What Drives Underdevelopment? *

Mark Huggett
Georgetown University
mh5@georgetown.edu

Wenlan Luo
Tsinghua University
luowenlan@gmail.com

Rodimiro Rodrigo
George Washington University
rodimiro.rodrigo@gwu.edu

THIS DRAFT: JANUARY 29 2026

Abstract
In Mexican census data, the mean labor wedge and mean labor tax wedge increase with plant size and the size gradient of the former equals or exceeds the gradient of the latter in most industries. These findings help to defend the claim that reallocating labor from small towards large plants produces marginal productivity gains and that size-dependent taxation is a driver of low productivity in Mexico.

Keywords: Development, Productivity, Misallocation, Labor Wedge, Manufacturing

JEL Classification: O1, O4, H2, E2

*Corresponding Author: Mark Huggett, Economics Department, Georgetown University, Washington, DC 20057, (email: mh5@georgetown.edu). We are grateful to Natalia Volkow and the Mexican Institute of Statistics and Geography (INEGI) for facilitating access to confidential microdata. All results have been reviewed to ensure that no confidential information is disclosed. We thank Miguel Messmacher and Santiago Levy for useful discussions, GCER for research support and Luis Felipe Perez Gomez for research assistance.

1 Introduction

What drives underdevelopment? One view is that a wide variety of policy-related distortions are key. In this view, underdevelopment is, in part, self-inflicted.

Levy (2018) "Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico" provides one articulation of this view. He argues that a combination of tax, labor contracting and enforcement policies imply that some large Mexican establishments face high marginal tax rates on labor costs and many small establishments face low tax rates. He then conjectures that alternative policies that reduce size-dependent taxation are likely to be important for increasing productivity.

This paper quantitatively assesses the Levy narrative. The first contribution constructs a labor tax wedge for each manufacturing plant in the Mexican census based on social security and value-added tax payments and on imputed income taxes and state payroll taxes. This wedge equals one plus the marginal tax rate on labor expenditures as a ratio to one minus the marginal tax rate on revenue. In theory a larger labor tax wedge reduces plant size, other things equal. The mean labor tax wedge robustly increases with plant size within industries for each year of the data and social security taxes are key to this finding.[1] This finding supports the first part of the Levy narrative: larger plants face larger labor tax wedges.

In theory, any dispersion in labor wedges (i.e. the ratio of price to the marginal cost of production using labor exclusive of taxes) is associated with production inefficiency so that a reallocation of labor inputs can increase output. Labor tax wedges are a source of variation in the labor wedge but other sources (e.g. differences in market power, extortion and regulation) are possible.[2] The mean labor wedge, inferred using theory and Mexican data, increases with size within the vast majority of industries. Importantly, the labor wedge size gradient typically equals or exceeds the size gradient of the labor tax wedge across the different methods we use to infer the labor wedge. This finding suggests that the tax system does not moderate misallocation but instead increases it by increasing the size gradient of the labor wedge.

The second contribution assesses the second part of the Levy narrative. First, it is argued that uniform percentage labor reallocations from small towards large plants increase productivity robustly within industries. This holds because in theory productivity increases when positive input reallocations occur on average at high wedge plants, for that input, and because the labor wedge has a positive size gradient. Quantitatively, a 10 percent reduction in labor input

[1] Levy argues that many small plants hire mostly non-salaried labor and pay no social security tax and the enforcement of social security taxes on salaried labor is concentrated at large plants.

[2] An account of extortion in tortilla manufacturing in Mexico is presented in "How Mexico's Cartels Infiltrated the Tortilla Business", Washington Post, May 23, 2024, by M. Sheridan, L. Rios and F. Ramos.

at small plants that is reallocated to large plants produces approximately a 1 percent median increase in marginal productivity across industries, based on using a Cobb-Douglas production function to infer wedges. It is useful to stress that labor wedges, not labor tax wedges, enter these calculations.

Second, a quantitative equilibrium model of the Mexican manufacturing sector is constructed. The model ties policy changes to equilibrium input reallocations, firm entry and exit and their productivity consequences. The model is calibrated to match gradients in labor and labor tax wedges as well as the fall in the ratio of labor to intermediate input expenditures with size, which is a robust feature of Mexican data. Production functions with both neutral and labor-augmenting technological heterogeneity across plants are key to match these facts. We analyze a policy reform that eliminates social security and state payroll taxes paid by firms and increases value-added taxes by a uniform percentage across firms to preserve revenue neutrality. These taxes are key sources of the positive size gradient of labor tax wedges. We find ... .

This paper is organized as follows. Section 2 discusses the related literature. Section 3 presents theoretical results on input wedges and productivity. Section 4 documents size variation in wedges and tax wedges based on Mexican census data and the theory in section 3. Section 5 analyzes productivity gains from input reallocations and from policy reforms that drive equilibrium reallocations. Section 6 discusses the main contributions.

2 Related Literature

Four literatures are most closely related to our two contributions.

Development Accounting

Many papers (Inter-American Development Bank (2010), Kehoe and Ruhl (2010), Hanson (2010) and Restuccia (2013)) calculate levels and growth rates of GDP per person and aggregate total factor productivity (TFP) in Latin American countries and in comparison countries. According to Restuccia (2013), the main source of low GDP per person in Latin American countries relative to the US is not differences in employment rates, hours per worker, capital-output ratios or measured education levels relative to the US. Instead, it is attributed to persistently low relative values of aggregate TFP. He concludes that the Latin American development problem is a TFP problem.[3] The Levy narrative suggests that low TFP in Mexico arises, in part, from size-dependent labor tax wedges.

[3]This conclusion is common in the development accounting literature (e.g. Hall and Jones (1999)).

Misallocation and Size Dependence

Syverson (2011), Hopenhayn (2014), Restuccia and Rogerson (2017), Cusolito and Maloney (2018) and Bergquist, Lashkari and Verhoogen (2026) review the misallocation literature. The early literature calculates productivity increases from the hypothetical elimination of size-dependent wedges.[4] Hsieh and Klenow (2009) analyze the hypothetical elimination of all wedges. Our work differs in several ways. First, we provide evidence on size dependence in tax wedges as well as evidence on size dependence of input wedges based on frameworks that move beyond the assumption of Cobb-Douglas production. Second, we analyze the magnitude of the marginal productivity gains from size-dependent input reallocations. These gains can be calculated (see Theorem 3) without making parametric assumptions on primitives.[5] Third, we analyze policy counterfactuals to quantitatively assess the Levy narrative.

Asker, Collard-Wexler and De Loecker (2014) document that the dispersion in the average revenue product of capital (ARPK) is greater in US industries with larger revenue TFP shocks. They argue that this fact is well explained by standard models with shocks impacting revenue TFP and with frictions (i.e. time-to-build or adjustment costs) to adjusting capital. They do not view ARPK dispersion as solely reflecting capital misallocation.[6] We focus on misallocation arising from variable inputs (e.g. labor and intermediate input) where such frictions seem less relevant.

The methods used (see Theorem 1) to infer that (mean) input wedges display size dependence are impacted either only in part or not at all by proportional measurement error with a size-invariant mean. Measurement error is a major issue in calculating allocative efficiency within an industry, comparing levels across countries and in determining trends over time. Hsieh and Klenow (2009), Rotemberg and White (2020), Blackwood, Foster, Grim, Haltiwanger and Wolf (2021) and Bils, Klenow and Ruane (2021) provide perspectives on this issue.

Markups and Production Function Estimation

DeLoecker and Warzynski (2012) provide theory and methods for estimating markups as the ratio of the output elasticity of a variable input to the expenditure on the variable input as a

[4]Guner, Ventura and Xu (2008) and Restuccia and Rogerson (2008) analyze the hypothetical elimination of size-dependent wedges. Garicano, Lelarge, and Van Reenen (2016) infer the magnitude of a size-dependent regulation cost to account for a bulge in the firm size distribution in France.

[5]Hsieh-Klenow productivity gain calculations use parametric assumptions (e.g. on production functions and aggregators) and treat capital as a variable input.

[6]David and Venkateswaran (2019) argue that much of the ARPK dispersion in data from China and the US arises from sources other than adjustment cost and informational frictions, such as a time-to-build friction. Kaymak and Schott (2025) analyze capital misallocation in US data based on firm-level corporate tax rates.

share of revenue. Input wedges (i.e. the ratio of price to the marginal cost of producing output with a given input exclusive of taxes) equal the ratio of output elasticities to expenditure shares.

We use several methods to estimate output elasticities in order to estimate wedges. One approach follows Raval (2023) and many others in using revenue data on output and two-step estimators for production functions, developed by Ackerberg, Caves and Frazer (2015), to calculate plant-specific output elasticities. DeRidder, Grassi and Morzenti (2024) apply these methods to estimate markups and output elasticities using both revenue and quantity measures of output. Based on French data, they conclude "... the use of revenue rather than quantity data to estimate production functions affects the level of the estimated markups, but has only modest effects on dispersion." Wedge dispersion is central to our assessment of the Levy narrative. We also employ cost share methods. A strength of cost share methods (see Proposition B.1 in Appendix B) for estimating output elasticities is that they are robust to plant heterogeneity in technology, factor prices and tax functions.

Misallocation and Mexican Data

Levy (2018) describes plant size and productivity distributions, legal forms of labor contracting, different types of firm-level taxes and provides a narrative for the productivity consequences of labor contracting, firm taxation and enforcement policies. Many papers use Mexican data and the Hsieh-Klenow methodology to calculate productivity gains from hypothetically eliminating all wedge dispersion.[7] Our work differs by providing evidence on wedge and tax wedge size dependence, among other differences. Documenting size dependence in any one tax rate does not document size dependence in any tax wedge as this property depends on multiple tax rates.[8] To the best of our knowledge, our work is unique is calculating tax wedges based on a fairly comprehensive account of taxes impacting firms.

Many papers analyze policy counterfactuals based on Mexican data: Sanchez-Vela and Valero-Gill (2011), Anton, Hernandez and Levy (2012), Leal-Ordonez (2014), and Alvarez and Ruane (2019). We differ by ... .

[7]See Inter-American Development Bank (2010), Busso, Madrigal and Pages (2013), Levy (2018), Misch and Saborowski (2018) and Orozco-Vazquez (2023) among others. Productivity gains typically are in the range of 100 to 200 percent based on Mexican data.

[8]Busso, Levy and Torres (2019) document that social security tax rates increase with plant size. Bachas, Fattal Jaef and Jensen (2019) document that tax enforcement and compliance increase with firm size in many countries including Mexico.

3 Wedges: Theory

This section (i) defines input wedges, (ii) shows how size dependence in input wedges can be inferred from observables, (iii) connects dispersion in input wedges to production inefficiency and (iv) provides a formula for the productivity gain from input reallocations that is stated in terms of wedges.

3.1 Infering Size Dependence in Wedges

Assume that plants within an industry differ in a state variable $v = (z_1, z_2, \tau) \in V$ and solve Problem P1 by choosing variable inputs $x = (x_1, ..., x_J)$ to maximize after-tax profits.

$$P1 \quad \max P(y, z_2, Y)y - \textstyle\sum_{j=1}^{J} w_j x_j - T(w_1 x_1, ..., w_J x_J, P(y, z_2, Y)y; \tau)$$

$$\text{s.t.} \quad y = z_1 F(x_1, ..., x_J)$$

In the analysis that follows (i) $P(y, z_2, Y)$ is the relationship between output, output quality, aggregate output (y, z_2, Y) and price, (ii) $y = z_1 F(x_1, ..., x_J)$ is a plant's production function, where z_1 is productivity, (iii) w_j is the price of input j and (iv) T is a tax function, where τ is a plant-specific parameter vector affecting taxes paid.[9]

$$P_1(y, z_2, Y)y + P(y, z_2, Y) = \tfrac{w_j}{z_1 F_j(x_1,...,x_J)}\left(\tfrac{1+T_j(w_1x_1,...,w_Jx_J,P(y,z_2,Y)y;\tau)}{1-T_{J+1}(w_1x_1,...,w_Jx_J,P(y,z_2,Y)y;\tau)}\right)$$

$$P(y, z_2, Y) = \tfrac{w_j}{z_1 F_j(x_1,...,x_J)} \underbrace{\frac{1+T_j}{1-T_{J+1}}}_{tax\ wedge_j} \underbrace{\frac{1}{1+\frac{P_1(y,z_2,Y)y}{P(y,z_2,Y)}}}_{markup\ wedge} = \tfrac{w_j}{z_1 F_j(x_1,...,x_J)} wedge_j$$

$$\ln wedge_j = \ln \tfrac{P(y,z_2,Y)y}{w_j x_j} + \ln z_1 F_j \tfrac{x_j}{y} = \ln \tfrac{P(y,z_2,Y)y}{w_j x_j} + \ln \alpha_j(x)$$

The first equation above is a first-order necessary condition for a solution to P1. It states that marginal revenue from an extra unit of output equals the marginal cost of producing an extra unit with variable input j after accounting for marginal tax rates on revenue and input expenditure. The marginal tax rate on input expenditure j and revenue are denoted $T_j = \partial T / \partial w_j x_j$ and $T_{J+1} = \partial T / \partial P(y, z_2, Y)y$, the marginal product of input j is $z_1 F_j = z_1 \partial F / \partial x_j$ and the derivative of price with respect to output y is $P_1(y, z_2, Y)$.

[9] Problem P1 may be part of the solution to a larger problem. The key assumption is that variable (flexible) input choices $x = (x_1, ..., x_J)$ are optimal, conditional on state variables and input prices. When capital k is preset, then the state variable is $v = (z_1, z_2, \tau, k) \in V$ and $y = z_1 F(x; k)$.

The second equation restates the first. It says that the price equals the product of the marginal cost of production exclusive of taxes and a wedge term. Thus, $wedge_j$ is the ratio of price to the marginal cost of production using input j exclusive of taxes. It can be decomposed as the product of a markup wedge, common across inputs, and a tax wedge specific to input j. The markup wedge is determined by the elasticity of demand, whereas $tax\ wedge_j$ equals one plus the marginal tax rate on input expenditure j as a ratio to one minus the marginal tax rate on output expenditure (i.e. total revenue).

The third equation restates the second. It is a foundation for inferring properties of input wedges that uses a straightforward empirical input, the revenue-expenditure ratio, and a less straightforward input, an output elasticity $\alpha_j(x) \equiv z_1 F_j \frac{x_j}{y}$ of input j. Two approaches to estimating $\alpha_j(x)$ are pursued. First, let F be a parametric and homogeneous function (e.g. Cobb-Douglas or translog) with neutral technology differences across plants, estimate the parameters of F, calculate $\alpha_j(x)$ and then $wedge_j$. This is an indirect way of measuring input wedges without measuring tax wedges, markup wedges or sources of wedge variation (e.g. extortion) not stated in P1. Second, let F differ nonparametrically across plants, estimate $\alpha_j(x)$ using a cost-share approach (see Proposition B.1 in Appendix B) with data on expenditures and marginal tax rates, and calculate $wedge_j$. This is a combination of direct and indirect measurement of wedges as marginal tax rates are used but no restrictions are put on markup wedges or production functions, beyond constant returns to scale and diferentiability. The approach uses cost minimization in an essential way and abstracts from wedge sources other than markup wedges and measured tax wedges.

Theorem 1 states assumptions so that differences in the mean and the variance of log wedges, conditioned on size, can be inferred from the differences in the mean and variance of log revenue-expenditure ratios. It assumes the output elasticity $\alpha_j(x) \equiv z_1 F_j \frac{x_j}{y}$ is a constant α_j for plants within an industry so that production is Cobb-Douglas: $y = z_1 F(x_1, ..., x_J) = z_1 \prod_{j=1}^{J} x_j^{\alpha_j}$. Assumption $A1$ restates the first-order condition, imposes constant output elasticities, and defines the measurement error ϵ_j. In $A1$, $\frac{Rev}{w_j x_j}$ and $\frac{\tilde{Rev}}{w_j x_j}$ denote true and measured values of the revenue-expenditure ratio. The revenue-expenditure ratio is stated as $\frac{Rev}{w_j x_j}$ rather than $\frac{P(y,z_2,Y)y}{w_j x_j}$. Assumption $A1$ implies that $(*)$ holds.

