[redacted]

 WP/18/112

IMF Working Paper

Western Hemisphere Department

Resource Misallocation and Productivity: Evidence from Mexico

Prepared by Florian Misch and Christian Saborowski

Authorized for distribution by Costas Christou

May 2018

Abstract

This paper explores the role for specific structural distortions in explaining Mexico's weak productivity growth through the resource misallocation channel. It makes two contributions. First, we validate the approach of measuring misallocation indirectly as in Hsieh and Klenow (2009) by illustrating a close correlation between misallocation and per capita incomes across Mexican states. Second, we exploit the large variation in resource misallocation within industries and across states together with unusually rich data at the firm, local, and industry level to shed light on its determinants. We identify several well-defined distortions that have a statistically and economically meaningful effect on productivity via resource misallocation.

JEL Classification Numbers: D24, O12, O47

Keywords: Resource Misallocation; Productivity; Distortions

Author's E-Mail Address: FMisch@imf.org, CSaborowski@imf.org

* The paper has benefitted from valuable guidance by Costas Christou and Era Dabla-Norris as well as support under the Fiscal Affairs Department's microdata initiative. An earlier and shorter version of the paper is included in the Selected Issues Paper accompanying the IMF's 2017 Article IV report for Mexico. We would like to thank Suman Basu, Francesca Caselli, Laura Jaramillo, Frederic Lambert, Frederico Lima, Robert Rennhack, Alejandro Werner, participants in seminars at the IMF, and Banco de México staff for useful comments and support. We are also grateful to Natalia Volkow and her team from the National Institute of Statistics and Geography (INEGI) for excellent and generous support in accessing and processing data. Carolina Correa and Changchun Wang provided excellent research assistance. The views expressed herein are those of the authors and should not be attributed to the International Monetary Fund, its Executive Board, or its Management. This paper uses confidential data from Mexico's Economic census, accessed through the INEGI's Microdata Laboratory. All results and information shown were compiled by the authors and are not part of INEGI's official statistics.

Contents Page

I. INTRODUCTION

Mexico implemented sweeping structural reforms during the mid-1990s following a series of economic and financial crises. The reforms succeeded in achieving macroeconomic stability, opened the economy up to trade and foreign investment, and boosted educational attainment. Against this backdrop, Mexico's low average per capita growth rate over the last two decades and in particular its negative productivity growth remain a puzzle (Levy and Rodrik, 2017).

The objective of this paper is to provide evidence supporting the view that resource misallocation rather than access to technology could lie at the heart of Mexico's low productivity (Busso et al, 2012; IMF, 2017; Keller, 2002).[1] In a second step, we exploit subnational- and industry variation to uncover specific distortions that explain the inefficient allocation of resources in Mexico. Finally, we ask how much productivity levels could have benefited from addressing these distortions. Our analysis follows Hsieh and Klenow (2009) in calculating resource misallocation indirectly, using exceptionally rich firm-level data from the Mexican Economic Census which comprises the universe of urban formal and informal firms with fixed establishments.

A first look at the data suggests that the potential productivity gains from fully eliminating all distortions that give rise to resource misallocation in Mexico - at 125 percent - are indeed large compared to other countries. These estimates are, however, more conservative than previous estimates for Mexico by Busso et al (2012). At the same time, we find that the aggregate results mask significant variation not only across industries but also at the state level. For example, the productivity gains from eliminating resource misallocation in Mexico's least efficient state are some two-and-a-half times larger than the potential gains in Mexico's most efficient state. These subnational differences are much larger than those reported in Calligaris et al. (2016) for Italy – the only other paper that examines resource misallocation at the subnational level in a comparable way. Interestingly, we also find that the subnational differences in misallocation correlate very closely with state-level incomes per capita, even when we control for the composition of industries across states.[2] These results provide empirical support for the economic relevance of measuring resource misallocation indirectly through a model-based approach as proposed by Hsieh and Klenow (2009).

---

[1] Busso et al. (2012) examine resource misallocation in Latin America and find that the associated TFP losses in Mexico are substantially higher than in the rest of Latin American countries they consider. IMF (2017) suggests that resource misallocation in Mexico is above the 50th/75th percentile in a sample of 57 developing and emerging market economies depending on the year considered. Relatedly, based on the arguments by Keller (2001), it appears doubtful that the lack of access to technology alone could explain low productivity growth in Mexico, given that Mexico successfully has opened its economy to international trade and investment since the mid-1990s.

[2] In terms of industry variation, we find that misallocation is somewhat more severe in the manufacturing than in the services sector, in line with evidence in previous studies (Diaz et al., 2016; Busso et al., 2012).

We run regressions at the industry-state level to explain the large variation in resource misallocation across industries and states, controlling for unobserved industry and state fixed effects. Our candidate regressors are chosen to represent distortions that, according to theory, matter for the allocation of resources across firms by benefitting some firms at the expense of others, independently of their relative productivity levels (see also Hanson, 2010). The regressors are calculated using firm-, municipal- and state-level data from the Mexican Economic Census as well as other data sources covering information on crime, demographics, and economic geography. We find compelling evidence suggesting that misallocation rises with the prevalence of labor informality, crime, corruption and market concentration as well as weaker access to financial and telecommunications services. Finally, we show that misallocation also increases when firms are geographically further away from major population centers. To illustrate the economic significance of our results, the median Mexican state would see TFP rise by some 13 percent in a hypothetical reform scenario where all distortions included in our baseline regression would be attenuated to levels close to the domestic frontier.

The role of resource misallocation in explaining productivity levels has recently received much attention following the seminal work by Hsieh and Klenow (2009).[3] Restuccia and Rogerson (2017) distinguish two broad approaches to quantify resource misallocation. The *direct approach*, quantifies the effects of specific and observable distortions by constructing a counterfactual scenario, either from a structural model, or from a quasi-natural experiment. The *indirect approach* employed in our analysis, in turn, infers resource misallocation from the dispersion of the marginal products of capital and labor which are calculated using a calibrated model with firm-level data.[4] While the *direct approach* has failed so far in finding evidence of distortions that can explain important shares of plausible levels of aggregate resource misallocation, the *indirect approach* has been criticized because its estimates of resource misallocation could reflect misspecification of production functions within industries or adjustment costs, and because estimates from different countries may not be comparable due to measurement error (Restuccia and Richard Rogerson, 2017). More recently, Haltiwanger et al. (2018) have argued that the framework by Hsieh and Klenow (2009) rests on strong assumptions which are often difficult to verify.

---

[3] See Restuccia and Rogerson (2017), Restuccia and Rogerson (2013) and Hopenhayn (2014) for surveys of the literature.

[4] Several papers have used the *indirect approach* to show that the TFP gains from eliminating the distortions that give rise to resource misallocation could be economically significant. For instance, Hsieh and Klenow (2009) show that the TFP gains in China and India could amount to around 80–130 percent. Our aggregate estimates of resource misallocation in Mexico are broadly comparable with these studies.

Our results contribute to the literature in two ways. First, we confirm that the *indirect approach* to measuring resource misallocation delivers strong and economically sensible predictions at the macro level despite the often-valid criticism of some of its underlying assumptions. In particular, we show that differences in per capita incomes are indeed closely correlated with differences in resource misallocation at the state level. We are thus able to validate the economic relevance of the model-based approach to measuring resource misallocation by confirming the basic conjecture in the paper by Hsieh and Klenow (2009).[5] The result is also consistent with Restuccia and Rogerson (2008), but stands in contrast to previous findings by Inklaar et al. (2017) at the cross-country level.[6] Our second contribution is to relate various theoretically motivated distortions to resource misallocation by exploiting variation across state-industry pairs. We show that the findings are indeed highly economically significant in that addressing the distortions included in our baseline could yield substantial increases in state-level productivity.

Mexico has implemented an ambitious structural reform agenda in recent years in a coordinated effort to lift productivity growth. The reforms have already achieved major transformations in network industries and have contributed to economic growth despite external headwinds (Saborowski, 2017).[7] The findings in our paper highlight the need to push ahead with the implementation of these reforms, underscoring the importance of boosting competition and access to financial and telecommunications services and strengthening the rule of law to root out corruption, crime and labor informality. The link between the geographic isolation of some regions and resource misallocation highlights the importance of policies that increase the mobility of production factors in some of Mexico's less developed regions. Such policies could include targeted physical or transportation infrastructure investments.

The remainder of the paper is organized as follows. In Section II, we revisit the conceptual framework. In Section III, we describe the underlying data. In Section IV, we report stylized facts on resource misallocation. In Section V, we provide the results from the econometric analysis, and Section VI concludes.

---

[5] The correlation remains high even after controlling for unobserved industry-level fixed effects which should, among other things, account for potential differences in industry composition (Dias et al., 2016).

[6] The authors do not find evidence in favor of a correlation between resource misallocation and a country's level of development. The difference between our result and theirs is likely driven by the fact that we take a subnational rather than a cross-country approach in which our firm- and state-level data allow us to consider a much broader set of sectors, measure resource misallocation within much narrower sectors and omit from state-level GDP figures sectors that we do not consider in our resource misallocation measures.

[7] The data used in our analysis precedes many of the structural reforms (e.g. the telecommunications reform) implemented in recent years. The latter may already help to partially address the resource misallocation we observe.

