FLORIAN MISCH
International Monetary Fund, Washington, D.C., USA
CHRISTIAN SABOROWSKI
International Monetary Fund, Washington, D.C., USA

The Drivers and Consequences of Resource Misallocation: Exploiting Variation across Mexican Industries and States

ABSTRACT This paper explores the role of specific structural distortions in explaining Mexico's weak productivity growth through the misallocation of resources across firms. The paper makes two contributions. First, we show that there is a close correlation between the level of resource misallocation and per capita income across Mexican states. Second, we exploit the large variation in resource misallocation within industries and across states, together with unusually rich data at the establishment, local, and industry levels, to shed light on its determinants. We identify several well-defined and observable distortions that have a statistically and economically meaningful effect on productivity via resource misallocation. In particular, we find that misallocation rises with the prevalence of labor informality, crime, corruption, and market concentration and with weaker access to financial and telecommunications services.

JEL Codes: D24, O12, O47
Keywords: Resource misallocation, productivity, distortions

Mexico's low average per capita growth rate over the last two decades, and in particular its negative productivity growth, remain puzzling (Levy, 2018).[1] The objective of this paper is to explore to what extent resource misallocation within sectors could lie at the heart of Mexico's low

ACKNOWLEDGMENTS Work was supported under the Fiscal Affairs Department's microdata initiative. 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 Microdata Laboratory. All results and information shown were compiled by the authors and are not part of INEGI's official statistics.

1. See also Santiago Levy and Dani Rodrik, "The Mexican Paradox," August 10, 2017 (www.project-syndicate.org/commentary/mexican-paradox-economic-orthodoxy-low-productivity-by-santiago-levy-and-dani-rodrik-2017-08?mod=article_inline&barrier=accesspaylog).

productivity, in line with existing papers.[2] First, we document substantial disparities in the level of resource misallocation, suggesting that an examination of aggregate levels obscures important variation that could provide clues about the underlying drivers. In a second step, we exploit subnational and industry variation to uncover specific and observable distortions that explain the inefficient allocation of resources in Mexico. Finally, we ask how much productivity levels could benefit from addressing these distortions. Our analysis follows Hsieh and Klenow (2009) in calculating resource misallocation, using exceptionally rich establishment-level data from the Mexican Economic Census, which comprises the universe of urban formal and informal firms with fixed establishments. These data—together with additional sources of aggregate subnational data—allow us to proxy for several distortions that are plausibly linked to resource misallocation.

A first look at the data suggests that the potential productivity gains from fully eliminating all observable and unobservable distortions that give rise to resource misallocation in Mexico—at 125 percent relative to actual total factor productivity (TFP)—are indeed large compared to other countries.[3] More important, 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 by Calligaris, Del Gatto, Hassan, and others (2016) for Italy—the only other paper that comparably examines resource misallocation at the subnational level. We also find that the subnational disparities in misallocation correlate very closely with state-level income per capita, even

2. For example, Busso, Fazio, and Levy (2012); IMF (2017); and Levy (2018). Busso, Madrigal, and Pagés (2013) examine resource misallocation in Latin America and find that the associated TFP losses in Mexico are substantially higher than in the rest of the Latin American countries they consider. IMF (2017) suggests that resource misallocation in Mexico is above the fiftieth or seventieth percentile in a sample of fifty-seven developing and emerging market economies, depending on the year considered. Relatedly, based on Keller's (2004) arguments, it appears doubtful that the lack of access to technology alone could explain low productivity growth in Mexico, insofar as Mexico has successfully opened its economy to international trade and investment since the mid-1990s.

3. These estimates are, however, more conservative than previous estimates for Mexico by Busso, Fazio, and Levy (2012). In the supplementary material, we compare the estimates of the aggregate TFP gains in the literature and test the robustness of our assumptions.

when we control for the composition of industries across states.[4] These results provide empirical support for the economic relevance of measuring resource misallocation through a model-based approach as proposed by Hsieh and Klenow (2009).

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 benefiting some firms at the expense of others, independently of their relative productivity levels.[5] The regressors are calculated using establishment-, 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 and with weaker access to financial and telecommunications services. Finally, we show that misallocation also increases when establishments are geographically distant from major population centers. To illustrate the economic significance of our results, the median Mexican state would see TFP rise by about 13 percent in a hypothetical reform scenario in which all distortions included in our baseline regression were 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).[6] Restuccia and Rogerson (2017) distinguish two broad approaches to quantifying 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

4. 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 from previous studies (for example, Dias, Richmond, and Robalo Marques, 2016).

5. See also Hanson (2010). For brevity, we use the terms firm and establishment interchangeably in the remainder of the paper, even though our data are at the establishment level rather than at the firm level, and we refer in the majority of cases to establishments when we use the term firm. While there is obviously a conceptual difference, the vast majority of firms in our data have only one establishment, so actual differences are relatively rare in our sample.

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

experiment. The indirect approach, 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.[7] While the direct approach has failed thus 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 because of measurement error (Restuccia and Rogerson, 2017). More recently, Haltiwanger, Kulick, and Syverson (2018) argue that Hsieh and Klenow's (2009) framework rests on strong assumptions that are often difficult to verify.

Our results contribute to the literature in two ways. First, we demonstrate that there are large disparities in resource misallocation at the state level. We confirm that the indirect approach to measuring resource misallocation delivers strong and economically sensible predictions at the macroeconomic level despite the often valid criticism of some of its underlying assumptions. In particular, we show that differences in per capita income are indeed closely correlated with differences in resource misallocation at the state level, which allows us to confirm Hsieh and Klenow's (2009) basic conjecture, namely, that resource misallocation matters for aggregate per capita income.[8] This result is consistent with the findings of earlier papers, including Restuccia and Rogerson (2008) and Restuccia (2019), but stands in contrast to the findings of Inklaar, Lashitew, and Timmer (2017).[9]

7. Several papers use 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.

8. The correlation remains high even after controlling for unobserved industry-level fixed effects that should, among other things, account for potential differences in industry composition (Dias, Richmond, and Robalo Marques, 2016).

9. Based on a simulation exercise, Restuccia (2019) shows that differences in resource misallocation can plausibly explain observed differences in per capita incomes. Inklaar, Lashitew, and Timmer (2017) 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 establishment- 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.