A1 $\ln wedge_j = \ln \frac{Rev}{w_j x_j} + \ln \alpha_j$ and $\ln \frac{\tilde{Rev}}{w_j x_j} = \ln \frac{Rev}{w_j x_j} + \epsilon_j$

A2 $E[\epsilon_j | s]$ is invariant to s

A3 $Var(\epsilon_j | s)$ and $Cov(\ln \frac{Rev}{w_j x_j}, \epsilon_j | s)$ are invariant to s

$$(*)\ \ln wedge_j = \ln \frac{\tilde{Rev}}{w_j x_j} + \ln \alpha_j - \epsilon_j$$

Theorem 1: *Assume A1. Then for all size values s, s_0 and any $j = 1, ..., J$*

(i) $E[\ln wedge_j|s] - E[\ln wedge_j|s_0] = E[\ln \frac{\tilde{Rev}}{w_j x_j}|s] - E[\ln \frac{\tilde{Rev}}{w_j x_j}|s_0]$ *provided A2 also holds.*

(ii) $Var(\ln wedge_j|s) - Var(\ln wedge_j|s_0) = Var(\ln \frac{\tilde{Rev}}{w_j x_j}|s) - Var(\ln \frac{\tilde{Rev}}{w_j x_j}|s_0)$ *provided A3 also holds.*

Proof:

(i) Take the conditional expectation of $(*)$ and calculate differences to produce the equation below. Assumption $A2$ produces the conclusion.

$$E[\ln wedge_j|s] - E[\ln wedge_j|s_0] = E[\ln \frac{\tilde{Rev}}{w_j x_j}|s] - E[\ln \frac{\tilde{Rev}}{w_j x_j}|s_0] - E[\epsilon_j|s] + E[\epsilon_j|s_0]$$

(ii) Take the conditional variance of $(*)$ to produce the first equation below. Apply $A1$ to produce the second. The conclusion follows by calculating differences and applying $A3$.

$$Var(\ln wedge_j|s) = Var(\ln \frac{\tilde{Rev}}{w_j x_j}|s) + Var(\epsilon_j|s) - 2Cov(\ln \frac{\tilde{Rev}}{w_j x_j}, \epsilon_j|s)$$

$$Var(\ln wedge_j|s) = Var(\ln \frac{\tilde{Rev}}{w_j x_j}|s) + Var(\epsilon_j|s) - 2[Cov(\ln \frac{Rev}{w_j x_j}, \epsilon_j|s) + Var(\epsilon_j|s)] \quad \| $$

3.2 Wedges and Production Efficiency

Production efficiency holds when it is not possible to reallocate inputs to increase output. To connect properties of wedges to production efficiency or inefficiency, theorize that a competitive bundler produces a final good using the differentiated output of all the plants. The bundler maximizes profit (i.e. output Y less input costs $\int_V p(v)y(v)d\mu$), taking input prices $p(v)$ as given. Input prices are stated in units of the bundler's output. The bundler's output Y satisfies the production relationship $\int_V z_2 \mathcal{Y}(y(v)/Y)d\mu = 1$, where $\mathcal{Y}$ is a Kimball (1995) aggregator, $v \in V$ is a state variable and μ is a measure. The measure μ describes the mass of plants with states in different subsets of the state space V. Profit maximization implies that the first-order condition $p(v) = z_2 \mathcal{Y}'(\frac{y(v)}{Y}) / \int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y}) \frac{y(v)}{Y} d\mu$ holds and generates the demand relationship $P(y, z_2, Y)$ perceived by the monopolistic firm in problem $P1$.[10]

$$P(y, z_2, Y) = z_2 \mathcal{Y}'(\tfrac{y}{Y}) / \textstyle\int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y}) \frac{y(v)}{Y} d\mu$$

[10]Profit maximization implies cost minimization. The Lagrangean for the cost minimization problem is $\mathcal{L}(y, Y) = \int_V p(v)y(v)d\mu + \lambda(1 - \int_E z_2 \mathcal{Y}(y(v)/Y)d\mu)$. Apply Luenberger (1969, Thm. 2, Ch. 7.7) to argue that at a minimum, there is a Lagrange multiplier λ so that the Gateaux derivative $\delta\mathcal{L}(y, Y; h)$ in direction h satisfies $\delta\mathcal{L}(y, Y; h) = \int_V [p(v) - \frac{\lambda}{Y} z_2 \mathcal{Y}'(\frac{y(v)}{Y})]h(v)d\mu = 0$ for all directions h. This condition and zero profits produces the condition in the text. A Kimball aggregator, where $\mathcal{Y}(1) = 1, \mathcal{Y}'(x) > 0, \mathcal{Y}''(x) < 0$, implies a downward sloping demand curve.

Definition: *$(x(v), y(v), Y)$ is a feasible allocation, given $(F, \mathcal{Y}, \mu)$, provided (i) $y(v) = z_1 F(x(v))$, where $x(v) = (x_1(v), ..., x_J(v))$, (ii) $\int_V x_j(v) d\mu = 1$ for $j = 1, ..., J-1$, (iii) $\int_V z_2 \mathcal{Y}(y(v)/Y) d\mu = 1$ and (iv) $Y - \int_V x_J(v) d\mu > 0$.*

With this formulation of the production structure of the economy, input $j < J$ is a primary input available in fixed supply normalized to equal 1 and input J is an intermediate input, where part of the total production Y of the output good is used as intermediate input.

Theorem 2 establishes two points. First, production efficiency implies that there is no dispersion in the marginal product of input j for the final output good when the marginal input is channeled to any plant and the plant's marginal output is then channeled to the bundler. Second, there can be no dispersion in a wedge for any variable input if production efficiency is to hold within this market structure.

Theorem 2: *Assume $(x(v), y(v), Y)$ is a feasible allocation.*

(i) If $(x(v), y(v), Y)$ is production efficient, then $z_2 \mathcal{Y}'(y(v)/Y) z_1 F_j(x(v))$ is constant for each $j = 1, ..., J$.

(ii) If $(x(v), y(v), Y)$ is production efficient and consistent with solutions to the plants' and the bundler's problem, then there is no dispersion in $wedge_j$ for $j = 1, ..., J$.

Proof:

(i) Let $H(Y, \alpha; x, \bar{x}) \equiv \int_V z_2 \mathcal{Y}(\frac{z_1 F(x(v) + \alpha \bar{x}(v))}{Y}) d\mu - 1$, where $\bar{x}$ is a zero-mean input reallocation (i.e. $\int_V \bar{x}(v) d\mu = \vec{0}$). By feasibility, $H(Y, 0; x, \bar{x}) = 0$. If $(x(v), y(v), Y)$ is production efficient, then there is no α and no $\bar{x}$ that increases the bundler's output. The Implicit Function Theorem and production efficiency imply $\frac{dY}{d\alpha}(0) = -\frac{H_2(Y,0;x,\bar{x})}{H_1(Y,0;x,\bar{x})} = \frac{\int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y}) \frac{\sum_j z_1 F_j(x(v)) \bar{x}_j(v)}{Y} d\mu}{\int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y}) \frac{z_1 F(x(v))}{Y^2} d\mu} = 0$ for all zero-mean input reallocations $\bar{x}$. For this to hold $z_2 \mathcal{Y}'(y(v)/Y) z_1 F_j(x(v))$ must be constant for each $j = 1, ..., J$.

(ii) Combine the plants' and the bundler's first-order condition to produce the system of equalities below. Thus, there is a price function solving the bundler's problem, common factor prices and wedge terms that satisfy all the plants' first-order conditions. This result and Theorem 2(i) imply that achieving production efficiency, within this market structure, requires that $wedge_j$ is equal across plants.

$$\frac{z_2 \mathcal{Y}'(\frac{y(v)}{Y})}{\int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y}) \frac{y(v)}{Y} d\mu} = P(y(v), z_2, Y) = \frac{w_j}{z_1 F_j(x(v))} wedge_j(v)$$

$$wedge_j(v) = \frac{1}{w_j} \frac{z_2 \mathcal{Y}'(\frac{y(v)}{Y}) z_1 F_j(x(v))}{\int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y}) \frac{y(v)}{Y} d\mu} \quad \| $$

3.3 Wedges and Productivity Gains

Using system $(**)$, define the productivity gain $\Omega(\lambda)$ of moving a fraction $\lambda \in [0,1]$ of the way from feasible allocation $(x(v), y(v), Y)$ towards feasible allocation $(x^*(v), y^*(v), Y^*)$.

$$(**) \quad 1 + \Omega(\lambda) = \frac{Y(\lambda) - \int_V (x_J(v) + \lambda(x_J^*(v) - x_J(v)))d\mu}{Y - \int_V x_J(v)d\mu}$$

$$\int_V z_2 \mathcal{Y}(\frac{z_1 F(x(v) + \lambda(x^*(v) - x(v)))}{Y(\lambda)})d\mu = 1$$

Theorem 3 provides formulae for the productivity gain $\Omega(\lambda)$ and the marginal gain of moving towards an alternative allocation. The marginal gain $\Omega'(0)$ is the weighted sum of the reallocation gains from adjusting each input. Weights are the ratio of industry input costs to industry final output. The reallocation gain, for a given input, is stated in terms of input wedges, input weights and inputs reallocated, where $\hat{E}_j$ is the input weighted expectations operator for input j. Marginal gains are zero for each primary input absent wedge dispersion, consistent with Theorem 2, and are positive when input is shifted towards higher wedge plants.

Theorem 3: *Assume $(x(v), y(v), Y)$ and $(x^*(v), y^*(v), Y^*)$ are feasible, where $(x(v), y(v), Y)$ is consistent with solutions to the plants' and the bundler's problem. Then $\Omega(\lambda) = \int_0^\lambda \Omega'(\bar{\lambda})d\bar{\lambda}$ and $\Omega'(0) = \sum_{j=1}^J \omega_j \hat{E}_j[wedge_j \frac{(x_j^* - x_j)}{x_j}] - \frac{\int_V (x_J^* - x_J)d\mu}{[Y - \int_V x_J d\mu]}$, where $\omega_j = \frac{\int_V w_j x_j d\mu}{Y - \int_V x_J d\mu}$ is the share of final output paid for input j.*

Proof: $\Omega(\lambda) - \Omega(0) = \int_0^\lambda \Omega'(\bar{\lambda})d\bar{\lambda}$ follows from the Fundamental Theorem of Calculus. The result follows as $\Omega(0) = 0$. Define $x(v;\lambda) = x(v) + \lambda(x^*(v) - x(v))$ and $y(v;\lambda) = z_1 F(x(v;\lambda))$. Feasibility implies

$$\Omega'(\lambda) = \frac{Y(\lambda) \int_V \frac{z_2 \mathcal{Y}'(\frac{y(v;\lambda)}{Y(\lambda)})}{\int_V z_2 \mathcal{Y}'(\frac{y(v;\lambda)}{Y(\lambda)}) y(v;\lambda) d\mu} \sum_{j=1}^J z_1 F_j(x(v;\lambda))(x_j^*(v) - x_j(v))d\mu - \int_E (x_J^*(v) - x_J(v))d\mu}{Y - \int_V x_J(v)d\mu}$$

The first and second equation below restate $\Omega'(0)$ using $P(y, z_2, Y) = z_2 \mathcal{Y}'(\frac{y}{Y}) / \int_V z_2 \mathcal{Y}'(\frac{y(v)}{Y})\frac{y(v)}{Y}d\mu$ and $P(y(v), z_2, Y) z_1 F_j(x(v)) = w_j wedge_j(v)$, based on first-order conditions for the bundler and firm. The remaining equations reorganize this result. Towards this end, let $\omega_j = \int_V w_j x_j d\mu / [Y - \int_V x_J d\mu]$, $E[\cdot]$ denote the expectation operator based on the probability measure $\mu(\cdot)/\mu(V)$ and $\hat{E}_j[\cdot]$ denote the input-weighted expectation operator for input j.

$$\Omega'(0) = \sum_j \frac{\int_V P(y, z_2, Y) z_1 F_j(x)(x_j^* - x_j)d\mu}{Y - \int_V x_J d\mu} - \frac{\int_V (x_J^* - x_J)d\mu}{Y - \int_V x_J d\mu}$$

$$\Omega'(0) = \sum_j \frac{\int_V w_j wedge_j (x_j^* - x_j) d\mu}{Y - \int_V x_J d\mu} - \frac{\int_V (x_J^* - x_J) d\mu}{Y - \int_V x_J d\mu}$$

$$\Omega'(0) = \sum_j (\frac{\int_V w_j x_j d\mu}{Y - \int_V x_J d\mu}) \int_V \frac{w_j x_j}{\int_V w_j x_j d\mu} wedge_j \frac{(x_j^* - x_j)}{x_j} d\mu - \frac{\int_V (x_J^* - x_J) d\mu}{Y - \int_V x_J d\mu}$$

$$\Omega'(0) = \sum_j \omega_j E[\frac{x_j}{E[x_j]} wedge_j \frac{(x_j^* - x_j)}{x_j}] - \frac{\int_V (x_J^* - x_J) d\mu}{Y - \int_V x_J d\mu}$$

$$\Omega'(0) = \sum_j \omega_j \hat{E}_j[wedge_j \frac{(x_j^* - x_j)}{x_j}] - \frac{\int_V (x_J^* - x_J) d\mu}{Y - \int_V x_J d\mu} \quad \| $$

Comment: The formula holds without making parametric assumptions on the primitives $(F, \mathcal{Y}, \mu)$. It is applied to data in section 5. One application is to small reallocations so that the marginal gain $\Omega'(0)$ may approximate the overall productivity gain or loss $\Omega(1)$. When reallocations are not small the formula might still be useful. For example, Proposition C.1 in Appendix C shows that it serves as an upper bound (i.e. $\Omega'(0) \geq \Omega(1)$) when it can be established that $\Omega''(\lambda) \leq 0$. Proposition C.3 in Appendix C shows that the formula for an economy with $S \geq 1$ industries is $\Omega'(0) = \sum_{s=1}^{S} \sum_{j=1}^{J} \omega_{js} \hat{E}_{js}[wedge_{js} \frac{(x_{js}^* - x_{js})}{x_{js}}] - \frac{\sum_s \int_V (x_{Js}^* - x_{Js}) d\mu_s}{Y - \sum_s \int_V x_{Js} d\mu_s}$. The framework employs a bundler for each industry s, a bundler for the final output good and monopolistic firms producing differentiated goods within each industry.

4 Wedges: Empirics

This section describes (i) sample selection, (ii) employment structure, (iii) average tax rates and (iv) properties of wedges and tax wedges.

4.1 Sample Selection

We analyze Mexican census data for the census years 1999, 2004, 2009, 2014 and 2019. Census years are by convention the year after the measurement year. The data include in principle all establishments with a fixed premises within a sufficiently large urban area. Our focus is on establishments (i.e. plants) in the manufacturing sector.

Sample 1 contains all manufacturing plants that satisfy some basic positivity restrictions and are in six-digit industries with at least 400 plants. Sample 2 is used to calculate tax wedges. Sample 2 eliminates all plants in census year 1999 and 2004 as value-added tax information is

Table 1: Sample Selection and Sample Properties

Description	1999	2004	2009	2014	2019
Manufacturing Establishments	342711	328718	436851	489530	579828
Sample 1	200372	196151	271802	276374	321240
Sample 2: Sample 1 and tax data	0	0	271802	276374	321240
Sample 1 Properties					
Firms w/ 1 Establishment (%)			99.9	99.8	99.8
Number 6-Digit Industries	75	64	73	74	78
Median Estab. per Industry	810	894	891	893	1040
Median Estab. Size (Employment)	3	3	3	3	3
Mean Estab. Size (Employment)	10	9	8	8	9
Median Estab. Size (Revenue)	168	218	192	241	360

Note: Sample 1 eliminates plants within six-digit industries when (1) gross output, intermediate input or value added is not positive, (2) total workers are less than two and (3) there are less than 400 plants within an industry. Sample 2 is Sample 1 after excluding entire years in which value added tax data is missing or not validated. Revenue is in units of thousands of Mexican pesos.

either missing or not validated.[11] Table 1 shows that there are roughly 70 six-digit industries in Sample 1 across census years and that more than 99 percent of firms have only one plant.[12] This last fact leads us to equate a plant with a firm when convenient. The plant size distribution could be described as consisting of many minnows and a small number of muskies (i.e. plants with 100 or more employees, roughly the 99th percentile).

4.2 Employment Structure

In Mexican census data, workers are put into three categories based on classes of worker contracts: (1) Owners, Family and Other Non-remunerated Workers, (2) Outsourced and Commisioned Workers and (3) Salaried Workers. Workers in Mexico can be hired on a salaried or a non-salaried basis. Levy (2018, Ch. 2) states that a *"salaried worker is a subordinated employee performing the tasks set by the boss/entrepreneur in the location and at the time of the boss's choosing"*. Salaried workers are entitled to social security benefits and the firm is responsible for making social security payments for such workers.[13] Non-salaried workers, listed in Figure 1

[11] In 1999, the tax data are missing. In 2004, INEGI started asking firms about value-added taxes paid or charged. It was not until 2009 that these tax variables were validated.

[12] Starting in 2009, the census has a firm identifier and an establishment identifier, which allows a calculation of firms with one establishment.

[13] Levy (2008, Table 3.7) and Levy (2018, Ch. 2) summarizes the bundled social security benefits (retirement, health, disability, housing loans, daycare, recreation centers and firing protections) associated with salaried

as Owners, Family and Other Workers, are not subordinated employees, at least in legal terms if not in reality. They receive some social benefits, but the firm does not pay social security taxes for these workers.

Figure 1 shows the distribution of worker contracts within size bins based on total workers at a plant. Small plants employ a much higher fraction of non-salaried workers than larger plants. Specifically, they employ a large fraction of workers in the category "owners, family and other non-remunerated workers" for which the plant does not pay social security taxes.[14]

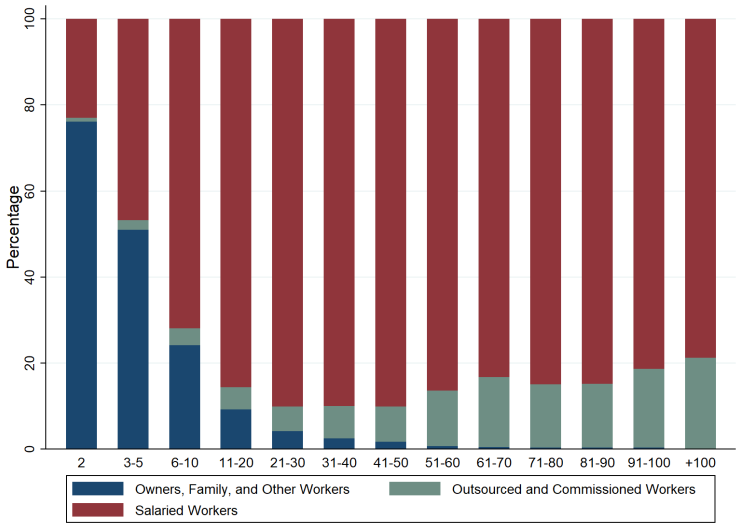


Figure 1: Employment Structure by Plant Size, 2019

4.3 Average Tax Rates

This section documents properties of tax rates for social security, value added, income taxes and state payroll taxes. A key feature of firm taxation in many Latin American countries is that small firms face a simplified income tax regime. In Mexico, this is called the Régimen de Pequeños Contribuyentes (Repeco) for census years 1999, 2004, 2009 and 2014 and the Régimen de Incorporacion Fiscal (RIF) for 2019. The vast majority of manufacturing firms in Mexico are Repeco and RIF firms. Under Repeco, firms with revenue below 2 million Mexican pesos pay a tax based on revenues that is administered at the state level. Table 1 shows that a plant with median revenue is well below the Repeco threshold in all census years.

