Author's calculations based on data from INEGI. *Note*: The columns show the percentage of establishments within each size category that would shrink or expand in the efficient benchmark. It also shows the average number of workers that establishments would employ in the efficient environment, assuming that the average wage per establishment is the same before and after reallocation.

# Appendix D

# Regressions on the scaled capital and output distortions

Table D.1: Regression of the scaled capital distortion on establishment size measured as the number of workers

| Number of workers | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| 2 | 0.20 | 0.29 | 0.36 | 0.34 | 0.26 |
| 3 | 0.09 | 0.23 | 0.34 | 0.39 | 0.16 |
| 4 | 0.01 | 0.16 | 0.28 | 0.36 | 0.09 |
| 5 to 9 | -0.02 | 0.12 | 0.11 | 0.18 | -0.15 |
| 10 to 19 | 0.16 | 0.23 | 0.03 | 0.10 | -0.26 |
| 20 to 49 | 0.39 | 0.38 | 0.07 | 0.08 | -0.24 |
| 50 to 99 | 0.53 | 0.54 | 0.17 | 0.12 | -0.25 |
| 100 to 249 | 0.44 | 0.51 | 0.21 | 0.06 | 0.01 |
| 250 to 499 | 0.46 | 0.58 | 0.24 | 0.12 | 0.29 |
| 500 to 999 | 0.57 | 0.92 | 0.54 | 0.25 | 0.26 |
| 1000 or more | 0.68 | 1.22 | 0.85 | 0.59 | 0.39 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.05 | 0.07 | 0.06 | 0.06 | 0.07 |
| Adjusted R-2 | 0.05 | 0.07 | 0.06 | 0.06 | 0.07 |

*Source*: Author's calculations based on data from INEGI.
*Note*: All coefficients are significant at the 0.01% level, except for the coefficient of the 1000 or more workers dummy in the 1993 regression. This coefficient is significant at the 5% level.

Table D.2: Regression of the scaled capital distortion on establishment age

| Age in years | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| 2 | -0.11 | -0.07 | -0.03 | 0.00 | -0.02 |
| 3 | -0.17 | -0.12 | -0.04 | -0.01 | -0.04 |
| 4 | -0.12 | -0.19 | -0.07 | -0.04 | 0.00 |
| 5 to 9 | -0.06 | -0.23 | -0.13 | -0.05 | -0.03 |
| 10 to 14 | 0.02 | -0.17 | -0.21 | -0.10 | -0.06 |
| 15 to 19 | 0.00 | -0.19 | -0.23 | -0.24 | -0.21 |
| 20 to 24 | 0.05 | -0.13 | -0.19 | -0.22 | -0.27 |
| 25 or more | 0.05 | -0.17 | -0.14 | -0.25 | -0.28 |
| Industry F.E. | YES | YES | YES | YES | YES |
| | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.05 | 0.06 | 0.06 | 0.05 | 0.07 |
| Adjusted R-2 | 0.05 | 0.06 | 0.06 | 0.05 | 0.07 |

*Source*: Author's calculations based on data from INEGI.
*Note*: All coefficients are significant at the 0.01% level, except for the coefficient of the 2-year-old dummy in the 2008 regression. This coefficient is statistically equal to zero.

Table D.3: Regression of the scaled capital distortion on establishment formality status

| Formality status | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| Formal | -0.29 | -0.22 | -0.31 | -0.30 | -0.46 |
| Semiformal | -0.34 | -0.16 | -0.22 | -0.20 | -0.36 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.06 | 0.06 | 0.06 | 0.05 | 0.07 |
| Adjusted R-2 | 0.06 | 0.06 | 0.06 | 0.05 | 0.07 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The excluded category is informal. Quintic terms for the number of workers and age are included as controls. All coefficients are significant at the 0.01% level.

Table D.4: Regression of the scaled output distortion on establishment size measured as the number of workers