Our second contribution is to relate various theoretically motivated distortions to resource misallocation by exploiting variation across state-industry pairs, thereby combining the direct and indirect approaches to measuring resource misallocation to some extent. Our proxies for distortions are almost all observable at the sector-state level, so that we can use standard fixed effects regressions, except for measures of corruption, which vary only at the state level. To establish a link between the former and resource misallocation at the sector-state level, we employ a difference-in-differences approach similar to the one proposed by Rajan and Zingales (1998). The results imply that boosting targeted physical or transportation infrastructure investments, competition, and access to financial and telecommunications services, and strengthening the rule of law to root out corruption, crime, and labor informality, are associated with lower resource misallocation. These effects are both statistically and economically significant. This suggests that Mexico should continue to pursue the ambitious structural reform agenda implemented in recent years.[10]

The remainder of the paper is organized as follows. In the next section, we revisit the conceptual framework, followed by a description of the underlying data. We then report stylized facts on resource misallocation and provide the results from the econometric analysis. The final section concludes.

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 reflects the presence of distortions. While these distortions are not necessarily observable, at least not directly, Hsieh and Klenow's (2009) framework can be used to quantify all observable and unobservable distortions indirectly by measuring the potential TFP gains that would arise in their absence.

We apply the Hsieh-Klenow framework to the state level and assume that each industry j in state s consists of N_{js} monopolistically competitive firms

10. See Saborowski (2017) for details. The data used in our analysis precede many of the structural reforms implemented in recent years (for example, the telecommunications reform). The latter may already help to partially address the resource misallocation we observe.

and that each state consists of J_s industries.[11] 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 $\Sigma_{j=1}^{J_s} \theta_{js} = 1$. Total output in each industry j and state s is given by a constant elasticity of substitution production function:

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

where y_{ijs} denotes the output of firm i (which, for exposition purposes, we assume to have only a single establishment) in industry j in state s, and 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:

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

where firm i's price is p_{ijs} and w, δ, and r denote the wage, depreciation, and interest rates, respectively. The parameter τ_{ijs}^{y} represents a firm-specific wedge that distorts output decisions, and τ_{ijs}^{k} represents a firm-specific wedge that distorts the capital-to-labor ratio; taken together, they reflect all observable and unobservable distortions. Restuccia and Rogerson (2017) distinguish three categories of distortions, including statutory provisions, such as the tax code and regulations (for example, size-dependent taxation); discretionary provisions made by the government or other private institutions such as banks that favor or penalize specific firms (for example, selective enforcement of

11. We use the terms sector and industry interchangeably.

taxation or outright government corruption); and market imperfections (for example, barriers to entry and enforcement of property rights). The modeling framework is based on the assumptions that there are no adjustment costs, no input price heterogeneity across firms, and no heterogeneity in terms of the production technology across firms within sectors.

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

$$\text{(5)} \qquad \mathrm{MRPL}_{ijs} = \left(\frac{1-\alpha}{\mu}\right)\left(\frac{p_{ijs}y_{ijs}}{l_{ijs}}\right) = \left(\frac{1}{1-\tau^{y}_{ijs}}\right)w$$

and

$$\text{(6)} \qquad \mathrm{MRPK}_{ijs} = \left(\frac{\alpha}{\mu}\right)\left(\frac{p_{ijs}y_{ijs}}{k_{ijs}}\right) = \left(\frac{1+\tau^{k}_{ijs}}{1-\tau^{y}_{ijs}}\right)(r+\delta),$$

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}:

$$\text{(7)} \qquad \mathrm{TFPR}_{ijs} = p_{ijs}A_{ijs} = \left(\frac{p_{ijs}y_{ijs}}{k^{\alpha}_{ijs}l^{1-\alpha}_{ijs}}\right) = \mu\left(\frac{\mathrm{MRPK}_{ijs}}{\alpha}\right)^{\alpha}\left(\frac{\mathrm{MRPL}_{ijs}}{(1-\alpha)}\right)^{1-\alpha}.$$

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

$$\text{(8)} \qquad A_{is} = \frac{\left(p_{ijs}y_{ijs}\right)^{\frac{\sigma}{\sigma-1}}}{k^{\alpha}_{ijs}l^{1-\alpha}_{ijs}},$$

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

$$\text{(9)} \qquad \text{TFP}_{js} = \left[\sum_{i=1}^{N} \left(A_{ijs} \frac{\overline{\text{TFPR}}_{js}}{\text{TFPR}_{ijs}} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}},$$

where $\overline{\text{TFPR}}_{js}$ 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 so that

$$\text{(10)} \qquad \text{TFP}^{*}_{js} = \left[\sum_{i=1}^{N} \left(A_{ijs} \right)^{\sigma-1} \right]^{\frac{1}{\sigma-1}}.$$

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

$$\text{(11)} \qquad \left(\frac{Y_s}{Y^{*}_s} \right) = \prod_{j=1}^{J_s} \left[\sum_{i=1}^{N_{js}} \left(\frac{A_{ijs}}{\overline{A_{js}}} \frac{\overline{\text{TFPR}}_{js}}{\text{TFPR}_{ijs}} \right)^{\sigma-1} \right]^{\theta_{js}/(\sigma-1)}$$

and

$$\text{(12)} \qquad \text{TFPGAIN}_s = 100 \times \left(Y^{*}_s / Y_s - 1 \right).$$

The level of actual to efficient TFP, referred to as the TFP gain from eliminating resource misallocation in each sector and state, can analogously be written as

$$\text{(13)} \qquad = 100 \times \left(\text{TFP}^{*}_{js} / \text{TFP}_{js} - 1 \right).$$

We also calculate aggregate resource misallocation for Mexico's entire economy analogously by effectively treating the entire country as one state in the appendix. In the subsequent sections, we calculate the TFP gains based on equation 12 for each state to illustrate the variation of resource misallocation within Mexico. Using equation 13, we also calculate the TFP gains for each industry-state pair, which is the left-hand-side variable in our econometric analysis.