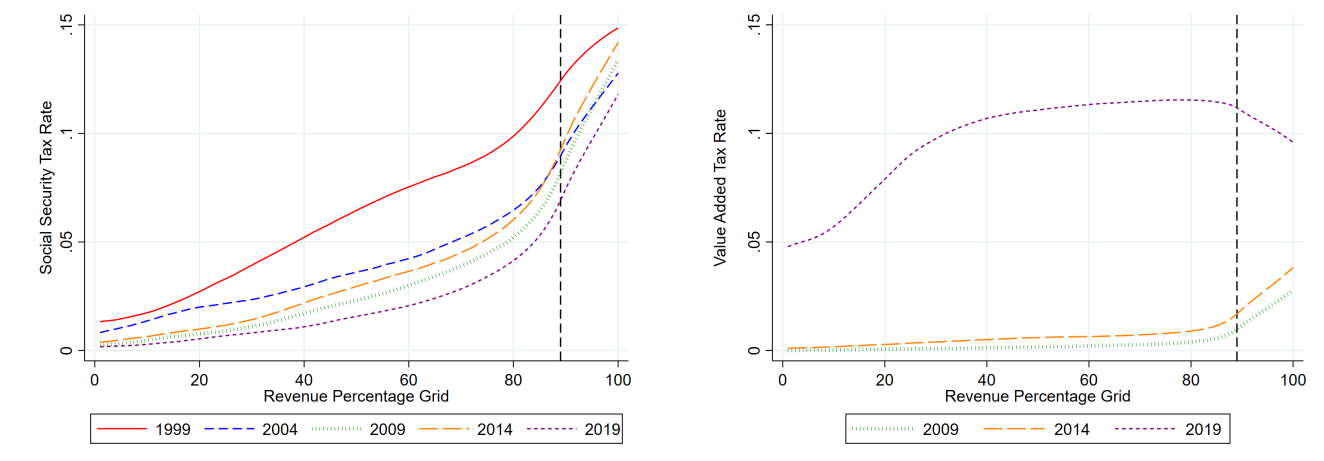
work.

[14]This category of workers receives labor market compensation, despite the direct translation of the spanish phrase describing this category, although the compensation is not recorded in census data.

Under RIF, firms with revenue below the same threshold pay income tax on profits under the general regime schedule but are entitled to a percentage reduction in these taxes (falling from 100 percent to 0 percent reduction after 10 years within RIF). RIF firms are exempt from paying value-added tax (VAT) if their annual gross income is less than 300 thousand Mexican pesos, otherwise they would pay VAT under the normal rules. RIF firms receive a reduction in social security taxes of 50 percent during the first two years, tapering to 10 percent in year 10. Appendix A provides a detailed discussion of firm taxation in Mexico.

4.3.1 Social Security and Value Added Tax Rates

Figure 2 plots average social security tax rates in each year as a function of a measure of plant size, measured by revenue. The social security tax rate for a plant within a revenue size bin is only calculated if the plant makes payments to salaried workers. Thus, it is important to stress that the tax rates are averages among the plants within a size bin that make payments to salaried workers. Across all census years, the median plant social security tax payment is zero. Thus, the size variation is more substantial as small plants are primarily those without salaried workers and those that pay zero social security taxes.



(a) Social Security (b) Value Added

Note: The social security tax rate is calculated as the ratio of total social security taxes to wages and salaries of salaried workers. The value added tax rate is calculated as the total value added tax charged divided by total revenue. Each point represents the average tax rate for plants within a specific size bin, measured by revenue. The vertical line marks the percentage grid for the 2-million-peso revenue threshold in 2019.

Figure 2: Average Social Security and Value Added Tax Rate by Plant Size

Figure 2 shows that average social security tax rates generally increase with size in all census years. Plants that are in the largest size bin in a given census year pay an average social security tax rate of between 10 and 15 percent. There is nothing obvious in the rules for the payment of

social security taxes (see Table A.1 in Appendix A) that would produce the increasing pattern of tax rates for the years 1999 - 2014. The size profile in 2019 is below the profiles in earlier census years. This is consistent with the structure of RIF whereby firms with revenue below the threshold are granted percentage reductions in social security taxes paid. Figure 2 plots a vertical line at the Repeco threshold. Roughly 90 percent of the plants in our sample are below the Repeco-RIF threshold in 2019.

Figure 2 plots average value-added tax rates on sales as a function of a measure of plant size. Value-added tax rates typically rise with plant size bin, although the pattern in 2019 is quite different. Before 2019, less than 10 percent of the plants in our sample charge positive value-added taxes, whereas in 2019 approximately 68 percent of the plants in our sample charge value-added taxes. Thus, a narrative is that enforcement reforms applying after the 2014 census year substantially impacted the fraction of plants charging value-added taxes. This is consistent with the fact that average value-added tax rates on intermediate inputs and on output share the same qualitative properties across years, see Figure B.2 in Appendix B, and the fact that the statutory value-added tax rate hardly changed. It was 15 percent up until the 2009 tax year and was constant at 16 percent afterwards.[15]

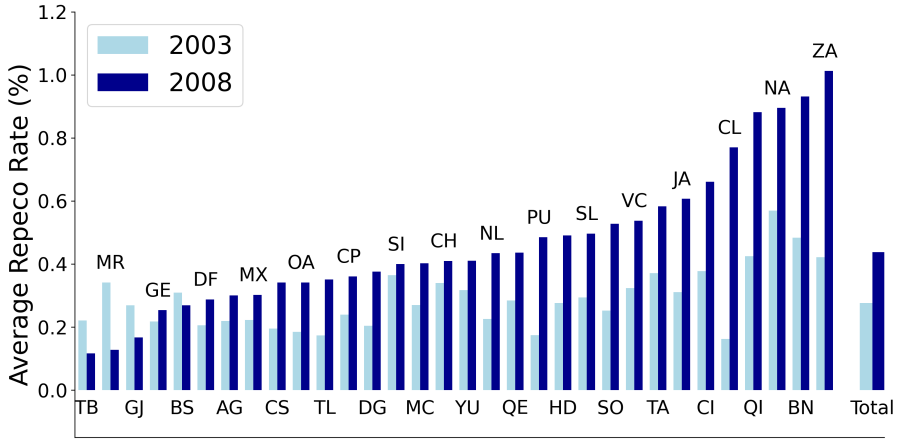
4.3.2 Repeco Tax, Profit Tax and Payroll Tax Rates

Figure 3 plots average Repeco tax rates by state. Individual states were in charge of setting and collecting Repeco revenue taxes. Average Repeco tax rates on revenue are approximately 1 percent or less in any state in any year.

Firms that are beyond the Repeco revenue threshold pay income taxes under the general regime. General regime income taxes are flat rate taxes based on firm profits. Across the census years, the flat rate on profits varied from a low of 28 percent to a high of 34 percent - see Appendix A. Once a firm crosses the Repeco threshold, it is taxed under the general regime thereafter, regardless of the value of future revenue.

Payroll taxes are determined, collected and administered by each state treasury office. The payment of this tax by firms is based on the sum of all salary paid to salaried workers. Appendix A documents the tax rates charged by each state. In the census years 2009, 2014 and 2019, the tax rate was typically between 2 and 3 percent. Even though payroll taxes are flat rate taxes in many states, they help to make labor tax wedges increase with size because salaried contracts are common for large plants.

[15]Firms that operate in border states and make sales in these border states were granted reduced value-added tax rates in some years. Independent of location, specific goods qualify for a tax rate of 0 percent or are exempt from value-added taxation. Anton and Hernandez (2010, Table 3) calculate that approximately 45 percent of the expenditures on goods from the manufacturing sector were either zero rated or exempt.



Note: Average Repeco tax rates equal state Repeco tax revenue, reported by Fuentes-Castro et al. (2011, Table 3), divided by total state revenue of Repecos, which we calculate by summing revenue within the universe of all establishments (not just in manufacturing) in the census in a state that have revenue below 2 million pesos. The 2004 and 2009 census correspond to tax years 2003 and 2008. State initials (e.g. GJ is Guanajuato) are employed.

Figure 3: Average Repeco Tax Rates by State

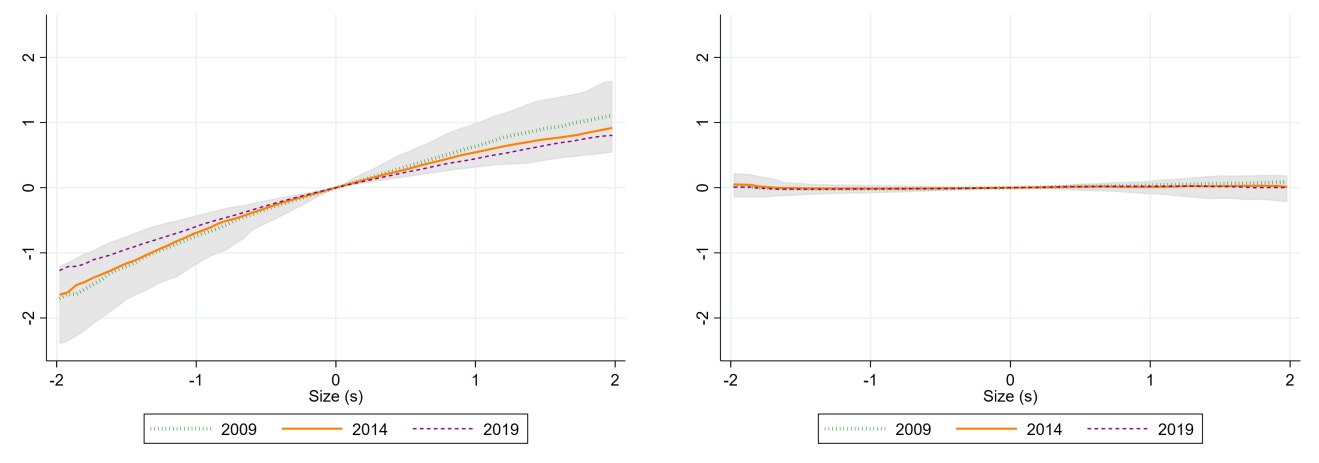
4.4 Size Dependence in Wedges

Based on Theorem 1, size dependence in the mean wedge $E[\ln wedge_j|s] - E[\ln wedge_j|s=0]$ is inferred via a calculation of $E[\ln \frac{\tilde{Rev}}{w_j x_j}|s] - E[\ln \frac{\tilde{Rev}}{w_j x_j}|s=0]$.[16] This relies on the assumption of a constant output elasticity of input j within an industry. Plant size measures are demeaned and standardized so that $s=0$ is the mean (log) industry size and $s=-1,1$ are one standard deviation below and above the mean (log) size measure for the industry, respectively.

Figure 4 plots the median (pointwise) value of $E[\ln wedge_j|s] - E[\ln wedge_j|s=0]$ across industries for each census year. The labor wedge increases with size each census year. Moreover, as size increases from $s=0$ to $s=1,2$ the labor wedge increases across the vast majority of industries. In contrast, the intermediate input wedge is quite flat in each census year.

Figure 5 documents size dependence in tax wedges $E[\ln taxwedge_j|s] - E[\ln taxwedge_j|s=0]$. Recall that $taxwedge_j = (1+T_j)/(1-T_{J+1})$ is one plus the marginal tax rate on the expenditures of input j divided by one minus the marginal tax rate on revenue - see Problem P1. Appendix B describes how tax wedges are computed using social security and value added tax data, imputing tax rates on income measures based on the simplified or the general income tax regime and imputing state payroll taxes. The benchmark tax wedge measure uses plant average tax rates,

[16] The conditional mean function $E[\ln \frac{\tilde{Rev}}{w_j x_j}|s]$ is estimated nonparametrically for each industry, given plant data $\{(\frac{\tilde{Rev}_i}{Exp_{ij}}, s_i)\}_{i=1}^n$. The estimator for $E[\ln \frac{\tilde{Rev}}{w_j x_j}|s]$ is $\frac{1}{n}\sum_{i=1}^n W_{hi}(s) \ln \frac{\tilde{Rev}_i}{Exp_{ij}}$. The weight $W_{hi}(s) = \frac{K_h(s-s_i)}{\frac{1}{n}\sum_i K_h(s-s_i)}$ employs the Epanechnikov kernel K_h with bandwidth h. See Hardle (1992, Ch. 3).



(a) Labor Wedge (b) Intermediate Input Wedge

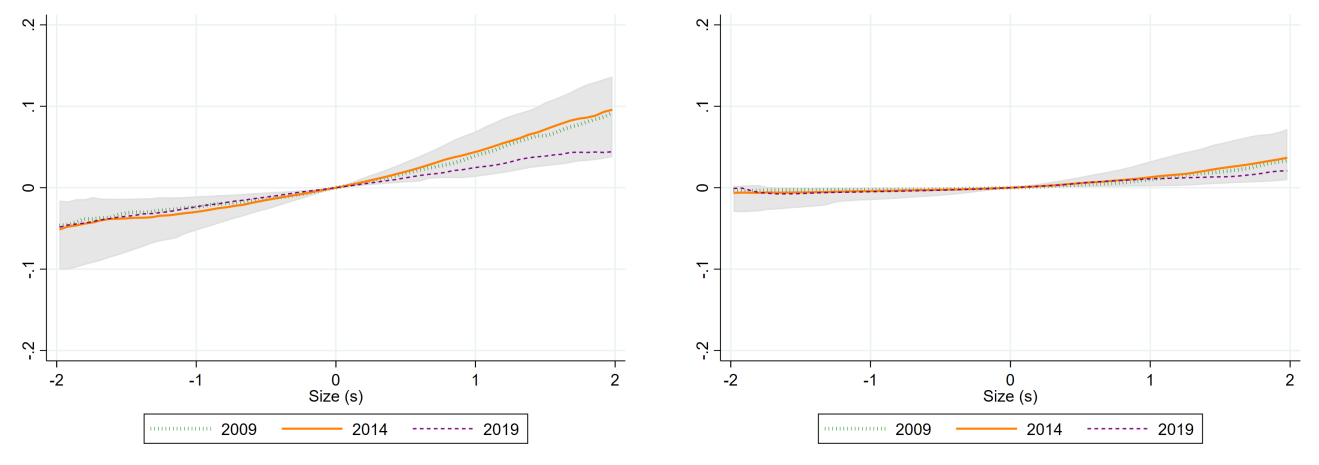
Note: The vertical axis plots $E[\ln wedge_j|s] - E[\ln wedge_j|s = 0]$. Size s is measured by (demeaned) log intermediate input expenditure, normalized to standard deviations from the industry mean, for the left panel and demeaned and normalized log labor expenditure for the right panel. The plotted line shows the median value of $E[\ln wedge_j|s] - E[\ln wedge_j|s = 0]$ across industries in a given year. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Conditional mean profiles are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth. Appendix B describes the construction of revenue, labor expenditures and intermediate input expenditures used in these calculations.

Figure 4: Mean Wedge Size Profiles: Constant Output Elasticities

for social security and value added, to approximate marginal tax rates. The conditional mean function $E[\ln taxwedge_j|s]$ is estimated nonparametrically.

Figure 5 shows that the labor tax wedge increases with size, whereas the intermediate input tax wedge is comparatively flat. The labor tax wedge displays roughly three times the variation (in log points) compared to the intermediate input tax wedge as size moves from $s = 0$ to $s = 2$. These patterns hold for three main reasons. First, the social security tax rate in Figure 2 shows substantial size variation. In addition, more than half of all plants pay no social security tax in any census year, whereas large plants hire salaried workers and have an average tax rate of between 10 and 15 percent. Small plants pay no tax partly because some hire only nonsalaried labor. Levy views non-salaried labor as a ruse that allows small firms to avoid paying social security taxes.[17] Second, state payroll taxes lead to size dependence in the labor tax wedge as the tax is on salaried labor and salaried labor is common at large firms. Third, the vast majority of plants charge more value-added tax on their output than they pay on inputs. Such plants

[17]Levy (2018, p. 38) states "Larger firms with salaried contracts are de facto taxed more than smaller ones. This is because enforcement of social insurance contributions by IMSS and Infonavit is concentrated in larger firms." Busso, Fazio and Levy (2012, p. 10) state " large firms will hire salaried workers mostly legally and therefore are taxed, small firms hiring salaried workers will do so mostly illegally and are therefore partly subsidized, and firms with nonsalaried contracts, regardless of size, are subsidized more than small firms hiring salaried workers (since they face no probability of being fined regardless of their number of workers)."



(a) Labor Tax Wedge (b) Intermediate Input Tax Wedge

Note: The vertical axis plots $E[\ln taxwedge_j|s] - E[\ln taxwedge_j|s=0]$. Size s is measured by demeaned and normalized log intermediate input expenditure for the left panel and demeaned and normalized log labor expenditure for the right panel. The plotted line shows the median value of $E[\ln taxwedge_j|s] - E[\ln taxwedge_j|s=0]$ across industries in a given year. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Conditional mean profiles are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth. Appendix B documents the computation of tax wedges.

Figure 5: Mean Tax Wedge Size Profiles

pay no marginal, value-added tax on a marginal increase in intermediate input expenditures.[18] Thus, the numerator of the labor tax wedge increases with size, whereas the numerator of the intermediate input tax wedge does not vary with size for the vast majority of Repeco-RIF plants.

The profits tax in the general income tax regime has not been highlighted as a key source of tax wedge dispersion. Why is this? A proportional tax τ_c on firm profits is a negative tax rate on marginal labor expenditure and a positive tax rate on marginal revenue. Thus, absent other taxes, $taxwedge_1 = (1+T_1)/(1-T_3) = (1-\tau_c)/(1-\tau_c) = 1$ for all such firms. Thus, while we account for it, it plays a small role in tax wedge dispersion.[19] With this said, the profits tax and the Repeco tax may be highly relevant for work that moves beyond static misallocation. For example, some firms may reduce investment in a range of productivity-increasing activities because a marginal investment makes it more likely that in future periods the firm crosses the Repeco threshold, is detected by the tax authority and permanently pays greater income taxes under a profits tax than under a Repeco revenue tax.[20]

[18] Appendix B documents the fraction of plants that charge more VAT than they pay or pay more VAT than they charge for each census year.

[19] The profits tax rate enters into the labor tax wedge only when VAT on output is less than VAT on inputs. This occurs, see Case 2 of Table B.2 in Appendix B, for a very small fraction of plants.