II. CONCEPTUAL FRAMEWORK

Aggregate TFP depends not only on the level of productivity of individual firms, but also on the allocation of labor and capital across firms within narrowly defined industries. Resource misallocation denotes a situation in which capital and labor are poorly distributed so that less productive firms receive a larger share of capital and labor than they should according to their level of productivity. Such misallocation arises in the presence of distortions. While these distortions are not necessarily observable, at least not in a direct way, the framework by Hsieh and Klenow (2009) can be used to quantify distortions indirectly by measuring the potential TFP gains that would arise in the absence of them.

We apply the Hsieh and Klenow framework to the state level and assume that each industry *j* in state *s* consists of N_{js} monopolistically competitive firms and that each state consists of J_s industries. In each state, there is a single final good derived from combining the output Y_{js} from each of the states' J_s industries using Cobb-Douglas production technology:

$$Y_{js} = \prod_{j=1}^{J_s} \left[Y_{js}\right]^{\theta_{js}} \tag{1}$$

with $\sum_{j=1}^{J_s} \theta_{js} = 1$. Total output in each industry *j* is given by a constant elasticity of substitution production function

$$Y_{js} = \left[\sum_{j=1}^{N_{js}} \left(y_{ijs}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{2}$$

where y_{ijs} denotes the output of firm i in industry *j* in state *s*. σ denotes the elasticity of substitution between output varieties in each industry. Each firm *i*'s output is produced by a Cobb-Douglas production function:

$$y_{ijs} = A_{ijs} {k_{ijs}}^{\alpha} {l_{ijs}}^{1-\alpha} \tag{3}$$

where *k*, *l*, and *A* denote capital, labor and physical productivity, respectively, and α represents the output elasticity of capital. Firms choose prices, capital and labor to maximize profits

$$\max \pi_{ijs} = \left(1-\tau_{ijs}^{y}\right) p_{ijs} y_{ijs} - \left(1+\tau_{ijs}^{k}\right)(r+\delta) k_{ijs} - w l_{ijs} \tag{4}$$

where firm *i*'s price is p_{ijs}, and *w*, δ and *r* denote the wage, depreciation and interest rates, respectively. τ_{fis}^{y} represents a firm-specific wedge that distorts output decisions, and τ_{fis}^{k} represents a firm-specific wedge that distorts the capital-to-labor ratio. Restuccia and Rogerson (2017) distinguish three categories of such distortion including statutory provisions, including elements of the tax code and regulations (e.g., size-dependent taxation), discretionary provisions made by the government or other private institutions (such as banks) that favor or penalize specific firms (e.g., selective enforcement of taxation or outright government corruption), and market imperfections (e.g., monopoly power, and enforcement of property rights).

The first-order conditions with respect to capital and labor of each firm are then given by

$$MRPL_{ijs} = \left(\frac{1-\alpha}{\mu}\right)\left(\frac{p_{ijs}y_{ijs}}{l_{ijs}}\right) = \left(\frac{1}{1-\tau_{ijs}^{y}}\right)w \tag{5}$$

$$MRPK_{ijs} = \left(\frac{\alpha}{\mu}\right)\left(\frac{p_{ijs}y_{ijs}}{k_{ijs}}\right) = \left(\frac{1+\tau_{ijs}^{k}}{1-\tau_{ijs}^{y}}\right)(r+\delta) \tag{6}$$

where $\mu = \frac{\sigma}{\sigma-1}$ is the constant markup of price over marginal cost and MRPL and MRPK represent the marginal products of labor and capital. The revenue productivity (TFPR) of each firm, in turn, is defined as the product of firm i's price p_{ijs} and physical productivity A_{ijs}:

$$TFPR_{ijs} = p_{ijs}A_{ijs} = \left(\frac{p_{ijs}y_{ijs}}{k_{ijs}{}^{\alpha}l_{ijs}{}^{1-\alpha}}\right) = \mu\left(MRPK_{ijs}/\alpha\right)^{\alpha}\left(MRPL_{ijs}/(1-\alpha)\right)^{1-\alpha} \tag{7}$$

Equation (7) implies that firms with larger distortions exhibit larger marginal revenue products and a higher TFPR. If all firms either face no distortions at all, or the distortions are the same across firms, more productive firms would be allocated more resources than less productive ones, and the marginal products of capital and labor will equalize. The presence of distortions leads to the dispersion of marginal revenue products and revenue productivity, thereby resulting in resource misallocation. By contrast, physical productivity is obtained from

$$A_{is} = \frac{y_{ijs}}{k_{ijs}{}^{\alpha}l_{ijs}{}^{1-\alpha}} = \frac{(p_{ijs}y_{ijs})^{\frac{\sigma}{\sigma-1}}}{k_{ijs}{}^{\alpha}l_{ijs}{}^{1-\alpha}} \tag{8}$$

where we derive quantities from observed revenues using an isoelastic demand function for each firm's output. Industry-level TFP in state *s* is defined as

$$TFP_{js} = \left[\sum_{i=1}^{N}\left(A_{ijs}\frac{\overline{TFPR}_{js}}{TFPR_{ijs}}\right)^{\sigma-1}\right]^{\frac{1}{\sigma-1}} \tag{9}$$

where $\overline{TFPR}_{js}$ is the geometric average of the average marginal revenue productivity of capital and labor in each industry. *TFP* in a given industry is maximized when marginal products are equalized across plants.

The level of resource allocation efficiency (which is the ratio of actual output to the level of output in the absence of distortions) and the TFP gain associated with eliminating resource misallocation in each state can be written as

$$\left(\frac{Y_s}{Y_s^*}\right) = \prod_{j=1}^{J_s} \left[\sum_{i=1}^{N_{js}} \left(\frac{A_{ijs}}{\overline{A_{Js}}} \frac{\overline{TFPR_{Js}}}{TFPR_{ijs}}\right)^{\sigma-1}\right]^{\theta j_s / (\sigma-1)} \tag{10}$$

and

$$TFPgain_s = 100 \times (TFP_s/TFP_s^* - 1) \tag{11}$$

where TFP^* is the hypothetical level of TFP when resources are efficiently allocated. We also calculate aggregate resource misallocation for Mexico's entire economy analogously by effectively treating the entire Mexico as one state. In the subsequent sections, we will calculate the TFP gains based on Equation (11) for each state to illustrate the variation of resource misallocation within Mexico and for each industry-state pair which we will use as the left-hand side variable in our econometric analysis.

III. Data

The paper uses establishment-level data from the latest wave of the Mexican Economic Census. The Mexican Instituto Nacional de Estadistica y Geografia (INEGI) compiles the data set every five years, with the survey responses in the latest wave referring to the year 2013. The data is unique in that it contains around 3.5 million observations covering the universe of non-agricultural formal and informal Mexican firms with fixed establishments in urban areas regardless of their industry and size. It includes a vast amount of information on firm characteristics and operations, allowing us to compute not only a measure of resource misallocation at the industry-state level, but also a broad range of proxies of potential distortions to serve as explanatory variables for our regression analysis. Previous rounds of the census have been used to compute resource misallocation in other studies such as Busso et al. (2012) who focus their analysis on productivity differences between formal and informal firms.

We compute resource misallocation at the 4-digit level based on the NAICS 2002 industry classification for the manufacturing and service sector economy. We exclude sectors in which productivity estimates could conceivably be misleading or difficult to compare to the remaining sectors, including financial services, construction, utilities, real estate, professional / technical services sectors as well as the management of shell companies. In addition, we omit health and education as well as arts and culture, and thus sectors in which an important share of the entities involved are unlikely to pursue profit objectives. This leaves manufacturing, retail and wholesale trade, transportation and warehousing, accommodation and food services, information and other services in our sample. As is standard in the literature, we also exclude all entities with negative or zero reported value added, capital, sales or labor input (including labor provided by the owner of the firm), and omit industries with fewer than 10 firms. In addition, we remove the 1 percent tails of the distribution of firm-specific output wedges, capital wedges, and total factor productivity. We end up with close to 3 million establishments and 3,139 industry-state pairs.

The output elasticities of labor and capital for each industry are approximated by their respective cost shares in the United States from the Bureau of Economic Analysis, in line with the literature. The idea here is to use cost shares that are independent of distortions in the Mexican economy itself. Moreover, we set the rental price of capital to 0.1, assuming real interest and depreciation rates of 5 percent, whereas we assume a uniform wage rate across firms; the elasticity of substitution between the outputs of different firms is set to 3. Capital and sales come straight from the data. In the baseline specification, we use firm-level employment as the labor variable in the production function. We choose employment over the wage bill given that many firms in Mexico use unpaid labor (e.g., family members), implying that the wage bill may be incomplete, missing or zero even if firms have one or more employees. In a robustness check, we use the firm-level wage bill as an alternative to somewhat relax the assumption of a uniform wage rate across firms.