Data

The paper uses establishment-level data from the latest wave of the Mexican Economic Census. The Mexican National Institute of Statistics and Geography (INEGI) compiles the data set every five years, with the survey responses in the latest wave referring to the year 2013. The database contains around 3.5 million observations covering the universe of formal and informal non-agricultural 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.[12]

We compute resource misallocation at the four-digit level based on the 2002 North American Industry Classification System (NAICS) for the manufacturing and service sectors in each state. We exclude sectors in which productivity estimates could conceivably be misleading or difficult to compare with the remaining sectors, including financial services, construction, utilities, real estate, professional and technical services, and the management of shell companies. We also omit health, education, and arts and culture, in which an important share of firms 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 sector-state pairs with fewer than ten firms. We remove the 1 percent tails of the distribution of firm-specific output wedges, capital wedges, and total factor productivity

12. Previous rounds of the census have been used to compute resource misallocation in other studies such as Busso, Fazio, and Levy (2012), who focus their analysis on productivity differences between formal and informal firms.

(based on equations 5, 6, and 7, respectively). We end up with close to three million establishments and 3,139 industry-state pairs.

The output elasticities of labor and capital for each industry are approximated by the cost shares of broader sectors (at most at the two-digit level) in the United States from the U.S. 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 at 0.1, assuming real interest and depreciation rates of 5 percent, whereas we assume a uniform wage rate across firms; and we set the elasticity of substitution between the outputs of different firms at 3.0. 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. In contrast to other papers, we choose employment over the wage bill for our labor variable l in equation 3, because many firms in Mexico use unpaid labor (for example, family members), such 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 to compute aggregate TFP gains from eliminating resource misallocation for Mexico as a whole (see the online appendix).[13]

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 and the 2010 State and Municipal Database System (SIMBAD). We describe these and present summary statistics in the appendix.

Stylized Facts: 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—larger even than the variation found by previous studies at the cross-country level. State-level TFP gains range from around 80 to 190 percent, which is a broader range than found by Busso, Madrigal, and Pagés (2013), for example, for a sample of ten Latin American countries. Even the interquartile range, which omits potential outliers, still

13. Supplementary material for this paper is available online at http://economia.lacea.org/contents.htm.

TABLE 1. State-Level TFP Gains from Eliminating Resource Misallocation
Percent

Statistic	*Aggregate uncorrected gains*	*Aggregate gains, corrected for industry fixed effects*
Minimum	78.1	95.7
10th percentile	89.3	103.6
Median	116.5	126.1
Mean	123.3	127.1
90th percentile	162.8	148.7
Maximum	192.2	159.6
Standard deviation	29.0	16.7
Interquartile range (IQR)	73.5	45.1

Source: Authors' compilation, based on data from the 2013 Mexican Economic Census.
Note: The table shows the distribution of the state-level TFP gains in percent based on equation 12, if resources were allocated optimally across establishments within all sectors under consideration within each state.

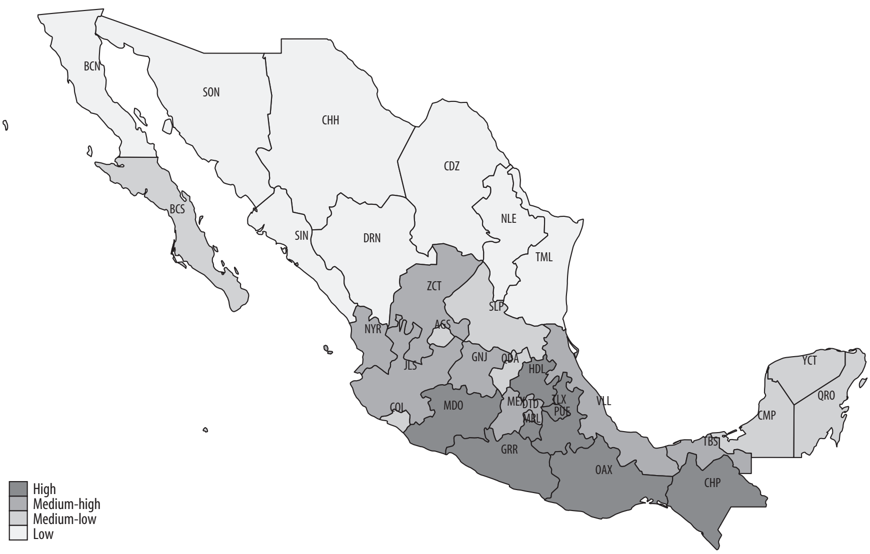
amounts to about 73.5 percentage points. This is more than the interquartile range of potential TFP gains in the manufacturing sector across advanced economies and amounts to two-thirds of the interquartile range for a large sample of developing countries reported by the IMF (2017). Table 1 contains the relevant summary statistics.

The variation in state-level TFP gains may be driven simply by differences in the industry composition of the economy of each state. For each industry-state pair, we therefore compute the level of TFP gains using equation 13 conditional on industry fixed effects, by running simple ordinary least squares (OLS) regressions and then recompiling aggregate state-level gains based on equation 12. In table 1, we also report the summary statistics for the state-level TFP gains corrected for industry-level fixed effects. The interquartile range of the TFP gains across states drops, but it is still large, reaching about 45 percent.

In figure 1, we distinguish four categories of the level of TFP gains by state. Broadly speaking, the level of TFP gains in northern states is low, whereas in southern states it is high. The level of TFP gains in central states and on the Yucatán Peninsula are somewhere between these extremes.

Given these geographic patterns, we proceed to evaluate whether resource misallocation would help explain income discrepancies 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 us to address issues related to measurement error and unobserved heterogeneity. Using establishment-level and national accounts data from a

FIGURE 1. Resource Misallocation by Mexican State



Source: Authors' compilation, based on data from the 2013 Mexican Economic Census.