[20] Static misallocation analyzes the productivity effects of a reallocation of variable inputs across a fixed set of firms. Analytical results (e.g. Theorem 2-3 in this paper, the Hsieh-Klenow (2009) efficiency gains formula

4.4.1 Robustness

The size gradient of the labor wedge and the labor tax wedge are positive and the gradient of the former exceeds that of the latter. The robustness of this benchmark (Cobb Douglas) result is explored in several directions. First, wedges are computed as $\ln wedge_j = \ln(\frac{Rev}{w_j x_j}\alpha_j(x))$, calculating output elasticities $\alpha_j(x)$ using plant inputs x and parameter estimates from an industry-level production function. Standard methods are used to estimate production functions in the translog class, which generalizes Cobb-Douglas.[21] Second, wedges are computed as $\ln wedge_j = \ln(\frac{Rev}{w_j x_j}\alpha_j(x))$, calculating output elasticities $\alpha_j(x)$ based on cost share methods and cost minimization. Appendix B describes theory and methods behind the calculation. An advantage of cost share methods is that they are robust to many types of plant heterogeneity (e.g. heterogeneity in technology, factor prices and taxes).

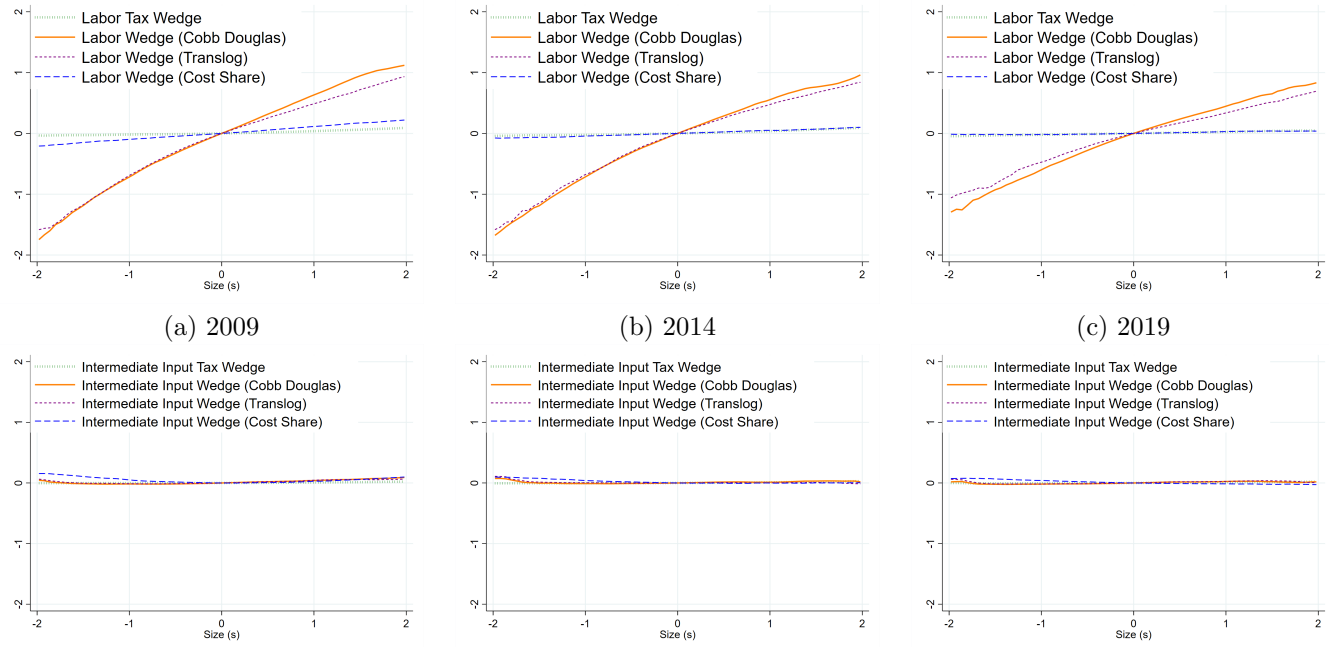
Figure 6(a)-(c) show that in each year the (median) size gradient is positive for all labor wedge measures (i.e. Cobb Douglas, translog and cost share) and these gradients typically equal or exceed the size gradient of the labor tax wedge. Size gradients differ substantially across the labor wedge measures. The fact that the translog specification has a smaller size gradient than the benchmark reflects the fact that the estimated output elasticity of labor decreases with plant size within industries. Thus, the Cobb-Douglas assumption underlying the benchmark results is not supported by the estimation. The labor wedge size gradient based on the cost share method is much smaller and reflects the fact that output elasticities of labor fall even more strongly with size. A key data property underlying this result (see Figure B.8 in Appendix B) is that the ratio of labor expenditure to intermediate input expenditure falls sharply within industries as plant size increases.

A third robustness check examines whether the labor wedge gradient result is sensitive to plausible magnitudes of overhead labor. For example, management and production workers have different roles. The Lucas span-of-control model (Lucas (1978)) assumes that managerial quality determines plant productivity and that labor input is not based on the manager's labor. For small plants, the revenue to labor expenditure ratio may be quite sensitive to including or excluding an imputation of a manager's compensation in labor expenditure.

All the calculations in Figure 6 measure labor expenditure excluding the value of a manager's time and, thus, allowing overhead labor. An alternative calculation that includes the value of

or Baqaee and Farhi (2020, Theorem 1)) focus on static misallocation and abstract from some margins of adjustment (e.g. firm entry and exit, physical capital investment, innovation investment). Ayerst, Nguyen and Restuccia (2024) provide a model-based decomposition of the magnitude of static and dynamic misallocation.

[21] These methods, see Appendix B, assume that plants share the same F and differ in a neutral technology shifter (i.e. $y = zF(x)$) that is Markovian, but are not designed to handle additional heterogeneity such as demand or product quality heterogeneity - see Problem P1 - or neutral and labor augmenting technological heterogeneity.



(a) 2009 (b) 2014 (c) 2019

(d) 2009 (e) 2014 (f) 2019

Note: The vertical axis plots $E[\ln taxwedge_j|s] - E[\ln taxwedge_j|s=0]$ or $E[\ln wedge_j|s] - E[\ln wedge_j|s=0]$. Size s is measured by demeaned and normalized log intermediate input expenditure for the top panel and demeaned and normalized log labor expenditure for the bottom panel. The plotted line shows the median value of the corresponding statistic across industries in a given year. Statistics are constructed based on a common sample of industries, for which (translog) output elasticities can be constructed.

Figure 6: Wedges and Tax Wedges

a manager's time in labor expenditures produces a labor wedge gradient, based on constant output elasticities, that is even larger than the Cobb-Douglas finding.[22]

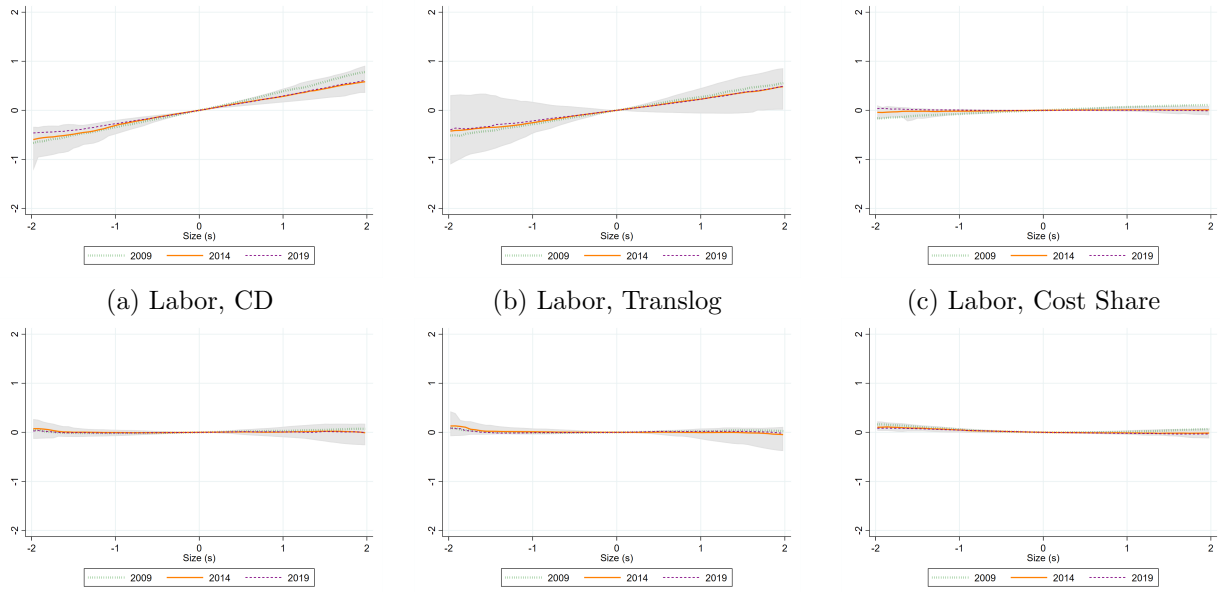
4.5 An Interpretation

We offer an interpretation of the evidence. First, a necessary condition for a solution to Problem P1 or the cost minimization problem implied by P1 is that wedges are the sum of tax wedges and a common markup per plant: $\ln wedge_j = \ln taxwedge_j + \ln markup$. Figure 6 suggests that adding up is strongly violated for the Cobb-Douglas (CD) and translog (TL) measures but not for the cost share (CS) measure. To clarify this point, define $\ln markup_j^m \equiv \ln wedge_j^m - \ln taxwedge_j$ for each plant, input j and measure $m = CD, TL, CS$.

Both the CD and TL results in Figure 7 strongly violate adding up required by the theory. For simplicity, focus on the CD results. If the CD constant output elasticity assumption is correct, then one possibility to square these results with the theory is that unmeasured

[22]Table B.1 in Appendix B describes each labor expenditure measure. The size gradient implied by abstracting from overhead labor is not plotted to avoid inessential clutter.

distortions account for the adding-up violations. For example, extortion could account for the size violations if plants face extortion tax rates that increase with plant size and are imposed on labor expenditures. What extortion tax rate T_1^{ext} would allow the "corrected" labor taxwedge $\ln(1+T_1+T_1^{ext})/(1-T_3)$ for a plant with size $s=2$ to account for all of the 100 log point increase in the markup wedge, from $s=-2$ to $s=2$, in Figure 7 (a)? A first-order approximation implies that $T_1^{ext}=1.0(1+T_1)$ so the required extortion tax rate is at least 100 percent for a plant with $T_1 \geq 0$.[23]



(a) Labor, CD (b) Labor, Translog (c) Labor, Cost Share

(d) Intermediate Input, CD (e) Intermediate Input, Translog (f) Intermediate Input, Cost Share

Note: The vertical axis plots $E[\ln markup_j^m|s]-E[\ln markup_j^m|s=0]$ for method $m\in\{CD,TL,CS\}$. The Cobb-Douglas (CD) wedge uses industry average cost shares as output elasticities (panels a, d). The translog (TL) wedge uses plant-specific output elasticities from estimated translog production functions (panels b, e). The cost share (CS) wedge uses plant-specific output elasticities from Proposition B.1 (panels c, f). Size s is measured by demeaned and normalized log intermediate input expenditure for the top panel and demeaned and normalized log labor expenditure for the bottom panel. The plotted line shows the median value across industries in a given year. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Conditional mean profiles are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth.

Figure 7: Markup Size Profiles

The CS measure in Figure 7 produces flat markup size profiles inferred from either input. Proposition B.1 in Appendix B establishes that the CS method imposes that adding up holds (i.e. the markup is the same for a plant across inputs). Based on the CS measure, the markup component of the wedge is flat with size, the labor wedge increases with size and the intermediate input wedge is flatter with size than the labor wedge. This holds across most industries

[23]The calculation assumes that plants at $s=-2$ face $T_1^{ext}=0$ and, therefore, the extortion tax for plants at $s=2$ solve $1.0=\ln(1+T_1+T_1^{ext})/(1-T_3)-\ln(1+T_1)/(1-T_3)$. A first-order approximation of the first term evaluated at $T_1^{ext}=0$ implies $T_1^{ext}\doteq 1.0(1+T_1)$.

in all years. The CS view suggests that (i) technological heterogeneity among plants accounts for the strong fall in the output elasticity of labor with size within most industries and (ii) aggregators $\mathcal{Y}(x) = x^{\gamma}$ may offer a useful approximation as they produce constant elasticity demand and, hence, an exactly flat size profile of the markup component of wedges.

Other evidence provides a useful perspective. Raval (2023) calculates labor wedges, which he calls markups, based on the CD method, whereas Raval (2023a) calculates labor wedges based on a version of the CS method that abstracts from taxes. Based on micro data sets from Chile, Colombia, India, Indonesia and a US retailer, Raval (2023, Table D1) and Raval (2023a Table B1) documents that CD labor wedges increase with size and CS labor wedges display much smaller increases with size for each data set.[24]

The principle of Occam's Razor favors the CS over the CD view. The CS view accounts for the results for Mexico and other countries based technological heterogeneity: a falling output elasticity of labor with size within industries. The CD view requires unmeasured distortions (e.g. tax or extortion) that increase dramatically with size and target labor expenditures. Technological differences across plants/firms that lead the output elasticity of labor to decrease with size are more plausibly universal across these data sets than the appropriate size variation in unmeasured distortions impacting labor wedges but not intermediate input wedges.

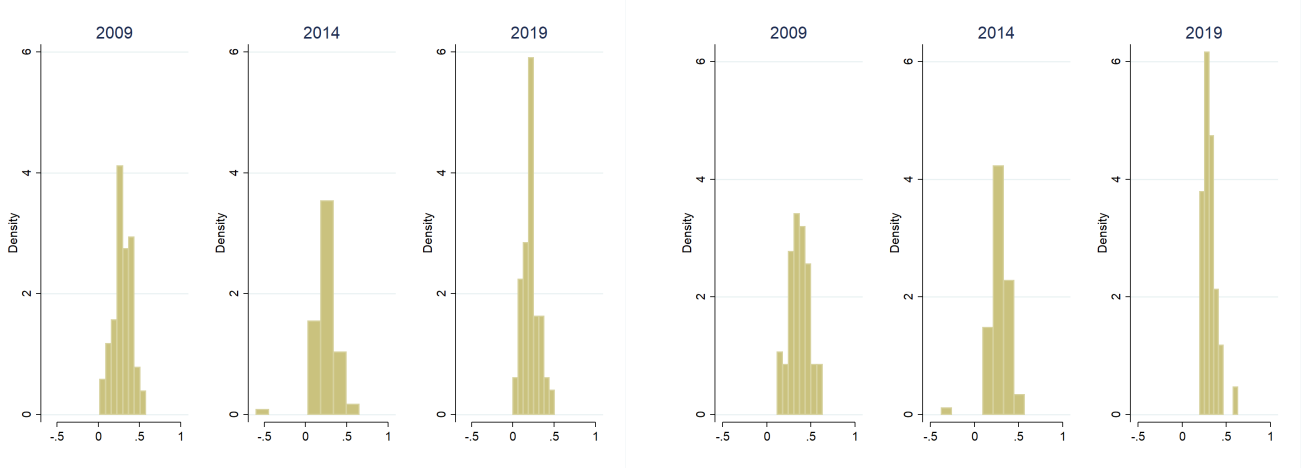
Regardless of which view more plausibly accounts for the evidence, the Levy narrative suggests that labor wedges and labor tax wedges are positively correlated and that an important component of the correlation arises from size dependence in labor tax wedges. We now determine whether these two properties hold. Figure 8 documents that the correlation is broadly positive.

To examine the importance of size dependence, decompose each industry correlation using $(*)$. Let $(X_i, Y_i) = (\ln wedge_{1i}, \ln taxwedge_{1i})$ and define $X_i = X_{1i} + X_{2i} = E[X|s(i)] + (X_i - E[X|s(i)])$ and $Y_i = Y_{1i} + Y_{2i} = E[Y|s(i)] + (Y_i - E[Y|s(i)])$, where X_{1i} and Y_{1i} are mean labor wedges and mean labor tax wedges conditional on the size $s(i)$ of plant i.

$$(*)\quad Corr(X,Y) = \frac{Cov(X,Y)}{SD_X SD_Y} = \frac{Cov(X_1,Y_1) + Cov(X_1,Y_2) + Cov(X_2,Y_1) + Cov(X_2,Y_2)}{SD_X SD_Y}$$

Table 2 shows that the majority of the positive correlation in column (2) is driven by size variation in the mean labor tax wedge (i.e. the variable $Y_{1i} = E[\ln taxwedge_1|s(i)]$) and is equal to the row sum of column (3) and (5). The remainder is driven by variation in the labor tax

[24]Demirer (2025) develops methods to estimate production functions that allow factor augmenting technological differences across firms and finds that the output elasticity of labor decreases strongly with firm size across a number of micro data sets.



(a) Labor Wedges: Cobb Douglas (b) Labor Wedges: Cost Share

Note: For each industry, $correlation(\ln taxwedge_1, \ln wedge_1)$ is computed and then histograms are displayed. Cobb Douglas labor wedges are calculated assuming constant output elasticities of labor within an industry, whereas cost share measures are calculated setting output elasticities to cost shares.