| Number of workers | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| 2 | -0.02 | 0.01 | 0.16 | 0.17 | 0.05 |
| 3 | -0.22 | -0.19 | 0.00 | 0.13 | -0.19 |
| 4 | -0.36 | -0.32 | -0.13 | 0.04 | -0.34 |
| 5 to 9 | -0.51 | -0.49 | -0.37 | -0.22 | -0.61 |
| 10 to 19 | -0.58 | -0.56 | -0.56 | -0.49 | -0.83 |
| 20 to 49 | -0.50 | -0.50 | -0.55 | -0.63 | -0.83 |
| 50 to 99 | -0.35 | -0.36 | -0.43 | -0.55 | -0.64 |
| 100 to 249 | -0.29 | -0.32 | -0.36 | -0.49 | -0.53 |
| 250 to 499 | -0.23 | -0.28 | -0.35 | -0.41 | -0.33 |
| 500 to 999 | -0.13 | -0.14 | -0.25 | -0.33 | -0.32 |
| 1000 or more | 0.05 | -0.01 | -0.06 | -0.08 | -0.21 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.05 | 0.04 | 0.05 | 0.04 | 0.06 |
| Adjusted R-2 | 0.05 | 0.04 | 0.05 | 0.04 | 0.06 |

*Source*: Author's calculations based on data from INEGI.
*Note*: All coefficients are significant at the 0.01% level except for the 1000 or more workers dummy from 1993 to 2008 and the 3 workers dummy in 2003.

Table D.5: Regression of scaled output distortion on establishment age

| Age in years | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| 2 | -0.42 | -0.24 | -0.18 | -0.38 | -0.41 |
| 3 | -0.43 | -0.24 | -0.18 | -0.40 | -0.43 |
| 4 | -0.40 | -0.23 | -0.21 | -0.38 | -0.31 |
| 5 to 9 | -0.34 | -0.22 | -0.21 | -0.38 | -0.27 |
| 10 to 14 | -0.23 | -0.12 | -0.19 | -0.34 | -0.24 |
| 15 to 19 | -0.22 | -0.08 | -0.19 | -0.40 | -0.30 |
| 20 to 24 | -0.20 | -0.01 | -0.14 | -0.38 | -0.31 |
| 25 or more | -0.08 | 0.02 | -0.11 | -0.37 | -0.34 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.06 | 0.03 | 0.03 | 0.04 | 0.04 |
| Adjusted R-2 | 0.06 | 0.03 | 0.03 | 0.04 | 0.04 |

*Source*: Author's calculations based on data from INEGI.
*Note*: All coefficients are significant at the 0.01% level, except for the coefficient of the 20 to 24 years old dummy in the 2003 regression. This coefficient is significant at the 5% level.

Table D.6: Regression of scaled output distortion on establishment formality status

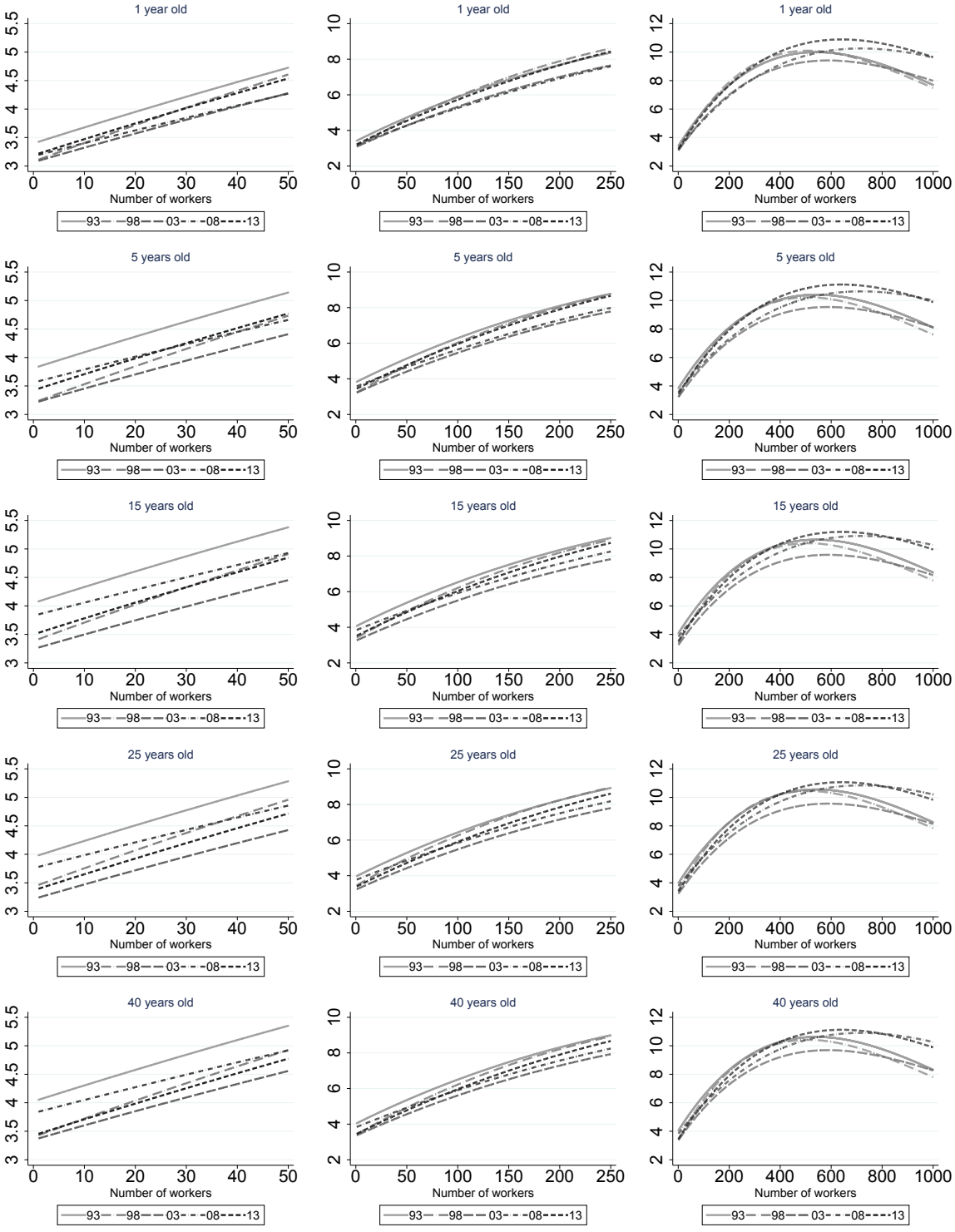
| Formality status | 1993 | 1998 | 2003 | 2008 | 2013 |
|---|---|---|---|---|---|
| Formal | -0.87 | -0.82 | -0.84 | -0.75 | -0.80 |
| Semiformal | -0.72 | -0.67 | -0.65 | -0.48 | -0.65 |
| Industry F.E. | YES | YES | YES | YES | YES |
| N | 2052505 | 2423725 | 2597924 | 2519113 | 3365946 |
| R-2 | 0.11 | 0.08 | 0.07 | 0.06 | 0.07 |
| Adjusted R-2 | 0.11 | 0.08 | 0.07 | 0.06 | 0.07 |