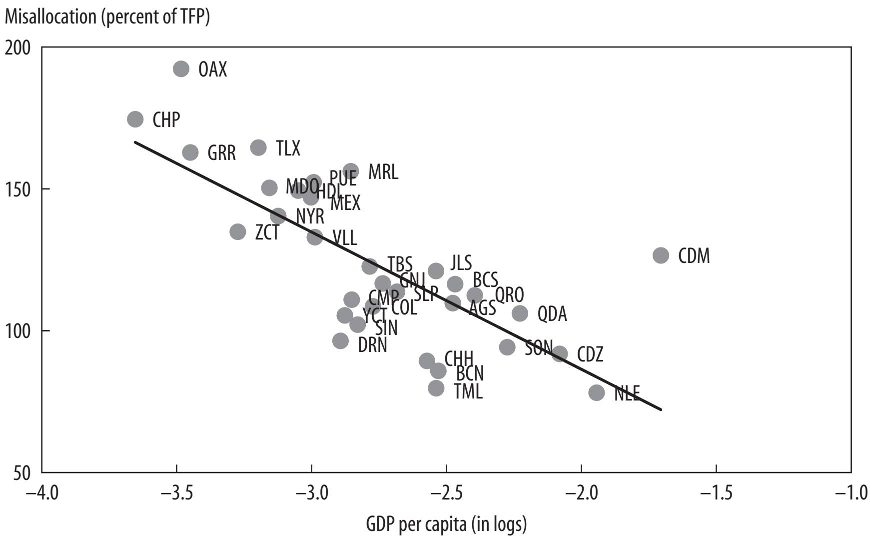
Note: The map shows the level of resource misallocation (expressed in quartiles) by state, which corresponds to the TFP gains if resources were allocated optimally across establishments within all sectors under consideration within each state. AOS, Aguascalientes; BCN, Baja California; BCS, Baja California Sur; CHH, Chihuahua; CHP, Chiapas; CMP, Campeche; COL, Colima; CDM, Ciudad de México; CDZ, Coahuila de Zaragoza; DRN, Durango; GNJ; Guanajuato; GRR, Guerrero, HDL, Hidalgo; JLS, Jalisco; MDL, Michoacán de Ocampo; ME, México; MRL, Morelos; NLE, Nuevo León; NYR, Nayarit; OAX, Oaxaca, PUE, Puebla; QDA, Querétaro; QRO, Quintana Roo; SIN, Sinoloa; SLP, San Luis Potosí; SON, Sonora; TBS, Tabasco; TML, Tamaulipas; TLX, Tlaxcala; VLL, Veracruz de Ignacio de La llave; YCT, Yucatán; ZCT, Zacatecas.

single source makes the data fully comparable across states. Most important, it allows us to exclude the same sectors from the national accounts data that we omitted from our establishment-level data in estimating resource misallocation, thus ensuring full consistency in the definition of the two measures we aim to correlate.

Figure 2 shows a scatter plot comprising all of Mexico's thirty-two states. After omitting Mexico City, which appears to be a clear outlier in the sense that its per capita income is higher than what one would expect based on its level of resource misallocation, we find that the correlation coefficient is a striking –0.84.[14] Using estimates of TFP gains that are corrected for

14. Dividing by the total population could bias our measure of per capita income if a large share of the population works 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, at –0.79.

FIGURE 2. State-Level per Capita Incomes and Resource Misallocation, 2013



Source: Authors' compilation, based on data from the 2013 Mexican Economic Census and the national account statistics.

Note: The figure shows the distribution of the state-level TFP gains (in percent) based on equation 12, if resources were allocated optimally across establishments within all sectors under consideration within each state.

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).

Econometric Results

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. To limit the effect of outliers on the results, we exclude observations with TFP gains from eliminating resource misallocation in the tenth and ninetieth percentiles of the distribution (based on equation 13). This leaves a total of 2,443 industry-state observations for the thirty-two Mexican states in our baseline regressions.

Our hypothesis is that there are observable proxies for regulation- and institution-related distortions in the Mexican economy, at both 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: (1) informality, given that the presence of a high share of informal firms plausibly implies that some firms enjoy unfair cost advantages, allowing them to attract more resources than they should according to their relative levels of productivity; (2) prevalence of crime, to capture the expectation that a high crime level would impose idiosyncratic costs on firms (for example, an establishment can be a victim of crime irrespective of whether or not it is relatively productive); (3) 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; (4) 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; and (5) geographic distance 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.[15] 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 variable that the fixed effects would not address is one whose impact on the dependent variable varied both across states and across industries (and also correlated 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 examples where this could be the case (for example, 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

15. 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.

rather than from productivity levels. As such, it does not appear straightforward to argue that the dependent variable would explain variation in our regressors.

We run simple regressions with heteroskedasticity-consistent standard errors.[16] The regression specification is given by

$$\mathrm{TFPGAIN}_{js} = \alpha_j + \alpha_s + \beta_1 \mathbf{X}_{js_{\mathrm{1stQuartile}}} + \beta_2 \mathbf{X}_{js_{\mathrm{2ndQuartile}}} + \beta_3 \mathbf{X}_{js_{\mathrm{3rdQuartile}}} + \varepsilon_{js},$$

where $\mathrm{TFPGAIN}_{js}$ refers to the TFP gain associated with eliminating resource misallocation in industry j and state s and is computed based on equation 13. State and industry fixed effects are given by α_s and α_j, respectively. The baseline regressions further include a vector of explanatory variables, $\mathbf{X}$, that contains our candidate distortions at the sector-state level (see the appendix for definitions, data sources, and summary statistics): (1) *Informality* is defined as the share of firms that did not make any social security or value added tax (VAT) payments in 2013; (2) *Crime* is defined as the share of firms located in high-crime municipalities (in which the number of robberies per capita is in the upper quartile of the distribution); (3) *No financial access* is defined as the share of firms without bank accounts; (4) *No internet use* is defined as the average share of employees who do not use the internet at work; and (5) *Distance* is defined as the average distance of firms in a given industry-state pair from the closest population center.[17]

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 dummy variables indicate whether an observation falls into the first, second, third, or fourth quartile of the underlying distribution of the continuous variable, where the first quartile has the least severe distortion and the fourth the most severe. For each distortion, we then include three of the four dummy variables in the regression, where the dummy variable pertaining to the fourth quartile is the omitted

16. The results are qualitatively robust to using standard errors that are clustered at the sector and state levels, as we report below.

17. Population centers are defined as cities with more than 500,000 inhabitants. As for the *Crime* variable, we exploit variation in the location of establishments across sectors within states, thereby ensuring that the effects of the *Distance* variable are not captured by the state fixed effects. We calculate only the "as the crow flies" distance, which could be misleading, especially in Mexico's mountainous center.

variable. The coefficient on each of the three included dummy variables 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 dummy variables to carry negative coefficients, with the most negative coefficient associated with the first quartile and the least negative coefficient associated with the third quartile.