Figure 8: $Corr(\ln taxwedge_1, \ln wedge_1)$ Across Industries

Table 2: Decomposition of the Labor Wedge Correlation

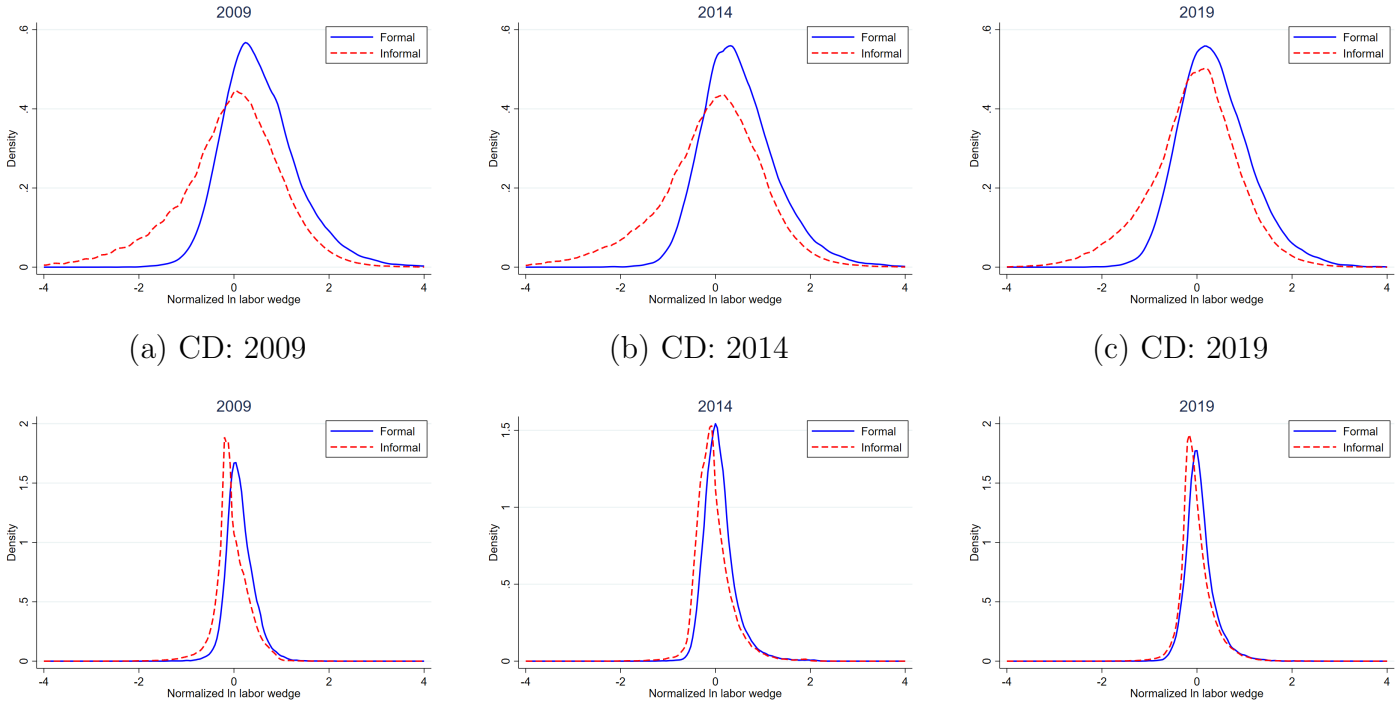
	Full Sample	Restricted Sample				
	$\frac{Cov(X,Y)}{SD_X SD_Y}$	$\frac{Cov(X,Y)}{SD_X SD_Y}$	$\frac{Cov(X_1,Y_1)}{SD_X SD_Y}$	$\frac{Cov(X_1,Y_2)}{SD_X SD_Y}$	$\frac{Cov(X_2,Y_1)}{SD_X SD_Y}$	$\frac{Cov(X_2,Y_2)}{SD_X SD_Y}$
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Cobb-Douglas Measure						
2009	0.298	0.297	0.246	0.055	-0.025	0.020
2014	0.244	0.243	0.230	0.050	-0.049	0.013
2019	0.214	0.214	0.151	0.041	-0.026	0.048
Panel B. Cost-Share Measure						
2009	0.368	0.369	0.261	-0.119	0.063	0.163
2014	0.287	0.286	0.239	-0.178	0.054	0.171
2019	0.315	0.315	0.140	-0.113	0.045	0.243

Note: Panel A uses the Cobb-Douglas measure of the labor wedge, whereas Panel B uses the cost share measure. Each entry is an average across industries. Column (1) is based on the full sample. Columns (2)-(6) restrict to industries where the local polynomial decomposition into size-dependent (X_1, Y_1) and residual (X_2, Y_2) components can be computed.

wedge within size groups. These findings are consistent with the view that variation in labor tax wedges drive a component of the variation in labor wedges and that size dependence in the labor tax wedge is particularly important.

A key reason why labor tax wedges increase with size is that many small plants pay zero social security taxes. Levy (2018, p. 38) argues that this occurs largely via size-varying enforcement of

social security taxes. A common definition for a firm/plant to be formal is that it is registered with the social security authority and, thus, pays positive social security taxes. Figure 9 shows that the density of the (normalized) labor wedge for informal plants, the majority of plants in all years, is shifted leftwards from the density for formal plants. This suggests that small reallocations of labor from informal to formal plants/firms within industries will produce productivity gains. Figure 9 shows that the dispersion in labor wedges is much larger for the CD than the CS measure.



(a) CD: 2009 (b) CD: 2014 (c) CD: 2019

(d) CS: 2009 (e) CS: 2014 (f) CS: 2019

Note: For each plant i in year t, the normalized log labor wedge is $\ln \tilde{\omega}_{1i} = \ln \omega_{1i} - E[\ln \omega_1 | \text{industry}(i)]$. Formal plants are those that pay positive social security taxes; informal plants pay zero social security taxes. Kernel densities are estimated pooling across industries after normalization. Panels (a)–(c) use the Cobb-Douglas (CD) measure; panels (d)–(f) use the cost-share (CS) measure.

Figure 9: Normalized Labor Wedge Density by Formality Status

5 Productivity Gains

This section provides two views on productivity gains. First, Levy (2018) suggests that reallocating labor to large plants may produce productivity gains. This conjecture is quantitatively assessed by applying the marginal gains formula in Theorem 3 to Mexican census data. Second, a quantitative model of the manufacturing sector in Mexico is used to gauge the equilibrium consequences of policy reforms on input reallocations and productivity change.

5.1 Productivity Gains: Formula Calculations

Based on Theorem 3, the marginal productivity gain from reallocating only the labor input (i.e. input $j = 1$) is given by the first equation below. The second equation, see Proposition C.2 in Appendix C, states the gain from a size-dependent reallocation x_1^* that scales labor for plants that are below a size threshold by a factor $(1+\gamma_1)$ and scales it by $(1-\gamma_2)$ for plants above the threshold. The threshold is set to the industry mean size (measured by labor or intermediate input expenditures) and γ_2 is set so the reallocation is feasible. The terms $\hat{E}_1[wedge_1|L]$ and $\hat{E}_1[wedge_1|H]$ denote the input weighted mean wedge, conditional on being in the low (L) or high (H) group (i.e. below or above the threshold). The third equation indicates how to calculate the gain using plant-level data.[25] [26]

$$\Omega'(0) = \omega_1 \hat{E}_1[wedge_1 \tfrac{(x_1^* - x_1)}{x_1}]$$

$$\Omega'(0) = \gamma_1 (\tfrac{\int_L x_1 d\mu}{\int_V x_1 d\mu}) \omega_1 \left[\hat{E}_1[wedge_1|L] - \hat{E}_1[wedge_1|H]\right]$$

$$\Omega'(0) = \gamma_1 (\tfrac{\sum_{i\in N_L} x_{1i}}{\sum_{i\in N_L \cup N_H} x_{1i}}) \tfrac{\sum_i w_1 x_{1i}}{\sum_i (Rev_i - w_2 x_{2i})} \left[\tfrac{1}{n_L} \sum_{i\in N_L} (\tfrac{x_{1i}}{E[x_1|L]}) wedge_{1i} - \tfrac{1}{n_H} \sum_{i\in N_H} (\tfrac{x_{1i}}{E[x_1|H]}) wedge_{1i}\right]$$

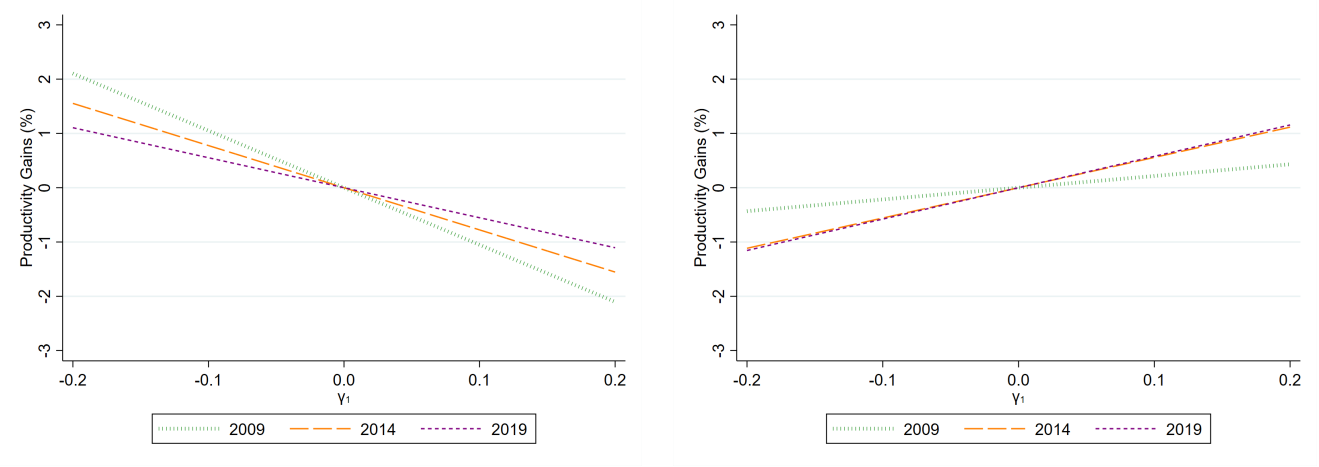
Figure 10(a) shows that reallocating labor towards large plants robustly produces productivity gains within industries. Gains are achieved because more labor is allocated to plants with higher labor wedges on average and away from plants with lower labor wedges. Plants with higher labor wedges have, in theory, a larger marginal product for the final output good when its marginal product of labor is channeled to the bundler. Figure 10(c) shows that reallocating labor in this way within all manufacturing industries produces an aggregate marginal productivity gain with the same qualitative properties as the median marginal productivity gains.

The reallocation gains are large even for small reallocations, based on Cobb-Douglas wedge measures.[27] For example, consider the case when small plants have a 10 percent reduction in labor input and large plants have a much smaller increase as they employ the majority of the labor. When $\gamma_1 = -0.1$, the median marginal productivity gain in Figure 10(a) is roughly a 1 percent gain whereas the aggregate gain in Figure 10(c) is roughly a 0.5 percent gain.

[25] Set $(wedge_{1i}, wedge_{2i}, \omega_1) = (\frac{Rev_i}{Exp_{1i}}\alpha_{1i}, \frac{Rev_i}{Exp_{2i}}\alpha_{2i}, \frac{\sum_i Exp_{1i}}{\sum_i (Rev_i - Exp_{2i})})$, where Rev_i, Exp_{1i} and Exp_{2i} are revenue, labor and intermediate goods expenditure for plant i. Set $(x_{1i}, x_{2i}) = (Exp_{1i}/w_1, Exp_{2i}/w_2)$ for arbitrary $w_1, w_2 > 0$. Table B.1 in Appendix B describes the construction of the revenue and expenditure measures. The terms n_L and n_H denote the number of plants below (L) and above (H) the threshold, whereas N_L and N_H denote the corresponding sets of plants. $E[x_1|L]$ and $E[x_1|H]$ are conditional means. $(\alpha_{1i}, \alpha_{2i})$ are plant output elasticities calculated using industry cost shares.

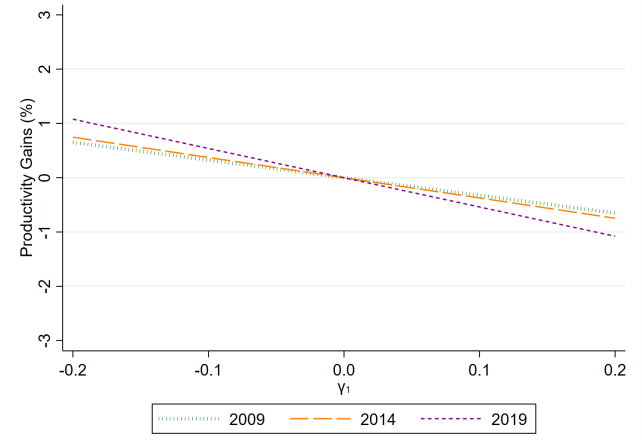
[26] Feasibility $\sum_{i=1}^n x_{1i}^* = \sum_{i\in N_L} x_{1i}(1+\gamma_1) + \sum_{i\in N_H} x_{1i}(1-\gamma_2) = \sum_{i=1}^n x_{1i}$ implies $\gamma_2 = \gamma_1 \frac{\sum_{i\in N_L} x_{1i}}{\sum_{i\in N_H} x_{1i}}$.

[27] Reallocation gains based on the cost share measure are smaller.

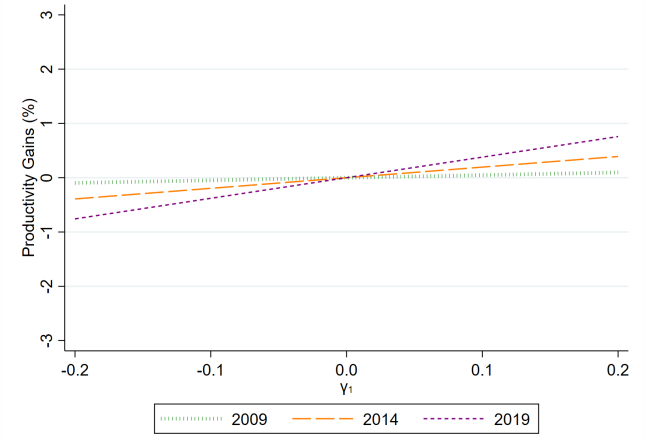


(a) Reallocate Labor

(b) Reallocate Intermediate Input



(c) Reallocate Labor, Agg. Gains



(d) Reallocate Intermediate Input, Agg. Gains

Note: Panels (a) and (b) plot the median gains across industries for each year as a function of the scale γ_1 of the reallocation. The high (H) and low (L) sets of plants N_H and N_L are defined based on whether intermediate input expenditure exceeds the industry mean. Panels (c) and (d) plot the aggregate gains by conducting such reallocations in all industries s, with (labor) gains $\Omega'(0) = \gamma_1 \sum_s \omega_{1s} (\frac{\int_{L_s} x_{1s} d\mu_s}{\int_V x_{1s} d\mu_s})[\hat{E}_{1s}[wedge_{1s}|L] - \hat{E}_{1s}[wedge_{1s}|H]]$ implied by Proposition C.2 and C.3. Appendix B describes the construction of all the variables that enter these calculations. All results are based on using the Cobb-Douglas measurement of input wedges.

Figure 10: Marginal Productivity Gains from Input Reallocation

Reallocating intermediate input in the same way produces marginal productivity gains that are smaller in absolute value. They are smaller mainly because the intermediate input wedge size gradient is relatively flat. One reason why the results are not even smaller is that the weight ω_2 attached to intermediate input reallocations is roughly three times the weight ω_1 associated with labor reallocations. This is consistent with the fact that intermediate input cost shares are typically much larger than labor cost shares.

The fact that productivity gains from reallocating labor and intermediate input have opposite

signs is noteworthy. It suggests that policy reforms that flatten the size gradient of VAT tax rates may not be well focused. If such a small reform acts to increase labor and intermediate input at large plants and decrease them at smaller plants, then Figure 10 implies that the resulting marginal productivity gain will be the result of partially offsetting marginal gains.

5.2 Productivity Gains: An Equilibrium Analysis

Consider the Lucas (1978) model but modified to allow monopolistic competition, an industry output aggregator and firm taxes. A probability measure μ describes the distribution of the state variable $v = (z, \tau) \in V$ across agents. If an agent with state $v = (z, \tau)$ chooses to be an entrepreneur, then (i) production $y = F(x_1, x_2; z)$ is determined by inputs (x_1, x_2) of labor and intermediate input and the agent's neutral and labor augmenting productivity $z = (z_1, z_2)$, (ii) taxes $T(w_1x_1, w_2x_2, py; \tau)$ are a function of labor expenditures w_1x_1, intermediate input expenditures w_2x_2, revenue py and a vector τ describing how the entrepreneur is taxed and (iii) profits are $\pi(v)$. If an agent chooses to be a worker, then the agent supplies one unit of labor inelastically to the market and receives earnings equal to w_1.

A competitive bundler maximizes profit by producing output Y and buying all the differentiated goods produced by entrepreneurs. Production is defined by a Kimball aggregator $\int_E \mathcal{Y}(y(v)/Y)d\mu = 1$, where $E \subset V$ is the set of active entrepeneurs. The bundler faces set prices $p(v)$ for a differentiated good supplied by an entrepreneur with state v. Entrepreneurs perceive the demand relationship $P(y, Y)$ induced by the bundler's choices. Prices are stated in units of the bundler's output so that $w_2 = 1$.

Definition: *An equilibrium is* $(c(v), s(v), x_1(v), x_2(v), \pi(v), P(y, Y), w_1, w_2, Y)$ *such that, given government spending* G*,*

1. *agents:* $(c(v), s(v)) \in argmax_{c,s}[u(c)|\ c = s\pi(v) + (1-s)w_1$ *and* $s \in \{0, 1\}]$ *for all* $v \in V$*.*

2. *entrepreneur: For all* $v \in V$*,*

   $(x_1(v), x_2(v)) \in argmax_{(x_1,x_2)} P(y, Y)y - \sum_j w_j x_j - T(w_1x_1, w_2x_2, P(y, Y)y; \tau)$ *s.t.* $y = F(x_1, x_2; z)$

   $\pi(v) = P(y(v), Y)y(v) - \sum_j w_j x_j(v) - T(w_1x_1(v), w_2x_2(v), P(y(v), Y)y(v); \tau)$

   $y(v) = F(x_1(v), x_2(v); z)$

3. *bundler:* $y \in argmax_y[Y - \int_E p(v)y(v)d\mu]$ *s.t.* $\int_E \mathcal{Y}(y(v)/Y)d\mu = 1$

   $p(v) = P(y(v), Y)$ *and* $P(y, Y) = \mathcal{Y}'(\frac{y}{Y}) / \int_E \mathcal{Y}'(\frac{y(v)}{Y})\frac{y(v)}{Y}d\mu$

   $E = \{v \in V : s(v) = 1\}$ *is the set of active entrepreneurs.*

4. *government:* $G = \int_E T(w_1 x_1(v), w_2 x_2(v), P(y(v), Y) y(v); \tau) d\mu$

5. *feasibility:* $\int_V c(v) d\mu + \int_E x_2(v) d\mu + G = Y$ *and* $\int_E x_1(v) d\mu = 1 - \mu(E)$

Let F be a constant elasticity of substitution production function with neutral z_1 and labor augmenting z_2 technology. This structure implies (see Proposition E.1 in Appendix E) that the ratio of labor to intermediate input expenditures falls as z_2 increases: $\ln \frac{w_1 x_1}{w_2 x_2} = C_0 - \sigma \ln(\frac{1+T_1}{1+T_2}) - (1-\sigma) \ln z_2$, when $\sigma = 1/(1-\alpha) \in (0,1)$. Intuitively, dispersion in neutral productivity z_1 accounts for much of the plant size distribution and a positive correlation between $\ln z_1$ and $\ln z_2$ accounts for the fall in the ratio with size.