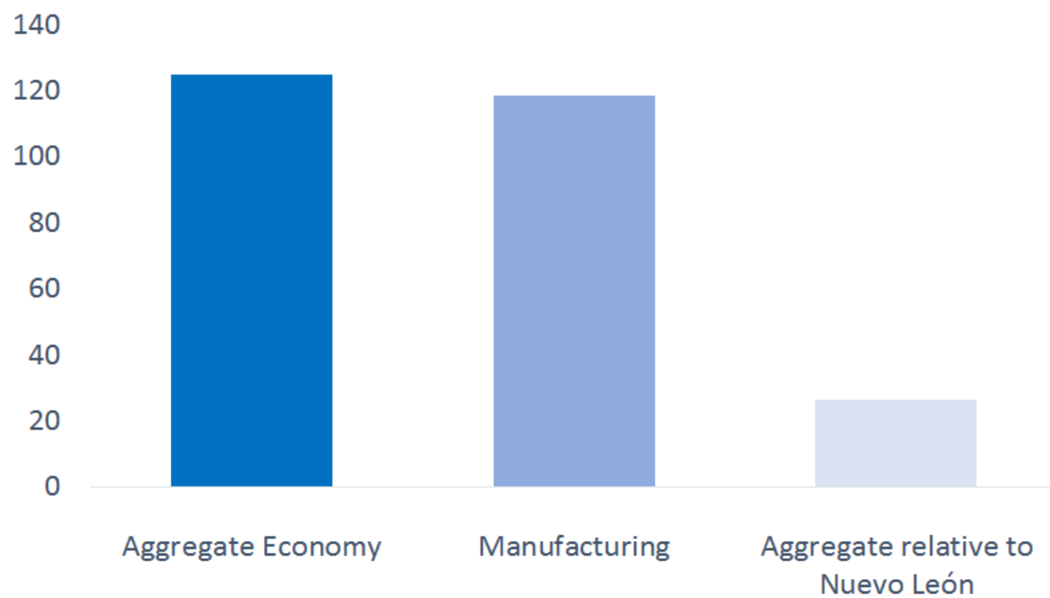
In compiling various proxies for candidate distortions and other control variables for the regression analysis, we use information both from the Economic Census itself and from other data sources, including the 2010 population census as well as the 2010 SIMBAD database. We describe these in Annex 2 and present summary statistics in Annex 3.

IV. Stylized Facts

A. Aggregate TFP Gains

We begin the empirical analysis by computing resource misallocation in Mexico at the country level, namely the TFP gains that would arise from eliminating any distortions that prevent the efficient allocation of labor and capital across firms within narrowly defined industries. As shown in Figure 1, fully eliminating resource misallocation would increase aggregate TFP by almost 125 percent. We test the robustness of this estimate in two ways: first, we use the firm-level wage bill instead of employment as the labor variable in firms' production functions. The advantage of this approach is that it controls for wage levels in addition to employment levels. The specification change reduces our sample size, however, as many small family-operated firms in Mexico do not have paid employees, and because we refrain from imputing wages for firms with missing observations given that their nature is different (contrary to Busso et al., 2013). Nevertheless, our estimate of aggregate resource misallocation in Mexico remains broadly in the same ballpark, at 116 percent. In a second robustness check, we expand the tails of firm-specific output wedges (and capital wedges as well as total factor productivity) that we remove from the sample to 2.5 percent, from 1 percent previously. Once again, our estimate of resource misallocation remains at around 125 percent.

Figure 1. TFP Gains from Efficient Allocation of Resources



Notes: The figure shows the TFP gains in percent based on Equation 11 if resources were allocated optimally across firms within all sectors under consideration.
Sources: 2013 Mexican Economic Census and own compilation.

While gains of more than 100 percent appear large at first sight, they are more conservative than those found by other studies. Estimates in Busso et al. (2012) – at around 200 percent - are larger than ours in part because the authors use data from earlier years, but also because our approach omits a greater number of industries, uses a broader industry classification and different data cleaning procedure as well as avoids the imputation of wages by using firm-level employment as the labor variable. Mayorga Garrido-Cortes (2017) uses data from the 2013 Economic census and finds that aggregate TFP gains amount to around 180 percent, but also imputes wages and uses a broader industry classification, among other differences. The notable differences between these results and ours highlight that the precise degree of misallocation is sensitive the underlying approach taken. Nevertheless, all three studies find that eliminating misallocation would more than double TFP and thus agree on the fact that resource misallocation (together with the underlying distortions) is likely to be a major factor in explaining weak productivity in Mexico.[8]

Several authors have pointed out that the modeling framework in Hsieh and Klenow (2009) is too simple and can lead to measurement errors that are likely to bias misallocation estimates upward. For example, the assumption that all firms in a given industry employ the same production technology may not reflect reality even in very narrowly defined industries such as in this paper. Moreover, it may be unrealistic to assume that reaching the efficient allocation of resources is an attainable policy goal given that a certain minimum level of distortions is likely to survive even in environments most conducive to allocative efficiency.

For both of these reasons, it is interesting to analyze the potential TFP gains Mexico could reap not by fully reaping resource misallocation but, at a minimum, by reaching a certain benchmark. In line with IMF (2017), we therefore use Sweden as a potential global benchmark, an economy where the efficiency of resource allocation corresponds to the 90th percentile of the distribution in the sample of countries analyzed. We find that the TFP gains that Mexico could reap by reducing distortions to the level of Sweden are still very large, at almost 70 percent. A less ambitious objective could be to bring the level of distortions in all states down to the level of Mexico's best performing (in terms of misallocation) state, Nuevo Leon (state-level estimates are discussed in more detail in the next section). Even in this case, potential TFP gains are still sizable, at some 25 percent (Figure 1).

B. State-Level TFP Gains

In this section, we compute the TFP gains associated with eliminating resource misallocation individually for each state. Our findings suggest that the variation across states is strikingly large, and larger even than the variation previous studies have found at the cross-country level. State-level TFP gains range from around 80 to 190 percent which is a broader range than the one found by Busso et al. (2013), for example, for a sample of ten Latin American countries. Even the interquartile range, which omits potential outliers, still amounts to some 73.5 percentage points. This is more than the interquartile range of potential TFP gains in the manufacturing sector across advanced economies and amounts to two thirds of the interquartile range for a large sample of developing countries reported in IMF (2017). Table 1 contains the relevant summary statistics.

[8] The TFP gains we report in Figure 1 for the manufacturing sector amount to 118 percent, less than those in Busso et al (2012) but more than those calculated by IMF (2017) based on Enterprise Survey data that do not include small and informal firms. We also find that misallocation is somewhat more severe in the manufacturing than in the services sector, in line with evidence in previous studies (Diaz et al., 2016; Busso et al., 2012).

Table 1. State-level TFP Gains from Eliminating Resource Misallocation (in percent)

	Min	10th perc.	Med	Mean	90th perc.	Max	SD	IQR
Aggregate, uncorrected	78.1	89.3	116.5	123.3	162.8	192.2	29.0	73.5
Aggregate, corrected for Industry FE	95.7	103.6	126.1	127.1	148.7	159.6	16.7	45.1

Notes: The table shows the distribution of the state-level TFP gains in percent based on Equation 11 if resources were allocated optimally across firms within all sectors under consideration within each state.
Sources: 2013 Mexican Economic Census and own compilation.

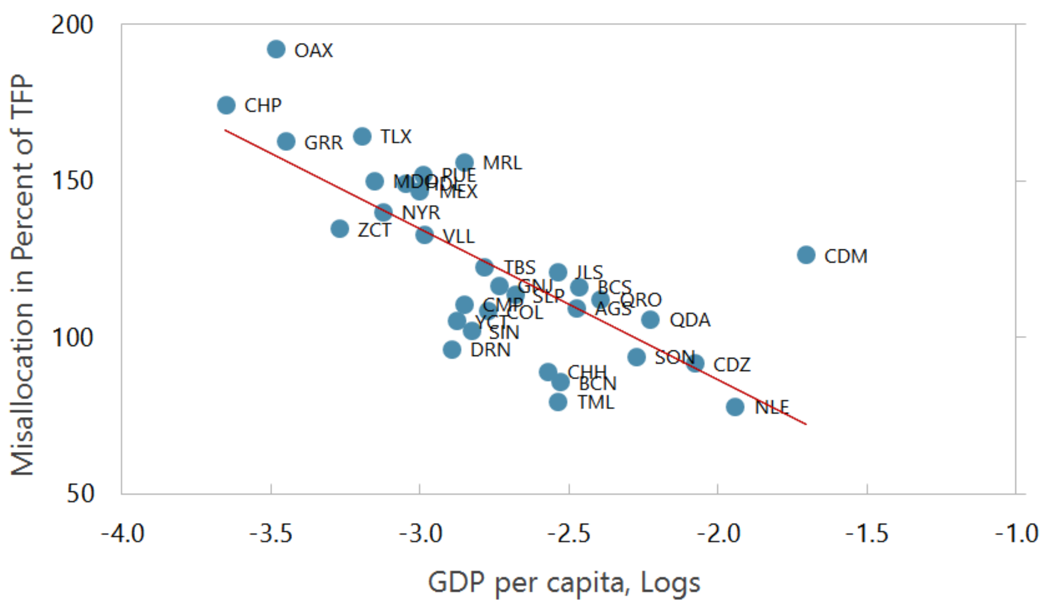
Of course, the variation in state-level TFP gains may be driven simply by differences in the industry composition of the economy of each state. For instance, if a given state has no manufacturing activity, its resource misallocation may be higher. For each industry-state pair, we therefore compute the level of TFP gains conditional on industry fixed effects through running simple OLS regressions and then recompiling aggregate state-level gains. In Table 1, we also report the summary statistics for the state-level TFP gains corrected for industry-level fixed effects. The interquartile of the TFP gains across states drops, but is still large, reaching some 45 percent.

We now proceed to evaluate whether resource misallocation would help explain income discrepancies, as predicted by Hsieh and Klenow (2009). While this claim is embedded in their model by construction, it has yet to be shown that misallocation measured indirectly through a model-based approach indeed correlates with income levels in a broader sample of countries or regions and is thus a meaningful and relevant economic concept in view of the methodological criticisms discussed above. Inklaar et al. (2017), for example, do not find evidence of a correlation between resource allocation and income levels in a sample of 52 developing and emerging market economies.