*Source*: Author's calculations based on data from INEGI.
*Note*: The excluded category is informal. Quintic terms for the number of workers and age are included as controls. All coefficients are significant at the 0.01% level, except for age to the third, fourth and fifth power in 2003.

# Appendix E

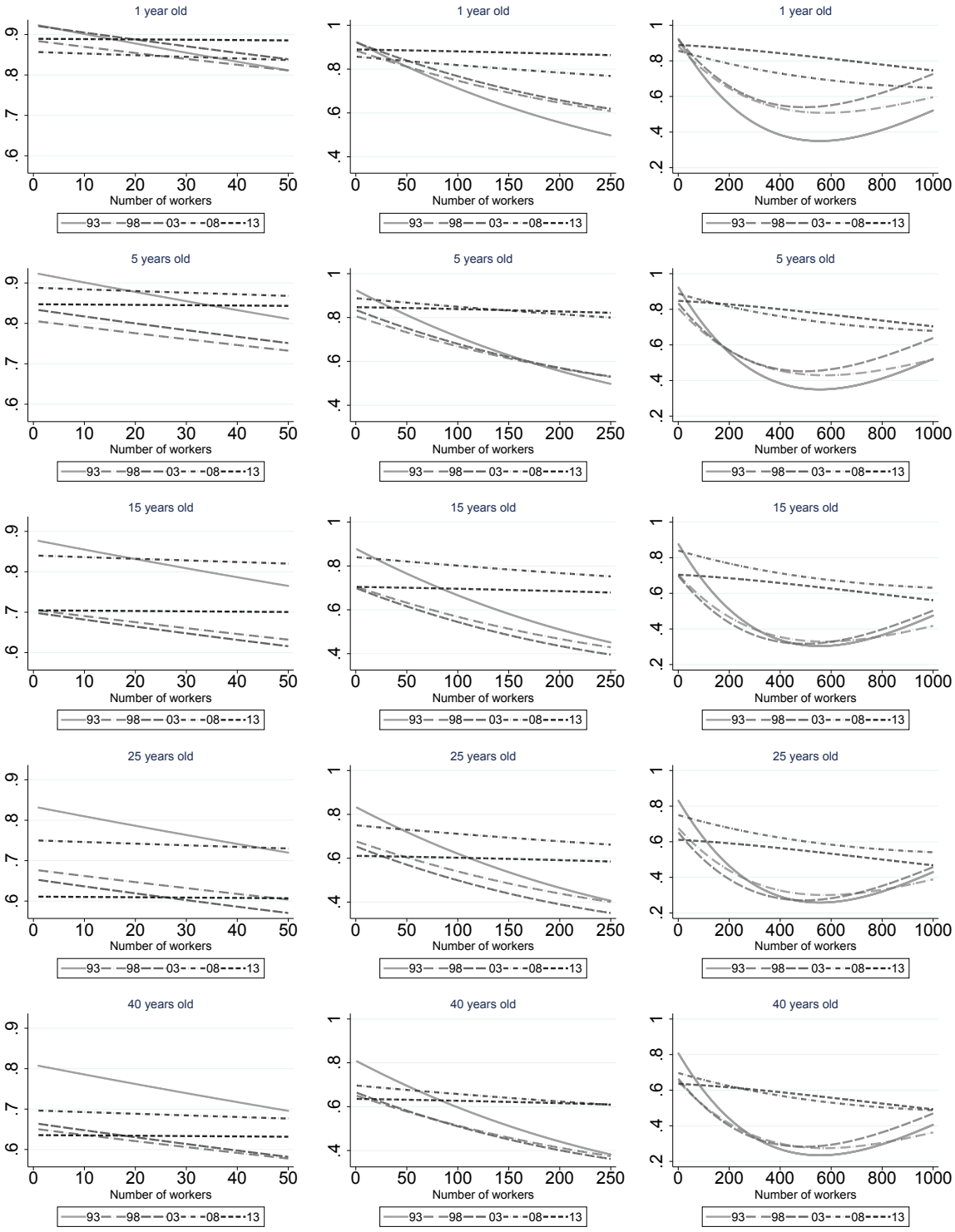
# Predicted values from regressions from section 4.3