1. Production: $y = F(x_1, x_2; z) = z_1 [a(x_1 z_2)^\alpha + (1-a) x_2^\alpha]^{1/\alpha}$ and $\mathcal{Y}(x) = x^\gamma$, where $\sigma = 1/(1-\alpha)$ is the elasticity of substitution

2. Agent Heterogeneity: $(\ln z_1, \ln z_2, \tau) \sim N(\mu, \Sigma)$, where $\tau = (\tau^{labor}, \tau^{va}, \tau^{va,m})$.

3. $T(\cdot; \tau) = \tau^{labor} w_1 x_1 + \tau^{va,m} w_2 x_2 + \tau^{va} py + \tau^c \left[(1-\tau^{va}) py - (1+\tau^{labor}) w_1 x_1 - (1+\tau^{va,m}) w_2 x_2 \right]$

4. Parameters: $\Gamma = (\alpha, \gamma, a, \mu, \Sigma)$; τ^c (set externally)

Covariance restriction: each tax covaries only with z_1. All tax–tax and tax–z_2 covariances are set to zero:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{12} & \sigma_{22} & 0 & 0 & 0 \\ \sigma_{13} & 0 & \sigma_{33} & 0 & 0 \\ \sigma_{14} & 0 & 0 & \sigma_{44} & 0 \\ \sigma_{15} & 0 & 0 & 0 & \sigma_{55} \end{pmatrix}$$

6 Discussion

We provide evidence that the mean labor tax wedge and labor wedge increase with size in Mexican data. The labor tax wedge finding is robust across the majority of industries in all time periods. The labor wedge finding is robust (i) across the majority of industries in all time periods, (ii) to calculating wedges based on three approaches to measuring output elasticities and (iii) to plausible magnitudes of overhead labor.

Two takeaways build on these findings.

1. The first part of the Levy narrative - labor tax wedges increase with size - is supported across time and industries. Social security taxes play a key role in this finding.

2. There are marginal productivity gains to reallocating labor to large plants within industries. These gains (see Figure 10) are quantified for small reallocations, based on Theorem 3, and their magnitude depends on measured size differences in average labor wedges.

Quantitative equilibrium models are constructed to address the second part of the Levy narrative. We posit a production function with neutral and labor augmenting technological heterogeneity and a constant elasticity aggregator. The model is calibrated to match the plant size distribution and the fall in the ratio of labor to intermediate input expenditures with size as well as the gradients of labor tax wedges and labor wedges. It is key that the two technologies covary positively as, other things equal, an increase in labor-augmenting technology reduces a plant's expenditure ratio when the elasticity of substitution between these inputs is below 1. The constant elasticity aggregator produces an exactly constant markup component of wedges, which closely approximate the flat size relationship of markups inferred from Mexican data using the cost-share methods.

Cobb-Douglas production functions, commonly used in the misallocation and markups literatures, are difficult to reconcile with Mexican data. For example, the ratio of labor to intermediate input expenditures robustly falls with plant size within industries in Mexico. In addition, size variation in the labor and intermediate input wedges, inferred using the Cobb-Douglas assumption, can only be squared with theory if there are large, size-varying distortions beyond those measured based on the tax system.

We use the quantitative equilibrium model to analyze the productivity effects of a policy reform that eliminates two firm-level taxes on labor expenditure, the social security tax and the state payroll tax, and replaces the lost revenue with a uniform increase in the value added tax rate. The policy reform eliminates all variation, both size variation and variation within size groups, in the numerator of the labor tax wedge arising from these taxes.

References

Ackerberg, D., Caves, K. and G. Frazer (2015), Identification Properties of Recent Production Function Estimators, Econometrica, 83, 2411-51.

Alvarez, J. and C. Ruane (2019), Informality and Aggregate Productivity: the Case of Mexico, IMF Working Paper 257.

Anton, A. and F. Hernandez (2010), VAT Collection and Social Security Contributions: Is There a Link?, IDB Working paper.

Anton, A., Hernandez, F. and S. Levy (2012), The End of Informality in Mexico? Fiscal Reform and Universal Social Insurance, Inter-American Development Bank.

Asker, J., Collard-Wexler, A. and J. De Loecker (2014), Dynamic Inputs and Resource (Mis)Allocation, Journal of Political Economy, 122, 1013- 63.

Ayerst, S., Nguyen, D. and D. Restuccia (2024), The Micro and Macro Productivity of Nations, NBER Working Paper 32750.

Bachas, P., Fattal Jaef, R. and A. Jensen (2019), Size-dependent Tax Enforcement and Compliance: Global Evidence and Aggregate Implications, Journal of Development Economics, 140, 203-22.

Bartelsman, E., Haltiwanger, J. and S. Scarpetta (2013), Cross-Country Differences in Productivity: The Role of Allocation and Selection, American Economic Review, 103, 305-334.

Baqaee, D. and E. Farhi (2020), Productivity and Misallocation in General Equilibrium, Quarterly Journal of Economics, 135, 105- 163.

Bergquist, L., Lashkari, D. and E. Verhoogen (2026), Wedges: A Microeconomic Perspective on Misallocation, NBER Working Paper 34756.

Bils, M., Klenow, P. and C. Ruane (2021), Misallocation or Mismeasurement?, Journal of Monetary Economics, 125, S39-56.

Blackwood, J., Foster, L., Grim, C., Haltiwanger, J. and Z. Wolf (2021), Macro and Micro Dynamics of Productivity: From Devilish Details to Insights, American Economic Journal: Macroeconomics, 13, 142-72.

Busso, M., Fazio, M. and S. Levy (2012), (In)Formal and (Un)Productive: the Productivity Costs of Excessive Informality in Mexico, IDB Working Paper 341.

Busso, M., Madrigal, L. and C. Pagés (2013), Productivity and Resource Misallocation in Latin America, BE Journal of Macroeconomics, 13, 903-32.

Busso, M., Levy, S. and J. Torres (2019), Establishment Dynamics and the Persistence of Resource Misallocation in Mexico: An Analysis of Longitudinal Data for 1998-2013, manuscript.

Cusolito, A. and W. Maloney (2018), Productivity Revisited: Shifting Paradigms in Analysis and Policy, World Bank Group.

David, J. and V. Venkateswaran (2019), The Sources of Capital Misallocation, American Economic Review, 109, 2531-67.

DeLoecker, J. and F. Warzynski (2012), Markups and Firm-Level Export Status, American Economic Review, 102, 2437- 71.

Demirer, M. (2025), Production Function Estimation with Factor-Augmenting Technology: An Application to Markups, manuscript.

DeRidder, M., Grassi, B. and G. Morzenti (2024), The Hitchiker's Guide to Markup Estimation: Assessing Estimates from Financial Data, manuscript.

Fuentes-Castro, H., Zamudio-Carrillo, A., Barajas-Cortes, S. and A. Brown del Rivero (2011), Estudio de Evasion Fiscal en el Regimen de Pequenos Contribuyentes, Tecnologico de Monterrey Working Paper.

Garicano, L., Lelarge, C. and J. Van Reenen (2016), Firm Size Distortions and the Productivity Distribution: Evidence from France, American Economic Review, 106, 3439-79.

Guner, N., Ventura, G. and Y. Xu (2008), Macroeconomic Implications of Size-Dependent Policies, Review of Economic Dynamics, 11, 721-44.

Hall, R. and C. Jones (1999), Why Do Some Countries Produce So Much More Output Per Worker Than Others?, Quarterly Journal of Economics, 114, 83-116.

Hanson, G. (2010), Why Isn't Mexico Rich?, Journal of Economic Literature, 48, 987-1004.

Hardle, W. (1992), Applied Nonparametric Regression, Cambridge University Press.

Hopenhayn, H. (2014), Firms, Misallocation, and Aggregate Productivity: A Review, Annual Review of Economics, 6, 735- 70.

Hsieh, C. and P. Klenow (2009), Misallocation and Manufacturing TFP in China and India, Quarterly Journal of Economics, 124, 1403-48.

Hsieh, C. and P. Klenow (2014), The Life Cycle of Plants in India and Mexico, Quarterly Journal of Economics, 129, 1035-84.

Inter-American Development Bank (2010), The Age of Productivity, Inter-American Development Bank, Editor C. Pages.

Kaymak, B. and I. Schott (2025), Tax Heterogeneity and Misallocation, manuscript.

Kehoe, T. and K. Ruhl (2010), Why Have Economic Reforms in Mexico Not Generated Growth?, Journal of Economic Literature, 48, 1005- 27.

Kimball, M. (1995), The Quantitative Analytics of the Basic Neomonetarist Model, Journal of Money, Credit and Banking, 27, 1241-77.

Leal-Ordonez, J. (2014), Tax Collection, the Informal Sector, and Productivity, Review of Economic Dynamics, 17, 262-86.

Levy, S. (2008), Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico, Brookings Institution Press.

Levy, S. (2018), Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico, Inter-American Development Bank.

Lucas, R. (1978), On the Size Distribution of Business Firms, The Bell Journal of Economics, 9, 508-23.

Luenberger, D. (1969), Optimization by Vector Space Methods, Wiley-Interscience.

Misch, F. and C. Saborowski (2018), Resource Misallocation and Productivity: Evidence from Mexico, IMF Working Paper 112.

Orozco-Vazquez, M. (2023), Misallocation of Resources, Firm Characteristics, and Structural Factors: Evidence from Mexico, Banco de Mexico Working Paper.

Raval, D. (2023), Testing the Production Approach to Markup Estimation, Review of Economic Studies, 90, 2592-2611.

Raval, D. (2023a), A Flexible Cost Share Approach to Markup Estimation, Economic Letters, 230, 111262.

Restuccia, D. (2013), The Latin American Development Problem: An Interpretation, Economia, 13, 69-100.

Restuccia, D. and R. Rogerson (2008), Policy Distortions and Aggregate Productivity with Heterogeneous Establishments., Review of Economic Dynamics 11, 707–20.

Restuccia, D. and R. Rogerson (2017), The Causes and Costs of Misallocation, Journal of Economic Perspectives, 31, 151-74.

Rotemberg, M and T. White (2020), Measuring Cross-Country Differences in Misallocation, manuscript.

Sanchez-Vela, C. and J. Valero-Gil (2011), The Effect of Firm-size Dependent Policies on the Economy: the Case of the Repeco Law in Mexico, IDB WP 216.

Syverson, C. (2011), What Determines Productivity?, Journal of Economic Literature, 49, 326-65.

Appendix: What Drives Underdevelopment?

Mark Huggett
Georgetown University

Wenlan Luo
Tsinghua University

Rodimiro Rodrigo
George Washington University

Appendix A Firm Taxation in Mexico **3**

A.1 Social Security Tax 3

A.2 State Payroll Tax 6

A.3 Value Added Tax 6

A.4 Corporate Profit Tax 8

A.5 Simplified Regimes: REPECO, RIF, and RESICO 10

Appendix B Mexican Census Data **11**

B.1 Variable Construction 11

B.2 Constructing Tax Wedges 12

B.3 Constructing Cost Shares 15

B.4 Estimating Output Elasticities 15

B.5 Properties of Tax Rates, Tax Wedges and Wedges 24

B.6 Reallocation Experiments 28

B.7 The Distribution of Revenue 29

Appendix C Additional Theory **31**

Appendix D Levy Narrative **35**

Appendix E Calibrating Model Parameters **36**

Appendix F Additional Descriptive Statistics **42**

F.1 Updated List (updated SSRate) 42

F.2 Updated List (updated L4) 42

F.3 Updated List (updated ii) 42

F.4 Updated List (1/29/2026) 42

Appendix A Firm Taxation in Mexico

The section summarizes five types of firm taxes in Mexico: social security, state payroll, value added, simplified regime and general regime income taxes.

A.1 Social Security Tax

Social security taxes in Mexico are established in the Social Security Law (SSL, *Ley de Seguridad Social*) and the National Workers' Housing Fund Law (*Ley del INFONAVIT*) are calculated based on the employee's base salary (*salario base de cotización*) that includes wages, bonuses, commissions, and other compensation (Article 27), and is capped at 25 times the minimum salary (Article 28).[1] [2] Taxes are shared between employers, employees, and in some cases, the government, and are collected by IMSS (*Instituto Mexicano de Seguridad Social*) and INFONAVIT (*Instituto del Fondo Nacional de la Vivienda para los Trabajadores*).

Social security taxes on firms collected by IMSS, consist of tax rates on base salary for five insurance categories (Article 11) that are described in Table A.1: labor risk; illness and maternity; disability and life insurance; retirement pensions and elder severance payment; and childcare and social benefits.

Table A.1: Social Security Tax Rates: IMSS and INFONAVIT

Category	Employer	Article, SSL
Labor risk	0.54 - 7.59%	73
Illness and maternity		
Fixed Rate: Salaries up to 3 UMAs*	20.40%	106, I.19
Marginal Rate: Salaries beyond 3 UMAs*	1.10%	106, II.19
In cash benefits	0.70%	107
In kind benefits	1.05%	25
Disability and life insurance	1.75%	147
Retirement fund		
Retirement (AFORE)	2.00%	168, I
Elder	3.15 - 11.87%	168, II
Childcare and social benefits	1.00%	211
Housing (INFONAVIT)	5.00%	29, II (Ley INFONAVIT)

Note: Elaborated by the authors with information from the SSL. (*) An UMA (*Unidad de Medida y Actualizacion*) is the economic reference value in pesos used to determine the amount of obligations, fines, taxes, and benefits defined by the government starting in 2016.

[1] Before 2016, the cap on base salary was 25 times the minimum wage. After 2016, the cap on base salary is 25 UMAs.

[2] https://www.ordenjuridico.gob.mx/Documentos/Federal/pdf/wo9056.pdf

The tax rate for the labor risk insurance category depends on the economic activity of the firm, which is classified into five classes as shown in Table A.2. The rate is determined annually and ranges from 0.54% to 7.59% of the employee's base salary. The risk premium is also adjusted based on the company's history of work-related accidents. The illness and maternity insurance tax consists of a fix rate of 20.4 percent for salaries below 3 UMAs, and a marginal rate for the portion of base salary beyond 3 UMAs of 1.1 percent, complemented by contributions for in cash and in kind benefits provided by law.

Table A.2: Risk premium by class (%)

Class	%	Examples
I	0.54	Clothing manufacturing
II	1.13	Hotels and restaurants
III	2.60	Agriculture
IV	4.65	Machinery production
V	7.59	Mining, plastic products, construction

Note: Elaborated by the authors with the rates from Article 73 of the SSL, and examples from Article 196 of the regulation of the SSL.

The SS system has remained without major changes after the 1997 pension reform introducing AFOREs (private individual retirement accounts), where contributions for retirement savings were shifted to individual accounts managed by private institutions (Articles 154-168 of the SSL).[3] Before 1997, the retirement pension component of social security was funded on a pay-as-you-go basis. In addition to the contributions to IMSS, firms need to pay a 5 percent rate of workers' base salary to INFONAVIT for housing benefits to aquire, build or renew their home.

Beyond social protection, Mexico's constitution grants salaried workers the right to a share of their employers' profits or PTU (*participación de los trabajadores en las utilidades*). The statutory PTU share is 10% of the firm's taxable income after deductions, reported to the Tax Administration Service (SAT), if positive. Firms have to pay PTU within 60 days after filing their annual tax return. Its compliance is enforced by the Ministry of Labor (STPS).[4]

A.1.1 Enforcement

Employers in Mexico are legally required to register themselves with the IMSS as well as report hiring, separations, and changes of salaries for each of their salaried workers within five business days of these

[3]Employers contributing 2% (see Table A.1) and employees 1.125% of the base salary.

[4]Employees need at least 60 days in the firm to receive PTU. PTU payments are capped at three months of an employee's salary and prorated for employees who have been in the firm for less than a year. The law exempts firms from profit sharing during their first two years of operation and authorized nonprofit organizations. All other firms are mandated to pay PTU to their workers including Repeco and RIF firms.

events (LSS, Art. 15, section I). Employees have the right to request their own registration with the IMSS and report salary changes independently of their employers, without absolving employers of their legal obligations (Art. 18).

Once registered, employers are responsible for accurately calculating and paying the required social security contributions for each worker (LSS, Art. 15, section III). The contribution base (known as the Salario Base de Cotización) includes cash wages, bonuses, commissions, and most in-kind benefits (LSS, Art. 27), subject to defined exclusions. The contribution base is capped at 25 times the applicable minimum wage (Art. 28). Employers must determine contributions monthly and submit their payment by the 17th of the following month (Art. 39). Employers are also responsible for withholding employee-side contributions from workers' pay and remitting them fully to the IMSS (Art. 304 A, section XVII). If contributions are paid late, interest and surcharges apply (Art. 40 C).