We perform a similar exercise across Mexican states. In particular, we calculate correlation estimates between state-level resource misallocation and state-level GDP per capita. Testing Hsieh and Klenow's prediction at the subnational level has several advantages, including that it allows addressing issues related to measurement error and unobserved heterogeneity. Using firm-level and national accounts data from a single source makes the data fully comparable across states. Most importantly, it allows us to exclude the same sectors from the national accounts data that we omitted from our firm-level data in estimating resource misallocation, thus ensuring full consistency in the definition of the two measures we aim to correlate.

As a first observation, the map in Annex 1 shows that high levels of misallocation are concentrated in Mexico's poorer South. Going a step further, we find that the variation in resource misallocation across states indeed correlates strongly with GDP per capita in the sectors we consider. Figure 2 shows a scatter plot comprising all of Mexico's 32 states. After omitting Mexico City (CDM) - which appears to be a clear outlier in the sense that its per capita income is higher than that one would expect based on its level of resource misallocation - we find that the correlation coefficient is a striking -0.84.[9] Using estimates of TFP gains that are corrected for industry-level effects, the correlation coefficient is almost unchanged, at -0.83. These results provide empirical support for the economic relevance of measuring resource misallocation indirectly through a model-based approach as proposed by Hsieh and Klenow (2009).

Figure 2. Correlation between State-level per Capita Incomes and Resource Misallocation, 2013



Notes: The figure shows the distribution of the state-level TFP gains in percent based on Equation 11 if resources were allocated optimally across firms within all sectors under consideration within each state.
Sources: 2013 Mexican Economic Census, National Account Statistics and own compilation.

V. Econometric Results

A. Baseline

In this section, we use industry-state-level data to examine the link between resource misallocation and observable proxies for potential distortions. For each industry in each state, the dependent variable is defined as the TFP gain that could be achieved if resources were allocated efficiently. In order to limit the effect of outliers on the results, we exclude observations with TFP

[9] Of course, dividing by the total population could bias our measure of per capita income if a large share of the population is occupied in sectors that we omit. However, the rank correlation between state-level per capita GDP and GDP per capita in the sectors under consideration is also high and amounts to -0.79.

gains from eliminating resource misallocation in the 10^{th} and 90^{th} percentiles of the distribution. This leaves a total of 2,443 industry-state level observations for the 32 Mexican states in our baseline regressions.

Our hypothesis is that there are observable proxies for distortions in the Mexican economy, both at the industry level and the subnational level, that can help explain the variation in resource misallocation across states and industries. In our baseline specification, we include the following variables as candidates for such distortions: (i) informality, given that a high share of informal firms plausibly implies that some firms enjoy unfair cost advantages, allowing them to attract more resources than they should as per their relative levels of productivity; (ii) prevalence of crime, to capture the expectation that high levels of crime would impose idiosyncratic costs on firms whose distribution across firms is unrelated to the entities' relative levels of productivity; (iii) access to finance, following the intuition that low levels of financial access imply that the financial sector's ability to help direct resources to their most productive use is impaired; (iv) access to internet technology, which can be thought of as attenuating limitations to factor movements and access to markets, especially in less densely populated areas; (v) geographical distance of firms' locations to regional population centers, given that large distances between firms and production factors could inhibit factor mobility.

In all specifications, we include a full set of state and industry fixed effects that would attenuate a potential bias from omitted variables.[10] Controlling for state fixed effects should also address the concern that the regression may suffer from an endogeneity bias arising from potential simultaneous correlations of the dependent variable and the regressors with state- or industry-specific variables such as income per capita. The only type of omitted variables the fixed effects would not allow dealing with is one whose impact on the dependent variable varies both across states and across industries (and also correlates with one or more of the explanatory variables). Such a situation could arise, for instance, in the case of state-specific distortionary policies directed at specific industries that are correlated with our state-industry-level regressors. While we can think of some examples where this could be the case (e.g., tax relief for some but not all firms in a state that is plagued by crime), we do not regard this as a first-order concern in our setup. The same holds for reverse causality. While reverse causality cannot be ruled out entirely, our dependent variable captures the efficiency of the allocation of resources across industries and states and is derived from the dispersion of firm productivities rather than from productivity levels. As such, it does not appear straightforward to argue that the dependent variable would explain variation in our regressors.

---

[10] They also allow us to zoom in more directly on the main question at hand, namely how to explain the significant variation in resource misallocation across states in narrowly defined industries, and across industries within a given state.

We run simple regressions with heteroscedasticity-consistent standard errors.[11] The regression specification is given by

$$TFP\ Gain_{js} = \alpha_j + \alpha_s + \beta_1 X_{js_{1st\ Quartile}} + \beta_2 X_{js_{2nd\ Quartile}} + \beta_3 X_{js_{3rd\ Quartile}} + \varepsilon_{js}$$

where $TFP\ Gain_{si}$ refers to the TFP gain associated with eliminating resource misallocation in industry *j* and state *s*. State and industry fixed effects are given by α_s and α_j, respectively. The baseline regressions further include a vector of explanatory variables *X* that contains our candidate distortions (see Annex 2 for definitions and data sources and Annex 3 for summary statistics). *Informality* is defined as the share of firms that did not make any social security payments or VAT payments in 2013; *Crime* is defined as the share of firms located in high crime municipalities (in which robberies per capita is in the upper quartile of the distribution); (iii) *No Financial Access* is defined as the share of firms without bank accounts; (iv) *No Internet Use* is defined as the average share of employees who do not use the internet at work; and (v) *Distance* is defined as the average distance of firms in a given industry-state pair from the closest population center.[12]

Instead of using the continuous variables themselves, our baseline specification employs a set of dummy variables. This approach reflects the finding that the relationship between most of our regressors and the dependent variable is not strictly linear (as illustrated by robustness regressions reported below that use the underlying continuous variables). The dummies indicate whether an observation falls into the first (i.e., distortion least severe), second, third or fourth (i.e., distortion most severe) quartile of the underlying distribution of the continuous variable. For each distortion, we then include three of the four dummies in the regression, where the dummy pertaining to the fourth quartile is the omitted variable. The coefficient on each of the three included dummies thus measures the effect on misallocation relative to the case where the distortion is most severe. For instance, the coefficient on the first quartile measures the difference in resource misallocation in industry-state pairs where the distortion is least severe relative to industry-state pairs where the distortion is most severe, conditional on other factors. In other words, we expect all dummies to carry negative coefficients, with the most negative coefficient associated with the first quartile, and the least negative one associated with the third quartile.

---

[11] The results are qualitatively robust to using standard errors that are clustered at the sector and state level as we report below.

[12] Population centers are defined as cities with more than 500,000 inhabitants. Similar to the crime variable, we exploit variation in the location of firms across sectors within states, thereby ensuring that the effects of the distance variable is not captured by the state fixed effects. We only calculate the ‘as the crow flies’ distance which could of course be misleading, especially in Mexico’s mountainous center.

The first regression in Table 2 presents the results of our baseline specification when all 15 dummies are included jointly in the regression. Most of them are statistically significant at least at the 90 percent level of confidence, and their coefficients carry the expected negative signs. In the case of the *Informality* variable, for example, the results suggest that moving from the fourth quartile to the third quartile lowers the TFP gains associated with eliminating resource misallocation by 11 percentage points while moving to the second or even the first quartile would reduce resource misallocation by an additional 3 or 5 percentage points, respectively. In other words, higher levels of informality are associated with higher resource misallocation, and the biggest reduction in misallocation would come with reducing informality from very high to high levels.

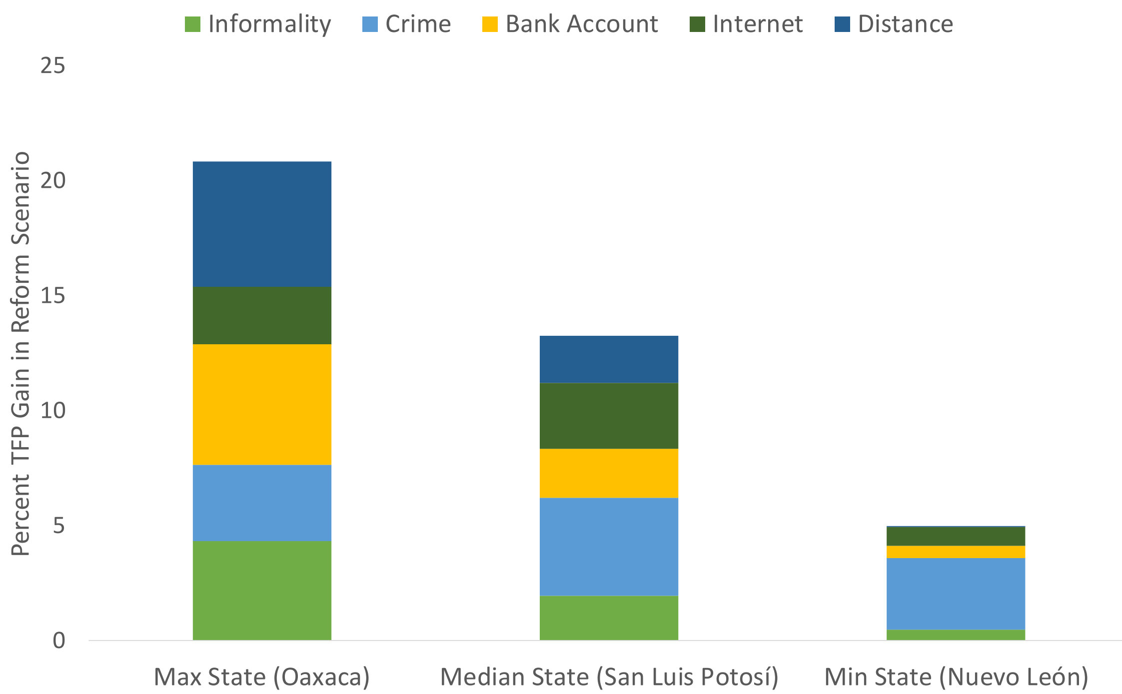
We find similarly clear-cut results in the cases of the *No Financial Access* and *Distance* variables. In both cases, moving from the fourth quartile to the third, second and first quartiles is associated with relatively gradually falling levels of the TFP gains associated with eliminating resource misallocation (for a total reduction of 18 and 12 percentage points, respectively). In the case of the *Crime* and *No Internet Use* variables, we also find that the first quartile is associated with the highest reduction in resource misallocation (for a total reduction of 14 and 9 percentage points of misallocation, respectively), but not all dummies are significant with a negative coefficient. It thus appears that both variables matter, but that the impact is not strictly linear. Intuitively, it is conceivable that only major reductions in crime matter for resource misallocation.[13]