Figure E.1: *Predicted TFPQ for formal establishments*



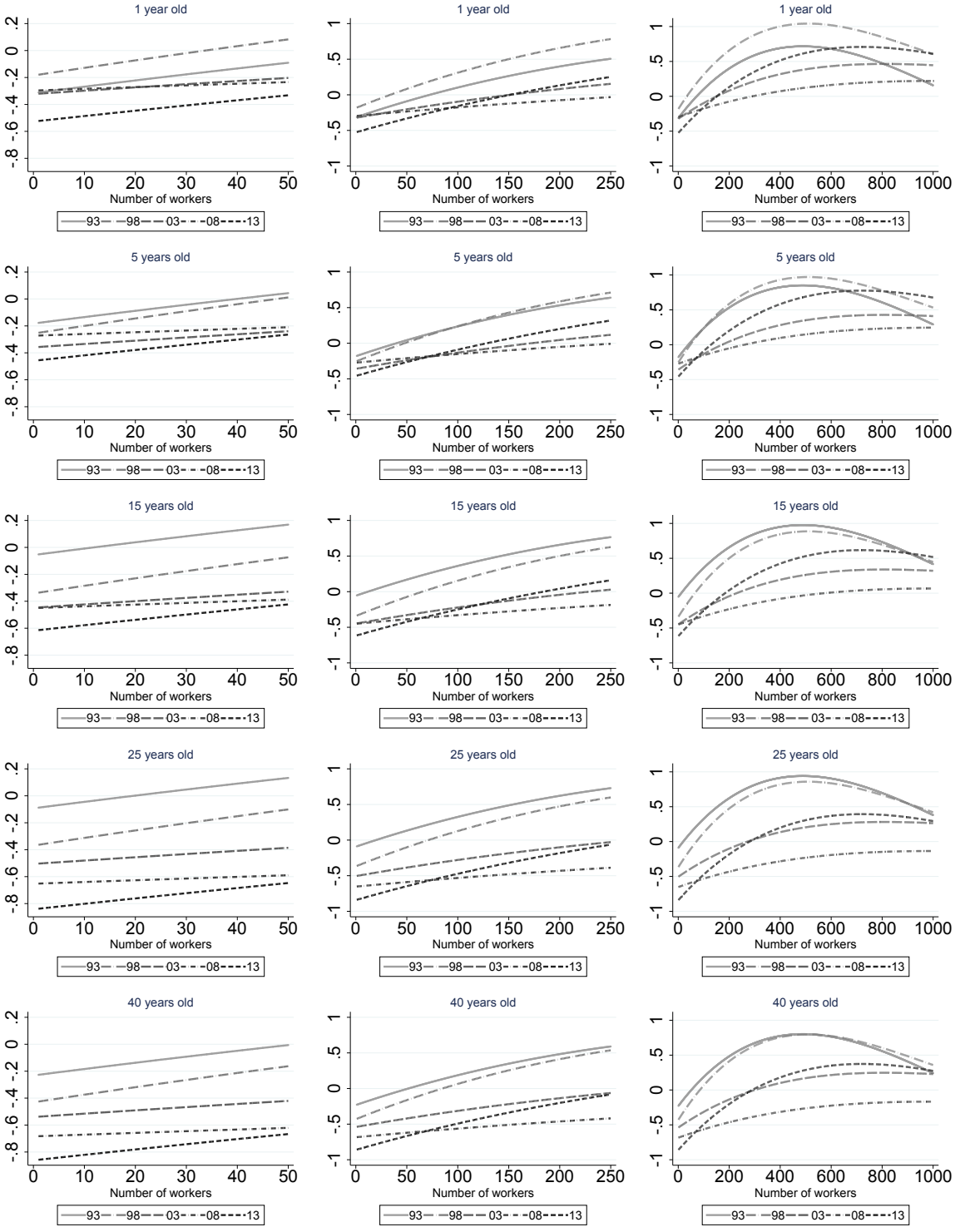

*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure E.2: *Predicted TFPR for formal establishments*