Table A.3: Violations and Penalties under the *Ley del Seguro Social*

Violation Type	LSS Article(s)	Penalty
Failure to register as employer	304 A I	20 to 350 UMA
Failure to register new workers or late registration	304 A II	20 to 350 UMA
Failure to report salary changes	304 A III	20 to 125 x min wage
Failure to calculate or report contributions	304 A IV	20 to 75 x min wage
Submitting false contribution or payroll data	304 A VI	20 to 210 x min wage
Failure to remit withheld employee contributions	304 A XVII	20 to 350 UMA
Failure to file occupational risk premium annual revision	304 A XV	20 to 210 x min wage
Failure to submit required external audit (firms with $>$ 300 workers)	304 A XX	20 to 350 UMA
Fraudulent evasion of contributions	307–308	3 months to 9 years in prison

Employers must also comply with occupation risk insurance (Seguro de Riesgos de Trabajo) requirements. Each year, firms are obligated to file an annual review of their occupational risk premium (Prima de Riesgo de Trabajo), reflecting their actual workplace accident experience and risk classification (LSS, Art. 71, and associated regulations). Failure to file the required risk premium report, or filing it late or with false information, is subject to sanctions (Art. 304 A, section XV).

Firms employing an annual average of 300 or more workers in the previous fiscal year, are subject to an additional compliance mechanism. They must submit an external audit report (dictamen) of their IMSS obligations, prepared by an independent authorized public accountant (LSS, Art. 16). Timely

submission of this report exempts the firm from domiciliary inspections for the audited years, unless material discrepancies are identified. Failure to submit the required audit, or submitting a false or incomplete audit, is subject to penalties (Art. 304 A, sections XIX and XX).

The LSS provides the IMSS with broad enforcement powers (Art. 251), including authority to conduct inspections, require documentation, determine employers' contribution obligations, and assess penalties. All unpaid contributions, interest, surcharges, and fines are treated as fiscal credits enforceable under Mexican tax law (Art. 287). Table A.3 summarizes the main types of violations and the corresponding penalties under the LSS. Anton, Hernandez and Levy (2012, p.103) claim that in practice the penalty imposed on a firm caught evading social security taxes is 150 percent of the upaid taxes.

A.2 State Payroll Tax

State payroll taxes (SPT) are determined, collected and administered by each state treasury office, established in their respective Tax Codes, and are the main source of revenue of state governments in Mexico. Table A.4 reports the rates by state for the census years analyzed (i.e. the data reported corresponds to the previous year when the census data was collected).

The payroll measure used to compute a firm's SPT payment is the sum of all salary paid to salaried workers considering vacation bonus, commissions, compensations, overtime, seniority bonus. The base of this tax is not based on the contributions to IMSS, INFONAVIT, AFORE, in kind payments, or other expenses on the workers like training or work equipment. Neither Repeco nor RIF firms were exempt from paying the SPT.

A.3 Value Added Tax

The Value Added Tax (VAT) in Mexico is regulated by the VAT Law (VATL, *Ley de Impuesto al Valor Agregado*) and is a general consumption tax applied to the sale of goods, provision of services, leasing of goods, and imported goods and services.[5] Firms charge and collect VAT on their sales and pay VAT on their purchases (Article 5). The difference is paid to the tax authority (*Servicio de Administración Tributaria* - SAT) and is filed monthly.

Table A.5 shows the VAT rate during our analysis period. Two fiscal reforms changed the VAT rate, the 1995 reform, which increased the VAT rate from 10% to 15%, and the 2009 reform, which increased the rate to 16%. The general VAT rate remained at 16% until December 2024.[6] [7]

Historically, the VAT rate has differed for firms in the border regions. The 1995 fiscal reform increased the VAT rate from 6 to 10% for all trade on the frontier regions (Article 2, except for real

[5]https://www.diputados.gob.mx/LeyesBiblio/pdf/LIVA.pdf
[6]https://www.diputados.gob.mx/LeyesBiblio/ref/liva/LIVA_ref20_27mar95.pdf
[7]https://www.diputados.gob.mx/LeyesBiblio/ref/liva/LIVA_ref40_07dic09.pdf

Table A.4: State Payroll Tax Rates

State	1999	2004	2009	2014	2019	2024
Aguascalientes	0	0	1.5	2	2	2
Baja California	1.8	1.8	1.8	1.8	1.8	1.8
Baja California Sur	2	2	2.5	2.5	2.5	2.5
Campeche	2	2	2	2.5	2.5	3
Chiapas	1.5	1.5	2	2	2	2
Chihuahua	1.8	1.8	1.8	3	3	4
Coahuila	1	1	1	2	2	2
Colima	0	0	2	2	2	2
Ciudad de México	2	2	2	2.5	3	3
Durango	1.1	1.1	2	2	2	2
Estado de México	2	2.5	2.5	2.5	3	3
Guanajuato	0	0	2	2	2	3
Guerrero	0	2	2	2	2	2
Hidalgo	0	0	3	3	3	3
Jalisco	2	2	2	2	2	2.5
Michoacán	0	2	2	2	2	3
Morelos	0	0	2	2	2	2
Nayarit	0	0	2	2	2	3
Nuevo León	2	2	2	3	3	3
Oaxaca	0	1	1	2	3	3
Puebla	1	1	2	2	3	3
Querétaro	0	0	1.6	1.6	2	3
Quintana Roo	2	2	2	2	3	3
San Luis Potosí	1	1	2	2	2.5	3
Sinaloa	1.5	1.5	1.5	1.5	1.5	3
Sonora	2	2	2	2	2	2
Tabasco	1	1	1	2.5	2.5	2.5
Tamaulipas	0	2	2	2	2	3
Tlaxcala	2	2	2	2	2	3
Veracruz	0	2	2	2	3	3
Yucatán	2	2	2	2.5	2.5	3
Zacatecas	0	1.9	1.9	1.9	1.9	3

Note: Elaborated by the authors with the rates from Chapa, Ayala, and Cabral (2016) and State's Fiscal Codes where we verified the years with changes in the State tax rates as well as the year taxes started in each State. The reported rates correspond to the prevalent rates the previous year, when the data for each census was collected. In states with progressive taxes, either based on salary or number of workers, we report the average of the lower and upper bounds.

Table A.5: Value Added Tax Rate, 1991-2024

Reform	VAT rate	VAT rate Frontier
November 1991	10	6
March 1995	15	10
December 2009	16	11
January 2014*	16	16
January 2019′	16	8

Note: Constructed by the authors using information from the VATL of 2021 and reforms from 2009, 1995, and 1991, and the Executive Orders of 2013 (*), 2018 and 2020 (').

estate).[8] The 2009 reform increased the VAT rate on the frontier to 11%. Later, three executive orders modified the VAT rate on frontier regions. The first in 2013 equalized the rate across the country from January 2014.[9] A second executive order in December 2018, reduced the VAT rate to 8% from January 2019 to December 2020, for around 40 municipalities on the northern frontier.[10] At the end of 2020, a third executive order extended the period of special treatment to the frontier until December 2024.[11]

The VATL has a 0% rate for food, prescribed medicines, and agricultural products (Article 2-A). In these cases, the seller charges no VAT but can still claim input VAT deductions. The law also exempts some goods and services (Articles 9 and 15, respectively) from VAT, namely residential mortgages, retirement funds, public education, and public transportation. The list of goods and services exempt from VAT or taxed at 0% has been adjusted over time.

A.4 Corporate Profit Tax

The Corporate Profit Tax (CPT) in Mexico is regulated by the CPT Law (CPTL, *Ley del Impuesto Sobre la Renta or Ley del ISR*).[12] Table A.6 describes the computation of taxable corporate profits and the tax paid for firms or legal entities. Since 1998, there have been special regimes for small firms. The profits tax does not apply to such firms. Table A.7 presents the changes to the CPT rate from 1990 to 2024, associated with the different fiscal reforms during those years.

[8]The 1995 reform defined the frontier as the 20-kilometer border strip parallel to the international limits to the north and south of the country, the northern States of Baja California and Baja California Sur, and in the south the State of Quintana Roo, and some additional municipalities of Sonora.

[9]https://www.dof.gob.mx/nota_detalle.php?codigo=5328038&fecha=26/12/2013#gsc.tab=0

[10]https://www.dof.gob.mx/nota_detalle.php?codigo=5547485&fecha=31/12/2018#gsc.tab=0

[11]https://www.dof.gob.mx/nota_detalle.php?codigo=5609181&fecha=30/12/2020#gsc.tab=0

[12]https://www.diputados.gob.mx/LeyesBiblio/pdf/LISR.pdf

Table A.6: Corporate Profit Tax, Computation

Concept	Description	Articles
Taxable Income	Income from sales, services, leases, and other activities	16-24
Allowable Expenses	Wages and contributions to social security, purchases, depreciation, interest, and operating expenses	25-30
Tax Credits	Credits for R&D, renewable energy, double taxation agreement	27-30
Profit	= Taxable Inc. - Allow. Exp. - Tax Credits	
PTU	= max{Profit x 0.10, 0}	9.I & Const 123
Cumulative Losses	Losses pending to apply from previous years.	9.II
Corporate Profit Tax	=(Profit – PTU* – Cum losses) x CPT rate	9

Note: Elaborated by the authors with information from the CPTL. PTU* is PTU paid one year prior to the PTU calculation in the table above. This is because the base in calculating taxable profit uses only expenditures occurring in the calendar year and because PTU is paid after the calendar year ends.

Table A.7: Corporate Profit Tax Rate, 1990-2024

Year	CPT rate
1990	34
1999	35
2003	34
2004	33
2005	30
2006	29
2007	28
2010	30

Note: Elaborated by the authors with information from the CPTL reporting years with tax rate changes.

A.5 Simplified Regimes: REPECO, RIF, and RESICO

Régimen de Pequeños Contribuyentes (Repeco) was established by 1998 Mexico's fiscal reform to simplify tax compliance for small businesses and individuals with low-income levels. Repeco taxes are based on firm revenue rather than firm profit. Under Repeco, firms with revenue below 2 million Mexican pesos (Article 137 - 140, CPTL of 1998) pay an income tax and VAT determined by state and municipal governments based on the type and revenue of their business activities. No detailed bookkeeping was required; instead, simplified reporting was allowed. Simplified reports allow firms to keep track of daily income, without tickets or invoices for sales less than or equal to $100 Mexican pesos. This regime did not alter social security taxes. Repeco firms pay the ordinary employer/employee contributions for salaried workers registered at the IMSS.

The 2013 fiscal reform repealed and replaced Repeco by the *Régimen de Incorporación Fiscal* (RIF) from January 2014, aiming to modernize tax compliance and bring more small firms into the formal economy. Under RIF, firms pay income taxes based on profit. Introduced by Articles 111-113 of the CPTL, RIF offered new registered firms at SAT with revenue below 2 million pesos a reduction in income taxes declining from 100 percent to 0 percent after 10 years within the regime. RIF firms were exempt from paying VAT if their annual gross income was less than 300,000 Mexican pesos, otherwise they would pay VAT under the normal rules. RIF also considered simplifications like exempting firms from issuing invoices for sales of less than 250 pesos, if not requested by clients, or allowing declarations bimonthly instead of monthly, as well as a package of benefits such as access to low-price life insurance for entrepreneurs. In April 2014, an executive order incentivized new RIF firms to register their workers and contribute to their social security, through a subsidy of 50 percent during the first two years, tapering to 10 percent in year 10.

In 2022, RIF was replaced by the RESICO (*Régimen Simplificado de Confianza*) through Articles 206 to 209 of the CPT Law for legal entities. RIF firms could continue using the regime until they exceed eligibility limits or choose to transition to RESICO. RESICO-eligible firms are those whose annual income does not exceed 35 million Mexican pesos. The regime eliminates deductions and administrative complexity while setting a CPT rate of 30% on profits. VAT computation is the same as in the general regime. RESICO do not consider breaks to social security contributions. Only firms remaining under RIF would keep the corresponding tax benefits.

Appendix B Mexican Census Data

B.1 Variable Construction

Table B.1 indicates how variables are constructed based on 2019 codes.

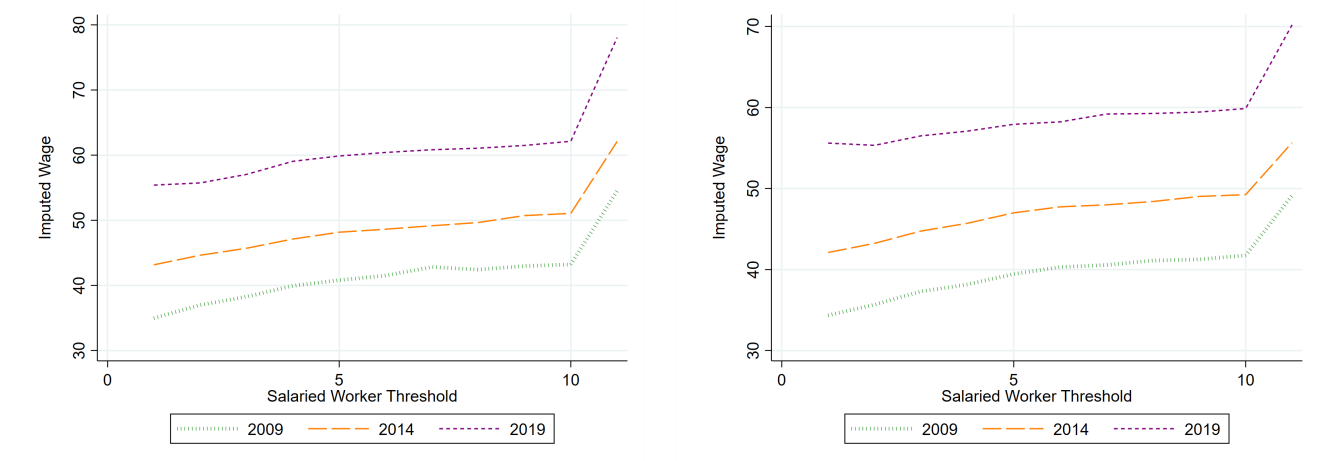
Table B.1: Variable Construction Using 2019 Census Codes

Variable	Description	Variable Construction and Code Variable Description
Rev	Gross Revenue	$Rev = \hat{Rev} + VATax = a111a + m001a$ a111a = Producción Bruta Total m001a = IVA Cobrado
I	Inter. Input Exp.	$I = a121a - k610a - k620a$ a121a = Consumo Intermedio k610a = Pagos a otra razón social que contrató y ... k620a = Gastos por honorarios o comisiones sin sueldo o salario fijo
L1	Total Workers	$L1 = h000a + i000a$ h000a = Total de personal dependiente de la razón social i000a = Total de personal no dependiente de la razón social
L2	Hours Worked	$L2 = h000d + i000d$ h000d = Total ... dependiente de la razón social. Horas trabajadas i000d = Total ... no dependiente de la razón social. Horas trabajadas
L3	Labor Exp.	$L3 = j010a + j203a + k610a + k620a + wage_i \times h020a$
L4	Labor Exp.	$L4 = j010a + j203a + k610a + k620a + wage_i \times (h020a - 1)$ j010a = Total de salarios al personal de producción, ventas y servicios j203a = Total de sueldos al personal administrativo, contable y de dirección h020a = Propietarios, familiares y otros trabajadores no remunerados $wage_i = \frac{\sum_{k \in K_i} (j010a + j203a)_k}{\sum_{k \in K_i} (h101a + h203a)_k}$ industry i compensation per salaried worker K_i = { plants k in industry i: $j010a + j203a > 0$ and $h101a + h203a \leq 10$ } h101a = Personal de producción, ventas y servicios. Total h203a = Personal administrativo, contable y de dirección. Total
VARate	VA Tax Rate	$VARate = VATax/Rev = m001a/Rev$
VARateI	VA Input Rate	$VARateI = VATaxPaid/I = k001a/I$ k001a = IVA Pagado
SSRate	SS Tax Rate	$SSRate = SSTax/Sal = j300a/(j010a + j203a)$ j300a = Contribuciones patronales a regímenes de seguridad social
PTU	PTU	$PTU = j500a$ j500a = Utilidades repartidas a los trabajadores

Note: Next to each code variable we paste the variable description from the Mexican codebook.

Gross revenue (Rev) is measured as inclusive of value-added taxes. Intermediate input expenditure (I) is computed by subtracting the expenditure in outsourced labor $k610a$ and freelancers $k620a$ from $a121a$, which includes both of these labor expenditures. To infer input wedges, revenue-expenditure ratios need to be calculated so that revenue is inclusive of value-added tax and so that labor and intermediate input expenditures are exclusive of taxes - see Problem P1. Figure B.1 shows that the

imputed industry wage rate ($wage_i$ for industry i) per salaried worker increases as the threshold value for the maximum number of salaried workers for a plant increases. This wage rate, calculated at the 6-digit industry level, is used to impute wage payments for non-remunerated workers and construct the labor expenditure measure L3 and L4. In practice, the measure based on 10 workers is used. The labor expenditure measure L4 is the benchmark measure used in the paper.[13] The variable j300a corresponds to the social security tax paid.[14] Table A.1 of Appendix A described the rules governing social security taxes paid for a salaried worker. The variable j500a is the profit sharing variable PTU described in Appendix A.1.



(a) 4-digit Industry (b) 6-digit Industry

Note: Plotted are the median across industries. The rightmost value eliminates the maximum worker threshold so effectively the threshold is set to infinity.

Figure B.1: Imputed Wage By Number of Salaried Workers Threshold

B.2 Constructing Tax Wedges

This section first explains how tax wedges for plant i in census year $t \in \{2009, 2014\}$ are constructed.

Step 1: Compute $(\tau_{s,t}^{Repeco}, \tau_t^{corp}, \tau_t^{payroll})$, where s denotes a state.

$$\tau_{s,t}^{Repeco} = Repeco\ Tax_{s,t} / Repeco\ Revenue_{s,t}$$

$Repeco\ Tax_{s,t}$ is taken from SAT data reported in Fuentes-Castro et al. (2011, Table 3).

$Repeco\ Revenue_{s,t} = \sum_{i \in K_{s,t}} \hat{Rev}_{it}$, where $K_{s,t}$ is the set of plants i located in state s in census year t (in the entire Census) with revenue measure $\hat{Rev}_{it} \leq 2M$.

[13] For plants with h020a =0, the $L4$ is set equal to $L3$ to avoid the posssibility that $L4 < 0$.

[14] Regarding the variable for social security contributions $j300a$, all 2008, 2013, and 2018 questionaries of Mexico's economic census, explicitly include in the question firm payments to "IMSS, INFONAVIT, AFORE and others", previous questionaries were ambiguous about what to consider social security contributions.