To examine the economic significance of the results, we conduct a simple policy experiment: we calculate the potential TFP gain for each Mexican state that would, according to our baseline regression, be associated with addressing the distortions reflected in the regressors of our baseline specification. In particular, we assume that the level of each distortion in each industry-state pair is lowered to match the level of distortions in the first (least severe) quartile of the distribution of the respective distortion, and calculate the implied impact on TFP. In some sense, we are thus estimating the gains each state could obtain by moving to the domestic frontier as represented by a synthetic state in which the severity of the distortions is low. While we do not identify the exact reform measures underlying this scenario, we believe that the objective may be achievable in the sense that it would imply moving to the domestic rather than the international frontier.

---

[13] In the case of the internet variable, the finding that the third quartile dummy is significant with the opposite sign is surprising and difficult to explain.

Figure 3. TFP Gains in Reform Scenario Across States



Notes: The figure shows the hypothetical state-level TFP gains in percentage points for a reform scenario that is based on the results from the baseline regression. We assume that the level of each distortion in each industry-state pair within a given state is lowered to match the level of distortions in the first (least severe) quartile of the distribution of the respective distortion across all industry-state pairs, and calculate the implied impact on TFP. For illustrative purposes, we choose three different states (Oaxaca, San Luis Potosí and Nuevo León) which correspond to the states where the level of misallocation corresponds to the maximum, median and minimum, respectively.
Sources: 2013 Mexican Economic Census and own compilation.

Figure 3 shows the percentage point change in resource misallocation that would result from the experiment, broken down into the contributions of the five types of distortions considered for the state with the largest expected TFP gain (Oaxaca), the state with the smallest gain (Nuevo León), and the state with the median gain (San Luis Potosí). The potential gains associated with the reform scenario are economically meaningful: For example, the TFP gain associated with the reform scenario in Oaxaca would exceed 20 percentage points. Even when excluding the *Distance* variable from these simulations – which may arguably not be directly responsive to reform initiatives – the impact lies still at around 15 percent of GDP.[14] At the same time, the variation across states is striking. For example, the gains for relatively richer states with less resource misallocation at the outset such as Nuevo León are in the order of magnitude of only 5

[14] The Distance variable can be thought of as capturing limitations to factor movements that could be addressed through policy initiatives such as targeted infrastructure investment.

percentage points. In other words, the reform scenario would not only boost productivity in Mexico as a whole, but it would also lower disparities in the level of resource misallocation across states.[15]

B. Robustness

In the remaining regressions of Table 2, we test the robustness of our results to alternative measures of the distortions included in our baseline. Regression 2 includes a narrower definition of informality under which firms are only considered informal if they do not make any social security payments, VAT payments, income tax payments, or excise tax payments. [16] All three informality dummies remain highly significant and the coefficients are almost unchanged compared to the baseline specification, suggesting that our findings are not sensitive to the precise definition of informality. In Regression 3, we use a variable measuring the incidence of homicides instead of robberies. The results are once again similar to the baseline. In Regression 4, we replace the *No Financial Access* variable by an alternative indicator of lack of financial access that measures the absence of bank credit. While the coefficient of the "very low" is negative as expected, it is smaller than in the case of the *No Financial Access* variable, and the remaining two dummies are insignificant and do not carry the expected signs. This suggests that access to credit does not contribute to the efficiency of resource allocation in the same way as access to a bank account. Finally, Regression 5 replaces our indicator of *No Internet Use* with an indicator measuring the share of employees not using computers at work. Once again, the results are very similar to the baseline.

Table 3 presents some additional robustness checks. Regression 1 weights observations by the size of each industry-state pair (based on the total sales of each firm), with the results only marginally affected. Regression 2 illustrates that the regressors remain statistically significant when we cluster the error terms by state and industry. In both regressions, our results remain qualitatively unchanged. Regression 3 replaces our baseline regressors with the underlying continuous variables. While all five variables show the expected coefficient signs, two of the five are not significant, signaling that the relationships between the regressors and the dependent variable are not strictly linear.

Regressions 4-8 include additional control variables in the baseline specification, but our results remain qualitatively robust. Regression 5 includes an indicator of the share of firms that reside in municipalities with a high population density to the specification. Intuitively, one might expect

---

[15] The correlation coefficient between the simulated TFP gain associated with the reform and the level of resource misallocation before the reform is 0.55; a regression of the former on the latter yields an R^2 of 0.3The results suggest that the reform gain is 1 percentage point higher for every 15 percentage points more in initial resource misallocation.

[16] Busso et al. (2012) use a broader definition and consider firms as informal if the level of social security payments falls short of the 18 percent of wages and salaries that should have been paid.

that a higher population density could lead to more competition and labor mobility. However, the variable is neither statistically significant nor does it carry the expected negative coefficient. Similarly, we include the ratio of firms per capita in Regression 6, but, once again, the hypothesis is not supported by the data. Regression 7 introduces a variable measuring the number of firms per industry-state pair. In this case, both dummies are significant with positive coefficients. While an interpretation of this finding is not straightforward, it is comforting that our key results remain qualitatively unaffected. Regression 8 includes an indicator of the share of firms that have received FDI. Intuitively, one may conjecture that FDI would be attracted by a more competitive environment in which resources are more likely to be allocated to their most productive use. In line with this hypothesis, the coefficient on the FDI variable is negative, but it is not statistically significant. Finally, Regression 9 includes municipal GDP per capita averaged across firms to control more precisely (in addition to the state fixed effect included in the regressions) for differences in income levels. The two dummies turn out to be insignificant and the results broadly unchanged.

C. Extension: Market Concentration and Resource Misallocation

We now consider the role of an additional potential distortion that could explain high levels of resource misallocation, namely market concentration. To the extent that market concentration is positively correlated with market power, one may expect dominant firms to attract a larger share of resources than warranted by their relative level of productivity. To test this prediction, we calculate the Herfindahl index by industry and state in two alternative ways, namely based on the number of firms' employees and on their total sales. Introducing the employee-based concentration index in Regression 1 in Table 4 yields a surprising result: its coefficient is negative and highly significant, suggesting that higher levels of concentration reduce rather than increase resource misallocation. A potential explanation is that concentration is not driven by market power but productivity differentials. This may be a particularly good explanation in Mexico where a large number of small, unproductive and informal firms attract an outsized share of the economy's resources.

Indeed, high levels of concentration could be associated with lower levels of resource misallocation in a industry in which a small number of productive and formal firms attempt to attract resources from a large number of unproductive informal firms. Industries in which the group of formal and productive firms manages to compete successfully with their informal and less productive counterparts would be characterized by both higher market concentration and lower resource misallocation. This implies that concentration has a positive impact on resource misallocation in relatively more formal industries, and a negative impact in relatively more informal industries.

The remaining regressions in Table 4 test this hypothesis. We include interaction terms between our concentration terms and our measure of informality in the specification. In order to limit the number of interactions, we replace the two informality terms in the baseline specification with the underlying continuous variable for the purposes of this exercise. Regression 2 confirms the negative unconditional link between concentration and resource misallocation in the modified baseline. Regression 3 adds the interaction term between informality and the Herfindahl index. All three variables of interest turn out to be highly significant. The informality measure retains its positive coefficient which is now somewhat larger than in the modified baseline. Importantly, however, the concentration term now carries the expected positive coefficient while the interaction term shows a negative coefficient. In other words, higher levels of concentration do appear to be associated with market power and resource misallocation for some industries, but the effect switches sign in industries with a high prevalence of informality.

These results are robust to using alternative measures of market concentration (Regressions 4 and 5). Once again, the coefficient on the concentration measure switches sign (although the variable is not significant in this case) when an interaction between informality and concentration is added into the regression. Finally, Regressions 6 and 7 show that a similar link exists when we include an indicator of average firm size in place of the concentration measure. It appears that larger firm size tends to be associated with more resource misallocation but not in industries in which informality is prevalent.

D. Extension: Corruption and Resource Misallocation

Another potentially important driver of resource misallocation in Mexico is corruption. For example, an official who awards a contract based on bribery rather than relative productivity and cost of production directly engages in a misallocation of resources. In this paper, we use survey data from the 2013 Encuesta Nacional de Calidad e Impacto Gubernamental (ENCIG) to compute perception- and experience based measures of corruption in public service provision with the purpose of linking them to resource misallocation (see Appendix 1).[17] The survey collects information from respondents on their experience with and their perception of procedures and services provided by different levels of government. Based on the survey responses, our indicators count the share of respondents by state who would agree, for example, that corruption is frequent or very frequent in public service provision.