The first regression in table 2 presents the results of our baseline specification when all fifteen dummy variables 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 a relatively gradual fall in the levels of 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 dummy variables 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.[18]

To examine the economic significance of the results, we conduct a simple policy experiment: we calculate the potential TFP gain for each Mexican state

18. 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.

TABLE 2. Baseline Regressions

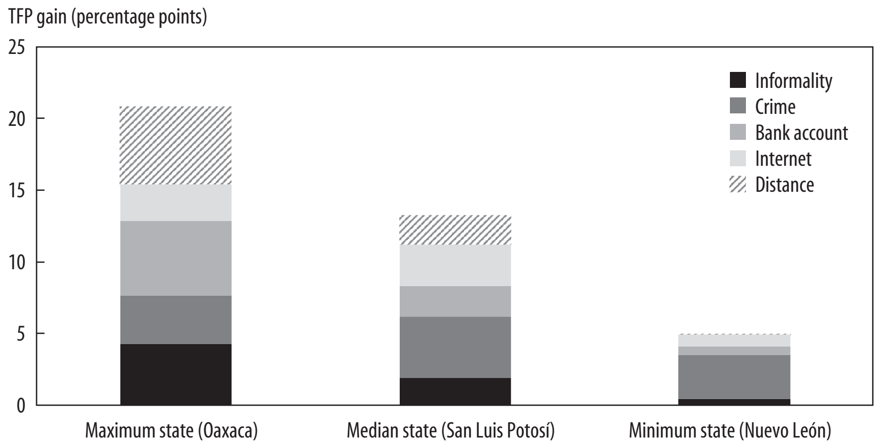
Explanatory variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*
Informality, 3rd quartile	−11.373***	−13.001***	−10.940***	−15.781***	−11.540***
	(3.008)	(2.981)	(2.985)	(2.521)	(2.995)
Informality, 2nd quartile	−14.647***	−17.054***	−14.319***	−23.878***	−14.386***
	(4.034)	(4.022)	(4.030)	(3.366)	(4.000)
Informality, 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)
Summary statistics					
No. observations	2,443	2,443	2,443	2,443	2,443
R^2	0.603	0.604	0.604	0.599	0.605
Alternative definition for		Informality	Crime	Financial	Internet

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Notes: Dependent variable: resource misallocation. All specifications include industry and state fixed effects. Standard errors are in parentheses.

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 we then calculate the implied impact on TFP. In some sense, we are thus estimating the gains

FIGURE 3. TFP Gains under Reform Scenario across States



Source: Authors' compilation, based on data from the 2013 Mexican Economic Census.

Notes: The figure shows the hypothetical state-level TFP gains in percentage points under a reform scenario that is based on the results of 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 we then calculate the implied impact on TFP. For illustrative purposes, we choose three states—Oaxaca, San Luis Potosí, and Nuevo León—where the level of misallocation corresponds to the sample maximum, median, and minimum, respectively.

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.

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 distortion 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).[19] 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 the *Distance* variable is excluded from these simulations—*Distance* may arguably not be directly responsive to reform initiatives—the

19. Figure A1 in the appendix shows the percentage point change in resource misallocation for all states.

impact still lies at around 15 percent of GDP.[20] 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 on the order of only 5 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.[21]

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 considered informal only if they do not make any social security payments, VAT payments, income tax payments, or excise tax payments.[22] All three informality dummy variables 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 with an alternative indicator of lack of financial access that measures the absence of bank credit. While the coefficient for the first quartile of the distribution is negative as expected, it is smaller than in the case of the *No financial access* variable, and the remaining two dummy variables 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 and that not all dimensions of financial development are equally important for resource misallocation. Finally, regression 5 replaces our indicator of *No internet use* with an indicator measuring the share of employees that do not use computers at work. Once again, the results are very similar to the baseline.

20. 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.

21. 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 0.3. The results suggest that the reform gain is 1 percentage point higher for every 15 percentage points more in initial resource misallocation.

22. Busso, Fazio, and Levy (2012) use a broader definition, considering firms to be informal if their social security payments fall short of the 18 percent of wages and salaries that should have been paid.

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 4 includes an indicator of the share of firms that reside in municipalities with a high population density. Intuitively, one might expect that a higher population density could lead to more competition and labor mobility. However, the variable coefficient is neither statistically significant nor of the expected (negative) sign. Similarly, we include the ratio of firms per capita in regression 5, but the hypothesis that the relative density of firms matters (due to more competition and labor mobility) is not supported by the data.

Regression 6 introduces a variable measuring the number of firms per industry-state pair to control for one important source of heterogeneity between sector-state pairs. In this case, both dummy variables 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 7 includes an indicator of the share of firms that have received foreign direct investment (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 8 includes municipal GDP per capita averaged across firms to control more precisely for differences in income levels (in addition to the state fixed effect included in the regressions). The two dummy variables turn out to be insignificant and the results broadly unchanged.

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 barriers to entry, one might 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, 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 driven not by barriers to entry but by 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 an 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 the firms' 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. To limit the number of interactions presented, we replace the three 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. Of note, however, the concentration term now carries the expected positive coefficient, while the interaction term has a negative coefficient. In other words, higher levels of concentration do appear to be associated with barriers to entry and resource misallocation for some industries, but the effect switches sign in industries with a high prevalence of informality.

TABLE 3. Robustness Checks

Explanatory variable	*(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 quartile	−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 quartile	−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 quartile	−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)			
No. firms per industry-state pair						−0.000		
						(0.000)		
FDI							−2.895	
							(5.626)	
GDP per capita								0.005
								(0.013)
Summary statistics								
No. observations	2,443	2,443	2,443	2,443	2,443	2,443	1,869	2,443
R^2	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

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Notes: Dependent variable: resource misallocation. All specifications include industry and state fixed effects. Standard errors are in parentheses.