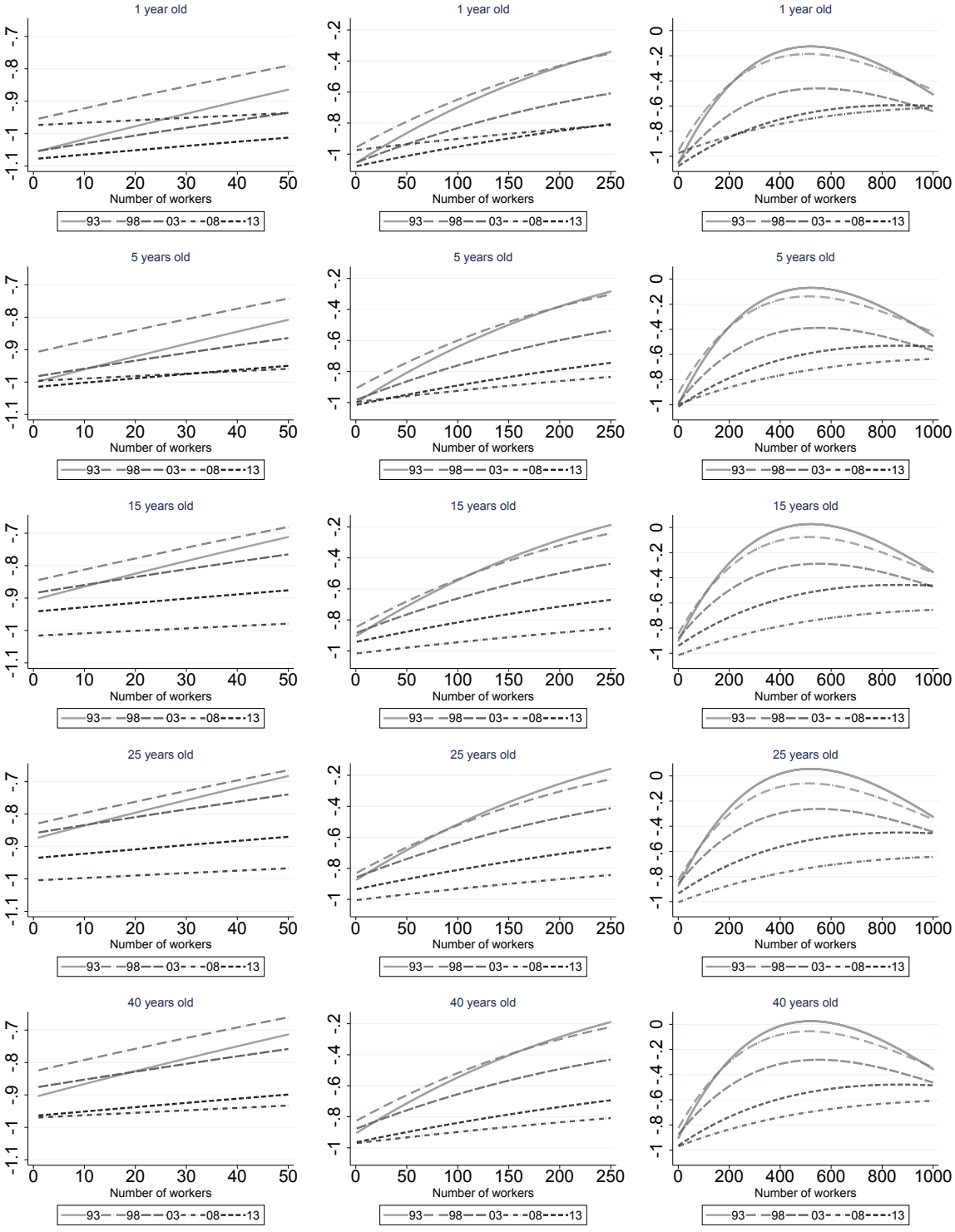

*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure E.3: *Predicted scaled capital distortion for formal establishments*




*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

Figure E.4: *Predicted scaled output distortion for formal establishments*




*Source*: Author's calculations based on data from INEGI.
*Note*: The numeraire in each year are 1-year-old informal establishments employing 1 worker.

# References


Azuara, O., and Marinescu, I. (2013). "Informality and the expansion of social protection programs: Evidence from mexico." *Journal of health economics*, *32*(5), 938–950.

Bellon, F., and Mallen-Pisano, J. (2013). *Is Misallocation Higher in France than in the United States?* (Tech. Rep. No. 38). GREDEG, Nice, France, Working Paper.

Bento, P., and Restuccia, D. (2016). *Misallocation, establishment size, and productivity* (Tech. Rep.). National Bureau of Economic Research.

Buera, F. J., Kaboski, J. P., and Shin, Y. (2011). "Finance and Development: A Tale of Two Sectors." *American Economic Review*, *101*(5), 1964–2002.

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

Busso, M., Madrigal, L., and Pagés, C. (2012). *Productivity and Resource Misallocation in Latin America* (Tech. Rep. No. 306).

Dias, D., Marques, C. R., and Richmond, C. (2015). *Misallocation and Productivity in the Lead Up to the Eurozone Crisis* (Tech. Rep. No. 1146). Board of Governors of the Federal Reserve System, Virginia, United States.

Dias, D., Marques, C. R., and Richmond, C. (2016). *Comparing Misallocation between Sectors in Portugal* (Tech. Rep. No. 14). Bank of Portugal Working Paper.