The reason why we did not include our corruption indicators in the baseline specification is that they only vary across—but not within—states and would thus drop out in any regression incorporating state-level fixed effects.[18] Our strategy in establishing a link between corruption and misallocation thus relies on a difference-in-differences approach similar to the one proposed by Rajan and Zingales (1998). To do this, we measure the exposure of a given industry to

[17] We thank Frederic Lambert for providing us with the indicators.

[18] The number of survey respondents is too low to construct measures of corruption at higher levels of geographical disaggregation.

corruption by the extent to which government procurement is an important source of demand in the industry. In doing so, we make use of firm-level information included in the Economic Census data indicating whether or not the government is the most important client for a given firm. We then define a dummy variable that takes the value one in industries in which the share of firms whose most important client is the government is in the upper quartile of the distribution, and zero otherwise. The interaction terms between this indicator and our measures of corruption is our variable of interest.

Table 5 presents the results of our difference-in-differences approach. Regression 1 includes the interaction term between the dummy variable for the importance of government procurement and our first corruption indicator in the baseline (note that both level terms drop out given state- and industry level fixed effects). The corruption indicator measures the share of respondents who have experienced corruption in their own interaction with public service providers or employees of the government. The interaction term turns out to be a highly significant determinant of resource misallocation. The coefficient is positive as expected, signaling that higher levels of corruption raise resource misallocation more in industries in which government procurement plays an important role in final demand. The following regressions (2, 3 and 4) provide additional confirmation for our hypothesis. Each of them includes an alternative indicator of corruption in the interaction term which remains significant in Regressions 2 and 3, and at least continues to carry a positive coefficient in Regression 4.

VI. Conclusion

Mexico implemented sweeping structural reforms during the mid-1990s, but productivity growth has remained puzzlingly low. The analysis in this paper suggests that resource misallocation may have played a key role in holding back productivity growth. Our main contribution is to analyze the determinants of resource misallocation in Mexico across industries and states. We find that it can be explained in part by some of Mexico’s main developmental challenges such as high levels of informality, crime, corruption and market concentration as well as insufficient access to financial and internet services and the degree of geographic dispersion of firms. The findings suggest that addressing these challenges could yield aggregate TFP gains that would be economically sizable even if potential reforms aim at reducing distortions to levels close to the domestic rather than the international frontier.

A second important finding arises from our focus on the subnational dimension of resource misallocation. The analysis suggests that the variation in resource misallocation across Mexican states rivals that found by previous studies at the cross-country level. We exploit this variation and find evidence of a close correlation between subnational income discrepancies and levels of resource misallocation. The finding validates the approach of estimating misallocation indirectly through a model based approach à la Hsieh and Klenow (2009).

The findings in our paper highlight the need for continued implementation of the structural reform program Mexico has embarked on (Saborowski, 2017). They underscore the importance of boosting competition and access to financial and telecommunications services and strengthening the rule of law to root out corruption, crime and labor informality. The link between the geographic isolation of some regions and resource misallocation highlights the importance of policies that increase the mobility of production factors in some of Mexico's less developed regions. Such policies could include targeted physical or transportation infrastructure investments.

Table 2. Baseline Regressions

Dependent Variable: Resource Misallocation

	(1)	(2)	(3)	(4)	(5)
Informality, 3rd Quartile	-11.373***	-13.001***	-10.940***	-15.781***	-11.540***
	[3.008]	[2.981]	[2.985]	[2.521]	[2.995]
Informality, 2nd Quartile	-14.647***	-17.054***	-14.319***	-23.878***	-14.386***
	[4.034]	[4.022]	[4.030]	[3.366]	[4.000]
Informality, 1st Quartile	-16.609***	-19.830***	-16.317***	-25.857***	-15.742***
	[5.174]	[5.230]	[5.165]	[4.530]	[5.132]
Crime, 3rd Quartile	-8.372***	-8.439***	-11.297**	-8.018**	-8.673***
	[3.065]	[3.064]	[4.847]	[3.115]	[3.054]
Crime, 2nd Quartile	-4.321	-4.200	-8.221	-3.665	-4.690
	[4.555]	[4.545]	[5.956]	[4.612]	[4.538]
Crime, 1st Quartile	-13.793*	-13.693*	-11.723*	-12.913*	-13.747*
	[7.450]	[7.447]	[6.503]	[7.506]	[7.504]
No Financial Access, 3rd Quartile	-7.234**	-6.649**	-7.074**	0.375	-7.194**
	[2.933]	[2.904]	[2.939]	[2.022]	[2.939]
No Financial Access, 2nd Quartile	-18.082***	-17.106***	-18.031***	1.478	-17.806***
	[3.872]	[3.809]	[3.867]	[2.350]	[3.880]
No Financial Access, 1st Quartile	-18.460***	-17.254***	-17.916***	-4.409	-18.062***
	[4.779]	[4.716]	[4.772]	[2.783]	[4.785]
No Internet Use, 3rd Quartile	4.275**	4.471**	4.233**	4.194*	4.824**
	[2.131]	[2.130]	[2.133]	[2.147]	[2.173]
No Internet Use, 2nd Quartile	-1.878	-1.438	-2.332	-3.424	-1.702
	[2.773]	[2.774]	[2.763]	[2.749]	[2.795]
No Internet Use, 1st Quartile	-8.513**	-8.052**	-8.624**	-10.288***	-11.891***
	[3.852]	[3.851]	[3.847]	[3.831]	[3.790]
Distance, 3rd Quartile	-7.552*	-7.479*	-7.964*	-6.794	-7.167*
	[4.150]	[4.145]	[4.180]	[4.130]	[4.156]
Distance, 2nd Quartile	-10.276**	-10.159*	-10.664**	-9.620*	-9.997*
	[5.212]	[5.202]	[5.278]	[5.205]	[5.196]
Distance, 1st Quartile	-12.095**	-11.864*	-12.355**	-11.728*	-12.014**
	[6.115]	[6.103]	[6.189]	[6.104]	[6.095]
Observations	2,443	2,443	2,443	2,443	2,443
R-squared	0.603	0.604	0.604	0.599	0.605
Alternative definition for		Informality	Crime	Financial	Internet

Robust standard errors in brackets; all specifications include industry and state fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Table 3. Robustness Checks

Dependent Variable: Resource Misallocation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Informality, 3rd Quartile	-10.649**	-11.051***		-11.058***	-10.958***	-11.061***	-10.090**	-11.067***
	[4.232]	[3.380]		[3.004]	[3.002]	[3.001]	[3.918]	[2.999]
Informality, 2nd Quartile	-12.443**	-14.424***		-14.438***	-14.340***	-14.430***	-11.964**	-14.499***
	[5.876]	[4.336]		[4.048]	[4.047]	[4.042]	[5.082]	[4.046]
Informality, 1st Quartile	-13.573*	-16.256***		-16.268***	-16.300***	-16.257***	-14.506**	-16.340***
	[7.275]	[5.273]		[5.184]	[5.179]	[5.177]	[6.126]	[5.187]
Crime, 3rd Quartile	-2.409	-5.288**		-5.283*	-4.716	-5.274*	-7.159**	-5.113*
	[4.189]	[2.167]		[3.040]	[3.106]	[3.038]	[3.246]	[3.074]
Crime, 2nd Quartile	-1.655	-5.786		-5.785	-5.298	-5.780	-13.303**	-5.588
	[7.340]	[5.459]		[4.681]	[4.709]	[4.680]	[5.301]	[4.703]
Crime, 1st Quartile	14.775	-9.628		-9.615	-8.984	-9.616	-14.869*	-9.486
	[13.618]	[9.328]		[7.062]	[7.098]	[7.054]	[7.870]	[7.059]
No Financial Access, 3rd Quart	-12.829***	-7.230**		-7.227**	-7.356**	-7.236**	-10.275***	-7.249**
	[4.240]	[3.302]		[2.950]	[2.951]	[2.950]	[3.833]	[2.952]
No Financial Access, 2nd Quart	-24.249***	-18.227***		-18.224***	-18.395***	-18.227***	-22.919***	-18.223***
	[5.521]	[4.337]		[3.900]	[3.906]	[3.900]	[4.795]	[3.899]
No Financial Access, 1st Quarti	-25.523***	-18.288***		-18.283***	-18.501***	-18.281***	-23.528***	-18.361***
	[7.050]	[4.452]		[4.806]	[4.811]	[4.807]	[5.722]	[4.809]
No Internet Use, 3rd Quartile	4.639	4.413**		4.414**	4.371**	4.408**	11.913**	4.397**
	[3.066]	[2.281]		[2.136]	[2.138]	[2.137]	[5.340]	[2.136]
No Internet Use, 2nd Quartile	-3.192	-1.959		-1.963	-1.952	-1.973	6.260	-1.997
	[4.157]	[2.957]		[2.774]	[2.769]	[2.774]	[5.610]	[2.770]
No Internet Use, 1st Quartile	-11.379**	-8.327**		-8.327**	-8.249**	-8.340**	0.425	-8.376**
	[5.169]	[3.973]		[3.845]	[3.834]	[3.845]	[6.199]	[3.840]
Distance, 3rd Quartile	-14.240**	-7.767**		-7.799*	-7.842*	-7.763*	-8.553*	-7.677*
	[7.158]	[3.852]		[4.198]	[4.166]	[4.173]	[4.736]	[4.174]
Distance, 2nd Quartile	-17.636**	-10.158**		-10.231*	-10.138*	-10.143*	-12.295**	-10.043*
	[7.992]	[4.710]		[5.310]	[5.222]	[5.234]	[5.988]	[5.238]
Distance, 1st Quartile	-21.562**	-11.945*		-12.096*	-11.752*	-11.914*	-14.339**	-11.826*
	[9.042]	[6.077]		[6.556]	[6.129]	[6.142]	[7.003]	[6.146]
Informality			47.662***					
			[9.647]					
Crime			22.498***					
			[8.257]					
No Bank Account			18.050*					
			[9.316]					
No Internet Use			0.105					
			[0.201]					
Distance			0.000					
			[0.000]					
High Population				0.452				
				[7.257]				
Firms per Capita					145.748			
					[136.287]			
Number of Firms						-0.000		
						[0.000]		
FDI							-2.895	
							[5.626]	
GDP per Capita								0.005
								[0.013]
Observations	2,443	2,443	2,443	2,443	2,443	2,443	1,869	2,443
R-squared	0.800	0.603	0.596	0.603	0.603	0.603	0.640	0.603
Weighted	Yes	No	No	No	No	No	No	No
Standard errors	Robust	Clustered	Robust	Robust	Robust	Robust	Robust	Robust