TABLE 4. Market Concentration Regressions

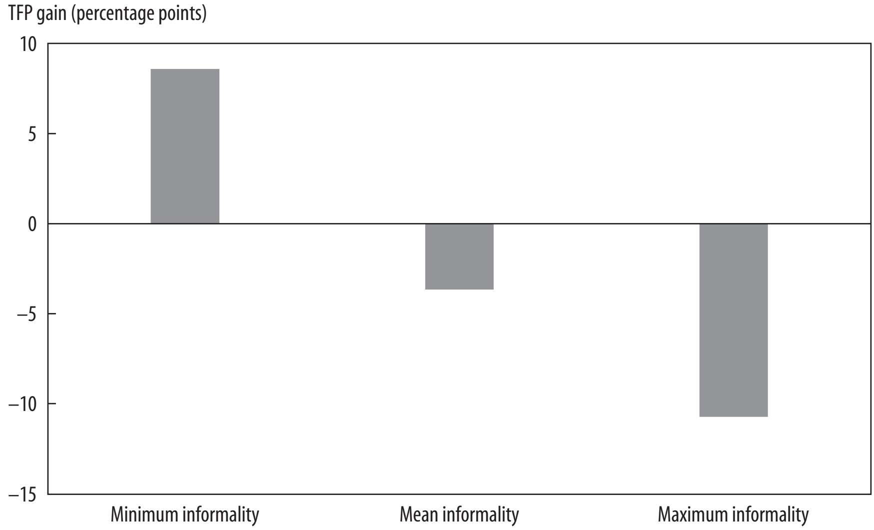
Explanatory variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Informality, 3rd quartile	−10.803***						
	(2.980)						
Informality, 2nd quartile	−13.994***						
	(4.015)						
Informality, 1st quartile	−15.633***						
	(5.170)						
Crime, 3rd quartile	−8.455***	−8.664***	−8.842***	−8.559***	−8.678***	−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 index 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)
Informality × Herfindahl index			−338.680***				
			(57.209)				
Herfindahl index by sales				−89.880***	49.020		
				(11.794)	(29.820)		
Informality × Herfindahl index					−220.676***		
					(43.047)		
Firm size						0.054**	0.140***
						(0.022)	(0.042)
Informality × Firm size							−0.618***
							(0.222)
Summary statistics							
No. observations	2,443	2,443	2,443	2,443	2,443	2,443	2,443
R^2	0.606	0.607	0.619	0.620	0.626	0.606	0.609

$^*p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: Dependent variable: resource misallocation. All specifications include industry and state fixed effects. Standard errors are in parentheses.

FIGURE 4. Effect of Market Concentration on TFP Losses, by Level of Informality



Notes: The figure shows the impact on TFP of a change in market concentration from the lowest to the average level in our sample. Informality is represented by three sector-state pairs from our sample: the pair with the lowest level of informality observed across all sector-state pairs, the average of all sector-state pairs, and pair the highest observed level.

In figure 4, we illustrate the results by simulating the impact of a change in market concentration from the lowest to the average level in our sample. We distinguish three sector-state pairs—one with a low level of informality (set to the minimum level of informality observed across all sector-state pairs), one with the mean level of informality, and one with a high level of informality (set to the maximum level of informality observed across all sector-state pairs). When informality is low, a higher concentration implies higher levels of TFP losses from resource misallocation. In contrast, when informality is high, increasing the level of concentration has beneficial effects and implies a decrease in the TFP losses from misallocation of around 10 percentage points.

These results are robust to using an alternative measure of market concentration based on sales rather than number of employees (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 to the regression. Finally, regressions 6 and 7 show a similar pattern 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.

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 National Government Quality and Impact Survey (ENCIG) to compute perception- and experience-based measures of corruption in public service provision with the purpose of linking them to resource misallocation (see table A1 in the appendix for definitions).[23] The survey collects information from respondents on their experience with and perception of procedures and services provided by different levels of government. 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.

We did not include our corruption indicators in the baseline specification because they vary across—but not within—states and would thus drop out in any regression incorporating state-level fixed effects.[24] 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 corruption by the extent to which government procurement is an important source of demand in the industry. We use 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, *Procurement*, 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

23. We thank Frederic Lambert for providing us with the indicators.

24. The number of survey respondents is too low to construct measures of corruption at higher levels of geographic disaggregation.

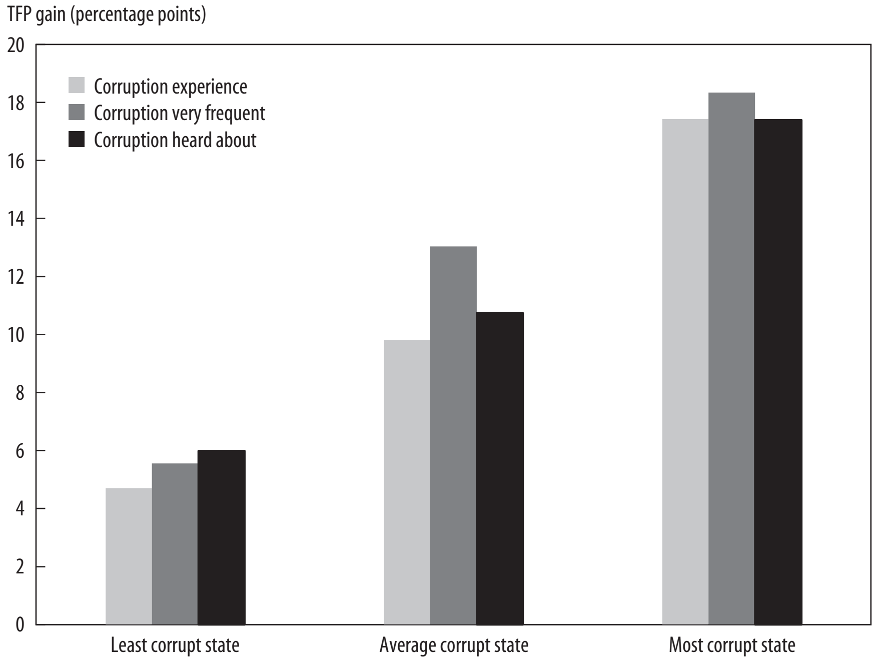
TABLE 5. Corruption Regressions

Explanatory variable	*(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)
Procurement × Corruption experience	1.050**			
	(0.469)			
Procurement × Corruption very frequent		0.295**		
		(0.144)		
Procurement × Corruption heard about			0.347*	
			(0.191)	
Procurement × Corruption top 3				0.033
				(0.279)
Summary statistics				
No. observations	2,443	2,443	2,443	2,443
R^2	0.605	0.605	0.605	0.604

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

Notes: Dependent variable: resource misallocation. See table A1 for definitions of the corruption variables. All specifications include industry and state fixed effects. Standard errors are in parentheses.