Ha, D. T. T., Kiyota, K., and Yamanouchi, K. (2016). "Misallocation and Productivity: The Case of Vietnamese Manufacturing." *Asian Development Review*, *33*(2), 94-118.

Hanson, G. (2010). "Why isn't Mexico Rich?" *Journal of Economic Literature*, *48*(4), 987-

1004.

Hsieh, C.-T., and Klenow, P. J. (2009a). "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, *123*(4), 1403-1448.

Hsieh, C.-T., and Klenow, P. J. (2009b). *Resource Misallocation in Mexico* (Tech. Rep.). Inter-American Development Bank, Washington, D.C., United States, Unpublished manuscript.

Hsieh, C.-T., and Klenow, P. J. (2014). "The Life Cycle of Plants in India and Mexico." *Quarterly Journal of Economics*, *129*(3), 1035-1084.

Inklaar, R., Lashitew, A. A., and Timmer, M. P. (2016). "The Role of Resource Misallocation in Cross-country Differences in Manufacturing Productivity." *Macroeconomic Dynamics*, *20*(1), 100-119.

Kim, M., Oh, J., and Shin, Y. (2016). "Misallocation and manufacturing tfp in korea, 1982–2007." *Federal Reserve Bank of St. Louis Review, Forthcoming*.

López-Calva, L. F., and Levy Algazi, S. (2016). "Labor earnings, misallocation, and the returns to education in mexico."

Machicado, C. G., and Birbuet, J. C. (2012). "Misallocation and Manufacturing TFP in Bolivia during the Market Liberalization Period." *The B.E. Journal of Macroeconomics*, *12*(1).

Oberfeld, E. (2013). "Productivity and misallocation during a crisis: Evidence from the Chilean crisis of 1982 ." *Review of Economic Dynamics*, *16*(1), 100-119.

Restuccia, D., and Rogerson, R. (2008). "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic Dynamics*, *11*(4), 707-720.