Standard errors in brackets; all specifications include industry and state fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Market Concentration Regressions

Dependent Variable: Resource Misallocation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Informality, 3rd Quartile	-10.803***						
	[2.980]						
Informality, 2nd Quartile	-13.994***						
	[4.015]						
Informality, 1st Quartile	-15.633***						
	[5.170]						
Crime, 3rd Quartile	-8.455***	-8.664***	-8.842***	-8.559***	-8.678***	-8.477***	-8.376***
	[3.060]	[3.026]	[2.960]	[2.997]	[2.964]	[3.027]	[3.036]
Crime, 2nd Quartile	-4.832	-5.200	-6.379	-5.800	-5.913	-4.695	-4.619
	[4.544]	[4.474]	[4.400]	[4.415]	[4.403]	[4.481]	[4.471]
Crime, 1st Quartile	-13.435*	-13.086*	-14.729**	-12.589	-13.825*	-13.024*	-13.151*
	[7.620]	[7.564]	[7.109]	[7.693]	[7.279]	[7.376]	[7.393]
No Financial Access, 3rd Quartile	-7.150**	-9.991***	-7.886***	-8.943***	-7.740***	-10.101***	-9.490***
	[2.915]	[2.438]	[2.413]	[2.378]	[2.377]	[2.455]	[2.449]
No Financial Access, 2nd Quartile	-17.828***	-18.862***	-15.730***	-17.908***	-16.370***	-18.896***	-17.429***
	[3.857]	[3.410]	[3.418]	[3.337]	[3.345]	[3.425]	[3.445]
No Financial Access, 1st Quartile	-18.709***	-15.760***	-12.853***	-15.399***	-13.879***	-15.201***	-13.573***
	[4.761]	[4.695]	[4.661]	[4.611]	[4.585]	[4.710]	[4.738]
No Internet Use, 3rd Quartile	4.767**	4.877**	4.971**	6.091***	5.850***	4.241**	4.753**
	[2.125]	[2.134]	[2.100]	[2.087]	[2.069]	[2.144]	[2.144]
No Internet Use, 2nd Quartile	-1.427	-0.756	-0.222	0.586	0.405	-1.568	-0.130
	[2.772]	[2.761]	[2.710]	[2.691]	[2.640]	[2.773]	[2.822]
No Internet Use, 1st Quartile	-8.061**	-5.282	-4.663	-3.891	-4.032	-6.430*	-4.126
	[3.856]	[3.882]	[3.793]	[3.822]	[3.728]	[3.889]	[3.961]
Distance, 3rd Quartile	-7.309*	-7.482*	-8.769**	-6.999*	-8.041*	-7.175*	-7.349*
	[4.121]	[4.116]	[4.099]	[4.133]	[4.210]	[4.154]	[4.148]
Distance, 2nd Quartile	-9.543*	-9.917*	-10.636**	-8.544	-9.685*	-10.123*	-10.347**
	[5.183]	[5.162]	[5.164]	[5.198]	[5.287]	[5.203]	[5.200]
Distance, 1st Quartile	-12.207**	-12.681**	-12.921**	-12.391**	-12.941**	-11.995**	-12.261**
	[6.094]	[6.049]	[6.053]	[6.072]	[6.124]	[6.073]	[6.053]
Herfindahl by Employment	-35.518***	-35.389***	150.770***				
	[13.571]	[13.623]	[35.236]				
Informality		40.515***	75.715***	41.782***	69.146***	44.429***	48.842***
		[9.213]	[10.474]	[9.210]	[10.095]	[9.178]	[9.374]
Interaction Informality/Herfindahl			-338.680***				
			[57.209]				
Herfindahl by Sales				-89.880***	49.020		
				[11.794]	[29.820]		
Interaction Informality/Herfindahl					-220.676***		
					[43.047]		
Firm Size						0.054**	0.140***
						[0.022]	[0.042]
Interaction Informality/Firm Size							-0.618***
							[0.222]
Observations	2,443	2,443	2,443	2,443	2,443	2,443	2,443
R-squared	0.606	0.607	0.619	0.620	0.626	0.606	0.609

Robust standard errors in brackets; all specifications include industry and state fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Table 5. Corruption Regressions

Dependent Variable: Resource Misallocation

	(1)	(2)	(3)	(4)
Informality, 3rd Quartile	-11.370***	-11.207***	-11.472***	-11.353***
	[3.002]	[3.010]	[3.006]	[3.016]
Informality, 2nd Quartile	-14.607***	-14.640***	-14.713***	-14.626***
	[4.029]	[4.032]	[4.031]	[4.043]
Informality, 1st Quartile	-16.455***	-16.480***	-16.769***	-16.586***
	[5.182]	[5.165]	[5.170]	[5.180]
Crime, 3rd Quartile	-8.388***	-8.390***	-8.417***	-8.358***
	[3.050]	[3.059]	[3.053]	[3.064]
Crime, 2nd Quartile	-4.176	-4.255	-4.286	-4.324
	[4.536]	[4.558]	[4.546]	[4.557]
Crime, 1st Quartile	-13.476*	-13.666*	-13.170*	-13.795*
	[7.401]	[7.422]	[7.410]	[7.451]
No Financial Access, 3rd Quartile	-7.199**	-7.276**	-7.154**	-7.242**
	[2.921]	[2.930]	[2.929]	[2.938]
No Financial Access, 2nd Quartile	-17.905***	-18.003***	-17.967***	-18.092***
	[3.862]	[3.867]	[3.871]	[3.876]
No Financial Access, 1st Quartile	-18.245***	-18.340***	-18.316***	-18.477***
	[4.770]	[4.770]	[4.777]	[4.786]
No Internet Use, 3rd Quartile	4.437**	4.494**	4.269**	4.287**
	[2.128]	[2.129]	[2.130]	[2.135]
No Internet Use, 2nd Quartile	-1.859	-1.739	-1.952	-1.871
	[2.771]	[2.767]	[2.772]	[2.776]
No Internet Use, 1st Quartile	-8.242**	-8.262**	-8.496**	-8.516**
	[3.850]	[3.842]	[3.851]	[3.853]
Distance, 3rd Quartile	-7.322*	-7.731*	-7.708*	-7.583*
	[4.151]	[4.151]	[4.147]	[4.166]
Distance, 2nd Quartile	-10.249**	-10.661**	-10.176*	-10.314**
	[5.206]	[5.210]	[5.206]	[5.228]
Distance, 1st Quartile	-12.214**	-12.343**	-12.113**	-12.126**
	[6.110]	[6.115]	[6.107]	[6.125]
Interaction Procurement/Corruption Experience	1.050**			
	[0.469]			
Interaction Procurement/Corruption Very Frequent		0.295**		
		[0.144]		
Interaction Procurement/Corruption Heard About			0.347*	
			[0.191]	
Interaction Procurement/Corruption Top 3				0.033
				[0.279]
Observations	2,443	2,443	2,443	2,443
R-squared	0.605	0.605	0.605	0.604

Robust standard errors in brackets; all specifications include industry and state fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

References

Busso, M., Fazio, M. and Levy, S., 2012, "(In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico," IDB Working Paper 341, Inter-American Development Bank.

Busso, M., and Madrigal, L., 2013, "Productivity and resource misallocation in Latin America," *The BE Journal of Macroeconomics*, Vol.13(1), 903–32.

Calligaris, S., Del Gatto, M., Hassan, F., Ottaviano, G. and Schivardi, F., 2016, "Italy's Productivity Conundrum. A Study on Resource Misallocation in Italy," European Economy - Discussion Papers 030, European Commission.

Dias, D., Richmond, C. and Robalo Marques, C., 2016, "A Tale of Two Sectors; Why is Misallocation Higher in Services than in Manufacturing?," IMF Working Paper 16/220 (International Monetary Fund, Washington, D.C.).

Haltiwanger, J., Kulick, R and Syverson, C., 2018, "Misallocation Measures: The Distortion That Ate the Residual," NBER Working Papers 24199.

Hanson, G., 2010, "Why Isn't Mexico Rich?", *Journal of Economic Literature*, Vol. 48(4), 987–1004.

Hopenhayn, H., 2014, "Firms, Misallocation, and Aggregate Productivity: A Review," *Annual Review of Economics*, Vol. 6(1), 735–70.