FIGURE 5. Effect of Corruption on TFP Losses in Government-Dependent Sectors, by Level of Corruption



Notes: The figure shows the impact of corruption on TFP, by measuring the difference in resource misallocation due to corruption in a sector-state pair that is fully dependent on government contracts or purchases versus a sector in the same state without any government dependence. Corruption is represented by three states from our sample: the state with the lowest observed level of corruption, the average of all states in the sample, and state the highest observed level.

fixed effects). The corruption indicator measures the share of respondents who have experienced corruption in their own dealings 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. Regressions 2, 3, and 4 provide additional confirmation for our hypothesis. Each one includes an alternative indicator of corruption in the interaction term, which remains significant in regressions 2 and 3 and continues to carry a positive coefficient in regression 4.

In figure 5, we further illustrate the effects of corruption using the three measures of corruption for which the interaction term is significant in the regressions. We distinguish three states—the one with the least corruption, one with an average level of corruption, and the one with the highest level of corruption based on these measures. Each bar shows the TFP losses from resource misallocation in a synthetic industry-state pair where the government is the primary client for all firms. In particular, a sector-state pair that is fully dependent on the government loses around 5 percentage points in TFP due to corruption relative to an industry without any government dependence when corruption is very low, and around 18 percentage points when corruption is very high. In the majority of sector-state pairs, the level of government procurement is substantially lower, implying that the TFP losses from corruption would also be smaller (see figure A1 in the appendix).

Conclusion

This paper argues that resource misallocation constrains productivity in Mexico. Our main contribution is to analyze the determinants of resource misallocation in Mexico across industries and states. We find that resource misallocation 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. In line with the findings of previous papers, this result shows the relevance of misallocation estimated through a model-based approach à la Hsieh and Klenow, although we cannot fully dismiss the criticism of the Hsieh-Klenow framework (2009). Another limitation of this approach is that it does not encompass all drivers of low growth and productivity in Mexico, such as resource misallocation across sectors, slow technological diffusion as described by Keller (2004), or slow within-firm growth. We examine the role of the latter in a companion paper (Misch and Saborowski, 2019).

The findings in our paper highlight the need for continued implementation of the structural reform program in Mexico (Saborowski, 2017). The results imply that boosting targeted infrastructure investments, competition, and access to financial and telecommunications services, and strengthening the rule of law to root out corruption, crime, and labor informality, are associated with lower resource misallocation. These effects are both statistically and economically significant. The link between the geographic isolation of some regions and resource misallocation points to the relevance 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.

Note: An earlier and shorter version of the paper is included in the Selected Issues Paper accompanying the 2017 IMF Article IV report for Mexico. The paper benefited from valuable guidance by Costas Christou and Era Dabla-Norris. We thank Suman Basu, Francesca Caselli, Laura Jaramillo, Frederic Lambert, Frederico Lima, Robert Rennhack, Alejandro Werner, participants in seminars at the International Monetary Fund, 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.

Appendix: Supplementary Tables and Figures

TABLE A1. 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 tax, or excise tax.	2013 Economic Census
Crime	Share of firms in high crime municipalities (in which the number of robberies per capita is in the upper quartile of the distribution).	2010 Sistema Estatal y Municipal de Base de Datos (SIMBAD); 2010 Population Census
—Alternative proxy	Share of firms in high crime municipalities (in which the number of homicides per capita is in the upper quartile of the distribution).	2010 Sistema Estatal y Municipal de Base de Datos (SIMBAD); 2010 Population Census
No bank account	Average across firms of a dummy variable that takes a value of one when firm has a bank account and zero otherwise.	2013 Economic Census
—Alternative proxy	Average across firms of a dummy variable that takes a value of one when firm has bank credit and zero otherwise.	2013 Economic Census

(*continued on next page*)

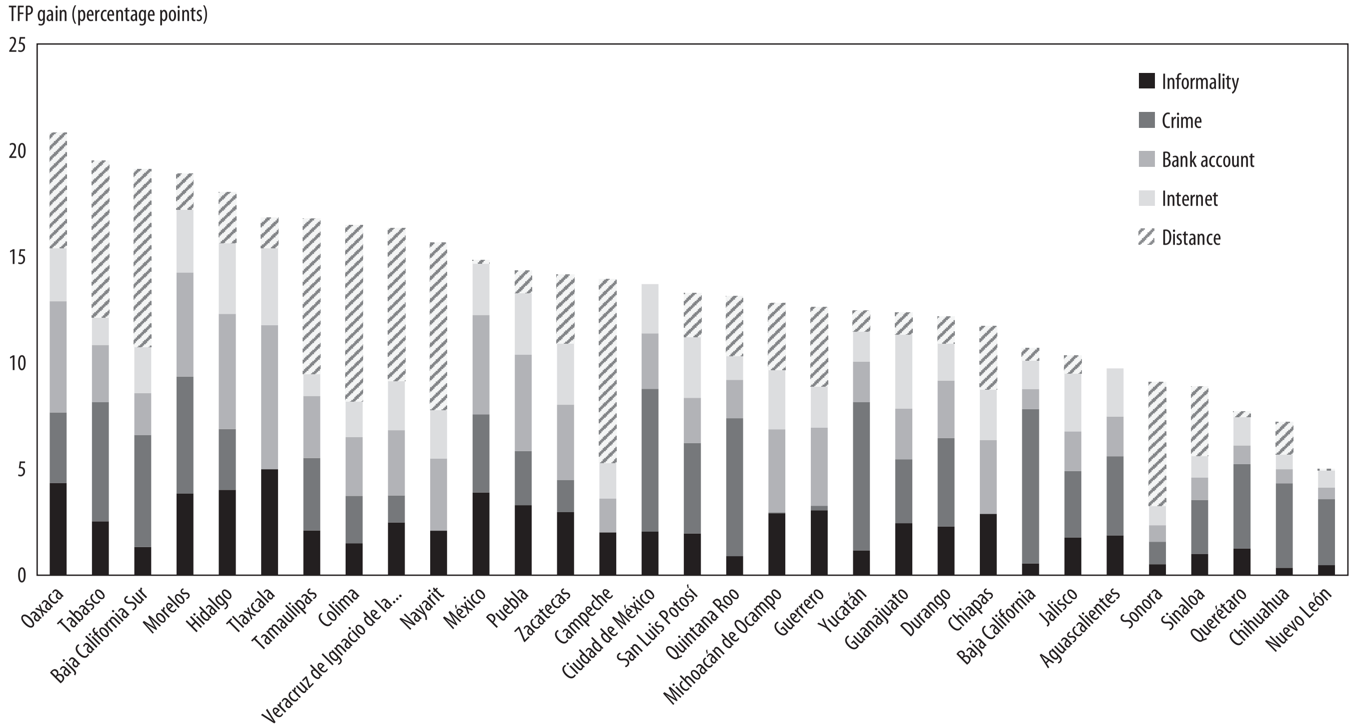
TABLE A1. Variable Definitions and Sources (*Continued*)