Hsieh, C. T., and Klenow, P. J., 2009, "Misallocation and manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–48.

IMF, 2017. "*IMF Fiscal Monitor: Achieving More with Less*," International Monetary Fund, Washington, D.C.

Inklaar, R., Lashitew, A. and Timmer, M., 2017, "The Role Of Resource Misallocation In Cross-Country Differences In Manufacturing Productivity," *Macroeconomic Dynamics*, Vol. 21(03), 733–56.

Keller, W., 2004, "International Technology Diffusion," *Journal of Economic Literature*, Vol. 42(3), 752–82.

Levy, S. and Rodrik, D. 2017, “The Mexican Paradox,” available at http://www.project-syndicate.org.

Mayorga Garrido-Cortés, J. 2017, “Resource misallocation in Mexico,” Tesis de maestría, Centro de Investigación y Docencia Económicas, available at http://repositorio-digital.cide.edu.

Rajan, R. Zingales. L., 1998, “Financial Dependence and Growth.” The American Economic Review, vol. 88(3), 559-586.

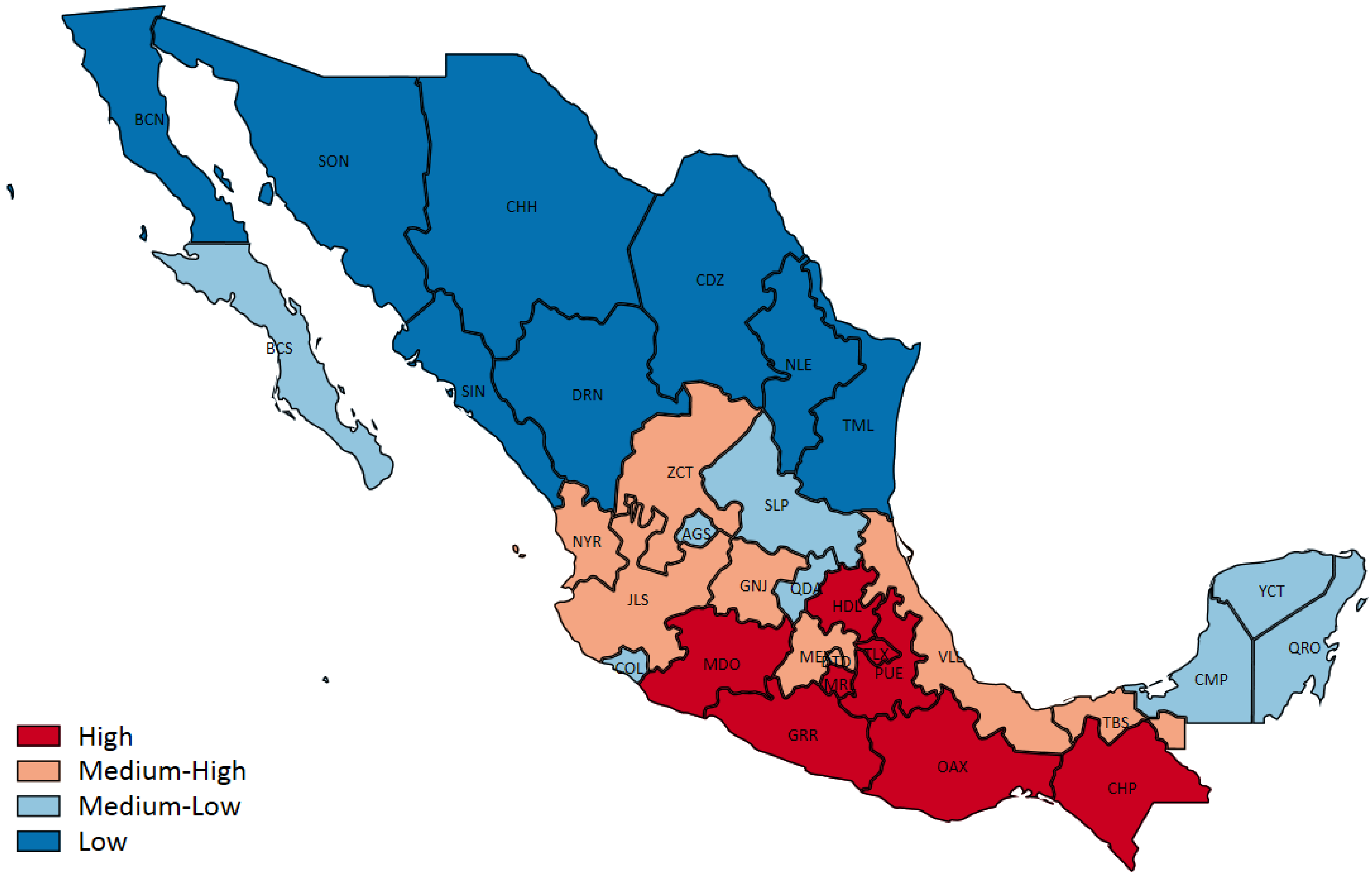
Restuccia, D., and Rogerson, R., 2008, “Policy Distortions and Aggregate Productivity with Heterogeneous Plants,” *Review of Economic Dynamics*, Vol. 11(4), 707–20.

Restuccia, D., and Rogerson, R., 2013, “Misallocation and productivity,” *Review of Economic Dynamics*, Vol. 16(1), 1–10.

Restuccia, D., and Rogerson, R., 2017, “The Causes and Costs of Misallocation,” *Journal of Economic Perspectives*, Vol. 31(3), 151–74.

Saborowski, C., 2017, “Mexico’s Structural Reform Agenda: Early Signs of Success,” Selected Issues Paper, 2017 IMF Art. IV Consultation for Mexico.

Annex 1. Resource Misallocation by Mexican State



Notes: The map shows the level of resource misallocation (expressed in quartiles) by state which corresponds to the TFP gains if resources were allocated optimally across firms within all sectors under consideration within each state.
Sources: 2013 Mexican Economic Census and own compilation.

Annex 2. Variable Definitions and Sources

Variable	Definition	Source
	Baseline Regressors	
Informality	Share of firms paying no social security and no VAT.	2013 Economic Census
alternative proxy	Share of firms paying no social security, VAT, income or excise tax.	
Crime	Share of firms in high crime municipalities (in which robberies per capita is in the upper quartile of the distribution).	2010 SIMBAD; 2010 Population Census
alternative proxy	Share of firms in high crime municipalities (in which homicides per capita is in the upper quartile of the distribution).	
No Bank Account	Avg. across firms of dummy that takes value 1 when firm has a bank account.	2013 Economic Census
alternative proxy	Avg. across firms of dummy that takes value 1 when firm has bank credit.	
No Internet Use	Average share of employees not using the internet for their work.	2013 Economic Census
alternative proxy	Average share of employees not using computers for their work.	
Distance	Average distance between the locality (e.g., city / town) of the firm and the closest population center (population > 500,000).	2013 Economic Census and own computation
	Regressors in Extensions	
Herfindahl by Employment	Herfindahl index calculated based on each firm's number of employees.	2013 Economic Census
Herfindahl by Sales	Herfindahl index calculated based on each firm's total sales.	2013 Economic Census
Firm Size	Average sales by firm.	2013 Economic Census
Procurement	Share of firms reporting the government as their most important client.	2013 Economic Census
Corruption Experience	Proportion of respondents who say that they experienced corruption in dealing with the government.	Encuesta Nacional de Calidad e Impacto Gubernamental 2013
Corruption Very Frequent	Proportion of respondents who answer that corruption is very frequent.	Encuesta Nacional de Calidad e Impacto Gubernamental 2013
Corruption Heard About	Proportion of respondents who heard from relatives/friends that there are people who had to pay bribes.	Encuesta Nacional de Calidad e Impacto Gubernamental 2013
Corruption Top 3	Proportions of respondents who consider corruption one of the top three issues in the state government.	Encuesta Nacional de Calidad e Impacto Gubernamental 2013
	Other Regressors	
High Population	Share of firms in municipalities within 4th quartile of population density.	2010 SIMBAD
Firms per Capita	Number of firms per capita.	2013 Economic Census; 2010 Population Census
Number of Firms	Number of firms.	2013 Economic Census
FDI	Share of firms engaged in FDI relationships.	2013 Economic Census
GDP per Capita	Average GDP per capita across firms of municipalities the firms are located in.	INEGI's National Account Statistics

Annex 3. Summary Statistics

Variable	Mean	Minimum	Maximum
Baseline Regressors			
Informality	0.634	0	1
alternative proxy	0.628	0	1
Crime	0.327	0	1
alternative proxy	0.334	0	1
No Bank Account	0.622	0	1
alternative proxy	0.815	0	1
No Internet Use	0.967	0	1
alternative proxy	0.942	0	1
Distance (in meters)	85879.9	114	370799.4
Regressors in Extensions			
Herfindahl by Employment	0.057	0.000	0.859
Herfindahl by Sales	0.075	0.000	0.717
Firm Size	11.950	1.244	1000.164
Procurement	0.005	000.364	1
Corruption Experience	9.345	4.484	16.596
Corruption Very Frequent	44.206	18.833	62.164
Corruption Heard About	0.309	0.173	0.501
Corruption Top 3	460.34	33.987	54.606
Other Regressors			
High Population	0.251	0	1
Firms per Capita	0.043	0.023	0.130
Number of Firms	880.9	10	127602
FDI	0.072	0	1
GDP per Capita	72.067	3.720	1500.0031