Variable	*Definition*	*Source*
No internet use	Average share of employees that do not use the internet for their work.	2013 Economic Census
—Alternative proxy	Average share of employees that do not use computers for their work.	2013 Economic Census
Distance	Average distance between the firm's location (e.g., city or town) and the closest population center (population > 500,000).	2013 Economic Census and authors' 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 have experienced corruption in dealing with the government.	2013 Encuesta Nacional de Calidad e Impacto Gubernamental (ENCIG)
Corruption very frequent	Proportion of respondents who answer that corruption is very frequent.	2013 Encuesta Nacional de Calidad e Impacto Gubernamental (ENCIG)
Corruption heard about	Proportion of respondents who have heard from relatives or friends that there are people who had to pay bribes.	2013 Encuesta Nacional de Calidad e Impacto Gubernamental (ENCIG)
Corruption top 3	Proportions of respondents who consider corruption one of the top three issues in the state government.	2013 Encuesta Nacional de Calidad e Impacto Gubernamental (ENCIG)
Other regressors		
High population	Share of firms in municipalities in the fourth quartile of population density.	2010 Sistema Estatal y Municipal de Base de Datos (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 the municipality in which the firms are located.	National account statistics from Instituto Nacional de Estadística y Geografía (INEGI)

TABLE A2. Summary Statistics

Variable	*Mean*	*Minimum*	*Maximum*
Baseline regressors			
Informality	0.634	0.000	1.000
—Alternative proxy	0.628	0.000	1.000
Crime	0.327	0.000	1.000
—Alternative proxy	0.334	0.000	1.000
No bank account	0.622	0.000	1.000
—Alternative proxy	0.815	0.000	1.000
No internet use	0.967	0.000	1.000
—Alternative proxy	0.942	0.000	1.000
Distance (in meters)	85,879.9	114.0	370,799.4
Regressors in extensions			
Herfindahl index by employment	0.057	0.000	0.859
Herfindahl index by sales	0.075	0.000	0.717
Firm size	11.950	1.244	1,000.164
Procurement	0.005	0.364	1.000
Corruption experience	9.345	4.484	16.596
Corruption very frequent	44.206	18.833	62.164
Corruption heard about	30.944	17.271	50.117
Corruption top 3	46.034	33.987	54.606
Other regressors			
High population	0.251	0.000	1.000
Firms per capita	0.043	0.023	0.130
Number of firms	880.9	10.0	127,602.0
FDI	0.072	0.000	1.000
GDP per capita	72.067	3.720	1,500.003

FIGURE A1. TFP Gains from Efficient Allocation of Resources under Reform Scenario across All States



Source: Authors' compilation, based on data from the 2013 Mexican Economic Census.

Notes: The figure shows the hypothetical state-level TFP gains in percentage points under a reform scenario that is based on the results of 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 we then calculate the implied impact on TFP.

References

Busso, Matías, María Victoria Fazio, and Santiago Levy. 2012. "(In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico." Working Paper 341. Washington, D.C.: Inter-American Development Bank.

Busso, Matías, Lucía Madrigal, and Carmen Pagés. 2013. "Productivity and Resource Misallocation in Latin America." *B.E. Journal of Macroeconomics* 13 (1): 903–32.

Calligaris, Sara, Massimo Del Gatto, Fadi Hassan, and others. 2016. "Italy's Productivity Conundrum: A Study on Resource Misallocation in Italy." European Economy Discussion Paper 030. Brussels: European Commission.

Dias, Daniel A., Christine J. Richmond, and Carlos Robalo Marques. 2016. "A Tale of Two Sectors; Why Is Misallocation Higher in Services Than in Manufacturing?" IMF Working Paper 16/220. Washington, D.C.: International Monetary Fund.

Haltiwanger, John, Robert Kulick, and Chad Syverson. 2018. "Misallocation Measures: The Distortion That Ate the Residual." Working Paper 24199. Cambridge, Mass.: National Bureau of Economic Research.

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

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

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124 (4): 1403–48.

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

Inklaar, Robert, Addisu A. Lashitew, and Marcel P. Timmer. 2017. "The Role of Resource Misallocation in Cross-Country Differences in Manufacturing Productivity." *Macroeconomic Dynamics* 21 (3): 733–56.

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

Levy, Santiago. 2018. "Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico." Washington, D.C.: Inter-American Development Bank.

Misch, Florian, and Christian Saborowski. 2019. "Firm Size, Life Cycle Dynamics, and Growth Constraints: Evidence from Mexico." Working Paper 19/87. Washington, D.C.: International Monetary Fund.

Rajan, Raghuram G., and Luigi Zingales. 1998. "Financial Dependence and Growth." *American Economic Review* 88 (3): 559–86.

Restuccia, Diego. 2019. "Misallocation and Aggregate Productivity across Time and Space." *Canadian Journal of Economics* 52 (1): 5–32.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." *Review of Economic Dynamics* 11(4): 707–20.

———. 2013. “Misallocation and Productivity.” *Review of Economic Dynamics* 16 (1): 1–10.

———. 2017. “The Causes and Costs of Misallocation.” *Journal of Economic Perspectives* 31 (3): 151–74.

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