$\tau_t^{corp} =$ `Statutory Corporate Profit Tax Rate in census year t`

$\tau_{s,t}^{payroll} =$ `State Payroll Tax Rate in state s and census year t`

Two clarifications are important. First, the census year refers to data for the previous calendar year. Thus, $Repeco\ Tax_{s,t}$ for state s in census year $t = 2004$ is the empirical state tax aggregate collected from Repeco firms over the calendar year 2003. Second, as Fuentes-Castro et al. (2011, Table 3) report tax aggregates for the calendar years 2000- 2010, some assumption is needed for census year 2014. We set $\tau_{s,2014}^{Repeco} = \tau_{s,2009}^{Repeco}$ for all states s. Repeco tax rates are displayed in Figure 3 in the main text. Statutory corporate tax rates are those reported in Table A.7 in Appendix A. State Payroll Tax Rates are reported in Table A.4 in Apendix A.

Step 2: Compute tax rates $(\tau_{it}^{ss}, \tau_{it}^{sp}, \tau_{it}^{va}, \tau_{it}^{va,m}, \tau_{it}^{r}, \tau_{it}^{c})$ for plant i for $t \in \{2009, 2014\}$.

$$\tau_{it}^{ss} = \begin{cases} SSTax_{it}/Sal_{it} & \text{if } Sal_{it} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{B.1}$$

$$\tau_{it}^{sp} = \begin{cases} \tau_{s(i),t}^{payroll} & \text{if } Sal_{it} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{B.2}$$

$$\tau_{it}^{va} = VATax_{it}/Rev_{it} \text{ and } \tau_{it}^{va,m} = VATaxPaid_{it}/I_{it} \tag{B.3}$$

$$\tau_{it}^{r} = \begin{cases} \tau_{s(i),t}^{Repeco} & \text{if } \hat{Rev}_{it} \leq 2M \\ 0 & \text{otherwise} \end{cases} \tag{B.4}$$

$$\tau_{it}^{c} = \begin{cases} \tau_{t}^{corp} & \text{if } \hat{Rev}_{it} > 2M \\ 0 & \text{otherwise} \end{cases} \tag{B.5}$$

The variables $(SSTax_{it}, Sal_{it}, Rev_{it}, \hat{Rev}_{it})$ are defined in Table B.1. The notation $s(i)$ denotes the state in which plant i is located. The notation $2M$ denotes the two million peso threshold defining Repeco plant/firms.

Step 3: Compute $(taxwedge_{1,i,t}, taxwedge_{2,i,t})$ for labor and intermediate input. Use the tax function T below and treat plant tax rates from Step 2 as proxies for plant marginal rates. For plants that are beyond the Repeco threshold, the tax function calculates taxable profit by subtracting salaried wages and social security taxes, which is consistent with Table A.5. The tax function T is a function of a tax-inclusive measure of revenue (Rev), consistent with problem P1 from section 3. Rev was constructed in Table B.1.

$$T = \begin{cases} (\tau_{it}^{ss} + \tau_{it}^{sp})Sal_{it} + \max\{\tau_{it}^{va} Rev_{it}, \tau_{it}^{va,m} I_{it}\} + \tau_{it}^{c}(Rev_{it}(1-\tau_{it}^{va}) - Sal_{it}(1+\tau_{it}^{ss}+\tau_{it}^{sp}) - I_{it}) & \hat{Rev}_{it} > 2M \\ (\tau_{it}^{ss} + \tau_{it}^{sp})Sal_{it} + \max\{\tau_{it}^{va} Rev_{it}, \tau_{it}^{va,m} I_{it}\} + \tau_{it}^{r} Rev_{it}(1-\tau_{it}^{va}) & \text{otherwise} \end{cases} \tag{B.6}$$

In computing tax wedges, cases 1-2 are for plants above the Repeco threshold, whereas cases 3-4 are for plants below the Repeco threshold.

Case 1: $\hat{Rev}_{it} > 2M$ and $\tau_{it}^{va} Rev_{it} \geq \tau_{it}^{va,m} I_{it}$

Case 2: $\hat{Rev}_{it} > 2M$ and $\tau_{it}^{va} Rev_{it} < \tau_{it}^{va,m} I_{it}$

Case 3: $\hat{Rev}_{it} < 2M$ and $\tau_{it}^{va} Rev_{it} \geq \tau_{it}^{va,m} I_{it}$

Case 4: $\hat{Rev}_{it} < 2M$ and $\tau_{it}^{va} Rev_{it} < \tau_{it}^{va,m} I_{it}$

$$taxwedge_{1,i,t} = \frac{1+T_1}{1-T_3} = \begin{cases} \frac{1+[\tau_{it}^{ss}+\tau_{it}^{sp}-\tau_{it}^{c}(1+\tau_{it}^{ss}+\tau_{it}^{sp})]}{1-[\tau_{it}^{va}+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{ss}+\tau_{it}^{sp}}{1-\tau_{it}^{va}} & \text{if case 1} \\ \frac{1+[\tau_{it}^{ss}+\tau_{it}^{sp}-\tau_{it}^{c}(1+\tau_{it}^{ss}+\tau_{it}^{sp})]}{1-[0+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{(1-\tau_{it}^{c})(1+\tau_{it}^{ss}+\tau_{it}^{sp})}{1-\tau_{it}^{c}(1-\tau_{it}^{va})} & \text{if case 2} \\ \frac{1+[\tau_{it}^{ss}+\tau_{it}^{sp}]}{1-[\tau_{it}^{va}+\tau_{it}^{r}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{ss}+\tau_{it}^{sp}}{(1-\tau_{it}^{r})(1-\tau_{it}^{va})} & \text{if case 3} \\ \frac{1+[\tau_{it}^{ss}+\tau_{it}^{sp}]}{1-[0+\tau_{it}^{r}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{ss}+\tau_{it}^{sp}}{1-\tau_{it}^{r}(1-\tau_{it}^{va})} & \text{if case 4} \end{cases} \tag{B.7}$$

$$taxwedge_{2,i,t} = \frac{1+T_2}{1-T_3} = \begin{cases} \frac{1+[-\tau_{it}^{c}]}{1-[\tau_{it}^{va}+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{1}{1-\tau_{it}^{va}} & \text{if case 1} \\ \frac{1+[\tau_{it}^{va,m}-\tau_{it}^{c}]}{1-[0+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{va,m}-\tau_{it}^{c}}{1-\tau_{it}^{c}(1-\tau_{it}^{va})} & \text{if case 2} \\ \frac{1+[0]}{1-[\tau_{it}^{va}+\tau_{it}^{r}(1-\tau_{it}^{va})]} = \frac{1}{(1-\tau_{it}^{r})(1-\tau_{it}^{va})} & \text{if case 3} \\ \frac{1+[\tau_{it}^{va,m}]}{1-[0+\tau_{it}^{r}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{va,m}}{1-\tau_{it}^{r}(1-\tau_{it}^{va})} & \text{if case 4} \end{cases} \tag{B.8}$$

The terms in square brackets above are marginal tax rates T_1 for (salaried) labor expenditure, T_2 for intermediate goods expenditure and T_3 for revenue.

One argument for using salaried worker marginal tax rates is that if a plant is off corners in its choice of salaried and non-salaried labor (i.e. hires both salaried and nonsalaried labor) and views both types as perfect substitutes in production that require the same direct wage, then the unobserved expected marginal tax rate on non-salaried labor must equal the marginal rate on salaried labor if it is minimizing expected cost. This argument views the expected cost of non-salaried labor as including probabilistic penalties. For example, when audited, penalties could be based on the portion of non-salaried labor that is determined to infact be salaried labor.

Tax wedges for the census year $t = 2019$ are constructed differently as in 2019 the RIF regime operated. The first step is to construct plant tax rates $(\tau_{it}^{ss}, \tau_{it}^{sp}, \tau_{it}^{va}, \tau_{it}^{va,m})$ exactly as described earlier. The second step is to construct an imputed profits tax rate τ_{it}^{c}, where a_i is plant i's "age"

Table B.2: Distribution of Tax Wedge Cases by Year

Year	Case 1	Case 2	Case 3	Case 4
2009	5.7	1.3	90.0	3.0
2014	7.3	1.8	84.0	6.9

Note: The table reports the percentage distribution of plants that satisfy Cases 1-4.

(i.e. number of years within the RIF regime).[15] This rate varies with age as RIF allows age-related tax rate reductions. The third step is to pose a tax function T and calculate tax wedges based on T. For $t = 2019$, two cases are relevant Case 1: $\tau_{it}^{va} Rev_{it} \geq \tau_{it}^{va,m} I_{it}$ and Case 2: $\tau_{it}^{va} Rev_{it} < \tau_{it}^{va,m} I_{it}$. Empirically, 92.4 percent of plants fall under Case 1.

$$\tau_{it}^{c} = \begin{cases} \tau_t^{corp} \times \dfrac{a_i}{10} & \text{if } 0 \leq a_i \leq 10 \ \& \ \hat{Rev}_{it} \leq 2M \\ \tau_t^{corp} & \text{if } a_i > 10 \ \& \ \hat{Rev}_{it} \leq 2M \\ \tau_t^{corp} & \hat{Rev}_{it} > 2M \end{cases} \tag{B.9}$$

$$T = (\tau_{it}^{ss} + \tau_{it}^{sp}) Sal_{it} + \max\{\tau_{it}^{va} Rev_{it}, \tau_{it}^{va,m} I_{it}\} + \tau_{it}^{c} (Rev_{it}(1-\tau_{it}^{va}) - Sal_{it}(1+\tau_{it}^{ss}+\tau_{it}^{sp}) - I_{it} - \cdots) \tag{B.10}$$

$$taxwedge_{1,i,t} = \frac{1+T_1}{1-T_3} = \begin{cases} \frac{1+[\tau_{it}^{ss}+\tau_{it}^{sp}-\tau_{it}^{c}(1+\tau_{it}^{ss}+\tau_{it}^{sp})]}{1-[\tau_{it}^{va}+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{ss}+\tau_{it}^{sp}}{1-\tau_{it}^{va}} & \text{if case 1} \\ \frac{1+[\tau_{it}^{ss}+\tau_{it}^{sp}-\tau_{it}^{c}(1+\tau_{it}^{ss}+\tau_{it}^{sp})]}{1-[0+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{(1-\tau_{it}^{c})(1+\tau_{it}^{ss}+\tau_{it}^{sp})}{1-\tau_{it}^{c}(1-\tau_{it}^{va})} & \text{if case 2} \end{cases} \tag{B.11}$$

$$taxwedge_{2,i,t} = \frac{1+T_2}{1-T_3} = \begin{cases} \frac{1+[-\tau_{it}^{c}]}{1-[\tau_{it}^{va}+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{1}{1-\tau_{it}^{va}} & \text{if case 1} \\ \frac{1+[\tau_{it}^{va,m}-\tau_{it}^{c}]}{1-[0+\tau_{it}^{c}(1-\tau_{it}^{va})]} = \frac{1+\tau_{it}^{va,m}-\tau_{it}^{c}}{1-\tau_{it}^{c}(1-\tau_{it}^{va})} & \text{if case 2} \end{cases} \tag{B.12}$$

B.3 Constructing Cost Shares

Table B.3 indicates how cost shares are constructed based on 2019 codes, whereas Table B.4 reports average cost shares across industries.

B.4 Estimating Output Elasticities

This section describes and implements two approaches to estimate plant-specific output elasticities.

[15] For the 2019 census year, $a_i = 2018 - 2014$ for plants i born on or before 2014 (i.e. $g111a \leq 2014$) and $a_i = 2018 - g111a$ otherwise.

Table B.3: Variable Construction Using 2019 Census Codes

Variable	Description	Variable Construction and Code Variable Description
K	Capital	K = fixed_asset + rented_capital/0.1 = q000a + k050a/0.1 q000a = Activos fijos k050a = Renta y alquiler de bienes muebles e inmuebles
Kcost	Cost of Capital	Kcost = K*0.1
I. Firm-level Cost Shares		
		total_factor_cost = L4 + I + Kcost
	Labor Cost Share	L4_cost_share = L4 / total_factor_cost
	Capital Cost Share	K_cost_share = Kcost / total_factor_cost
	Inter. Inputs Share	I_cost_share = I / total_factor_cost
II. Industry-level Cost Shares		
		$K_i = \{k : \text{plant } k \text{ is in industry i }\}$
	Labor Cost Share	$L4_Share_i = \frac{\sum_{k \in K_i}(L4)_k}{\sum_{k \in K_i}(L4+I+Kcost)_k}$
	Inter. Inputs Share	$I_Share_i = \frac{\sum_{k \in K_i}(I)_k}{\sum_{k \in K_i}(L4+I+Kcost)_k}$
	Capital Cost Share	$K_Share_i = \frac{\sum_{k \in K_i}(Kcost)_k}{\sum_{k \in K_i}(L4+I+Kcost)_k}$

Note: Fixed assets ($q000a$) and rented capital ($k050a$) are in thousands of Mexican pesos. Next to each variable code we paste the variable description from the Mexican codebook. Some variables (I and $L4$) were previously defined in Table B.1 and are in units of thousands of Mexican pesos.

Table B.4: Average Factor Cost Shares

year	mean K_share	mean L4_share	mean I_share
2009	0.066765	0.2182	0.715036
2014	0.070258	0.206626	0.723116
2019	0.061085	0.19332	0.745595

Note: Factor cost shares are first calculated by industry and then averaged across industries.

B.4.1 Cost Share Approach

$M1$ $\min_{(x_1,...,x_J)} \sum_{j=1}^{J} w_j x_j + T(w_1 x_1, ..., w_J x_J, py)$ *s.t.* $y = F(x_1, ..., x_J)$

$M2$ $\min_{(x_1,...,x_{J-1})} \sum_{j=1}^{J-1} w_j x_j + T(w_1 x_1, ..., w_J x_J, py)$ *s.t.* $y = F(x_1, ..., x_J)$ *and* x_J is preset.

If a plant solves the profit maximization Problem P1 from the main text, then the same inputs minimize the cost of producing the chosen output level y. Problem M1 and M2 differ in that a capital input (i.e. input x_J) may be set before knowing the production function F, due to a time-to-build friction. Proposition B.1 says that the output elasticity $\alpha_j(x) \equiv \frac{\partial F(x)}{\partial x_j}\frac{x_j}{y}$ equals the appropriate cost share of input j at a solution to M1 and it equals the product of the cost share of input j, among variable inputs, and the sum of the output elasticities of all variable inputs at a solution to M2. A strength of cost share approaches is that the output elasticity result holds regardless of whether technology, factor prices and tax functions are common or plant specific. Thus, it is robust to many types of plant heterogeneity. In Proposition B.1, $\Delta_s E[f|s] = E[f|s] - E[f|s=0]$ for any function f and $\Delta_j \ln wedge_j = \ln wedge_j - \ln wedge_1$.

Proposition B.1. *Assume F is differentiable and constant returns to scale in $x = (x_1, ..., x_J)$.*

(i) If $(x_1, ..., x_J)$ solves M1, then $\alpha_j(x) = \frac{w_j x_j (1+T_j)}{\sum_{j=1}^{J} w_j x_j (1+T_j)}$ for $j = 1, ..., J$.

(ii) If $(x_1, ..., x_{J-1})$ solves M2, then

(a) $\alpha_j(x) = \frac{w_j x_j (1+T_j)}{\sum_{j=1}^{J-1} w_j x_j (1+T_j)}(1 - \alpha_J(x))$ for $j = 1, ..., J-1$.

(b) $\ln wedge_j \equiv \ln \frac{Rev}{w_j x_j}\alpha_j(x) = \ln \frac{Rev}{w_j x_j}(\frac{w_j x_j (1+T_j)}{\sum_{j=1}^{J-1} w_j x_j (1+T_j)}(1 - \alpha_J(x)))$ for $j = 1, ..., J-1$.

(c) $\Delta_s E[\ln wedge_j|s] = \Delta_s E[\ln \frac{Rev(1-T_{J+1})(1-\alpha_J)}{\sum_{j=1}^{J-1} w_j x_j (1+T_j)}|s] + \Delta_s E[\ln taxwedge_j|s]$ for $j = 1, ..., J-1$.

(d) $\Delta_j \ln wedge_j = \Delta_j \ln taxwedge_j$ for $j = 2, ..., J-1$.

Proof: (i) The first equation is a first-order condition to a solution to M1, where λ is a Langrage multiplier and T_j is the partial derivative of the tax function with respect to expenditure on input j. The leftmost equality in the second equation is Euler's Theorem for homogeneous functions and constant returns to scale, whereas the rightmost equality holds by the first equation. The conclusion follows by the definition of the output elasticity $\alpha_j(x)$ and by substitution.

$\frac{\partial F(x)}{\partial x_j} = \frac{w_j(1+T_j)}{\lambda}$ for $j = 1, ..., J$

$1 = \sum_{j=1}^{J} \frac{\partial F(x)}{\partial x_j}\frac{x_j}{y} = \sum_{j=1}^{J} \frac{w_j x_j (1+T_j)}{\lambda y}$

$\alpha_j(x) \equiv \frac{\partial F(x)}{\partial x_j}\frac{x_j}{y} = \frac{w_j x_j (1+T_j)}{\lambda y} = \frac{w_j x_j (1+T_j)}{\sum_{j=1}^{J} w_j x_j (1+T_j)}$

(ii) Part a. Repeat the steps in the proof of Prop. B.1(i).

$\frac{\partial F(x)}{\partial x_j} = \frac{w_j(1+T_j)}{\lambda}$ for $j = 1, ..., J-1$

$1 - \alpha_J(x) = \sum_{j=1}^{J-1} \alpha_j(x) = \sum_{j=1}^{J-1} \frac{\partial F(x)}{\partial x_j}\frac{x_j}{y} = \sum_{j=1}^{J-1} \frac{w_j x_j(1+T_j)}{\lambda y}$

$\alpha_j(x) \equiv \frac{\partial F(x)}{\partial x_j}\frac{x_j}{y} = \frac{w_j x_j(1+T_j)}{\lambda y} = \frac{w_j x_j(1+T_j)}{\sum_{j=1}^{J-1} w_j x_j(1+T_j)}(1 - \alpha_J(x))$ for $j = 1, ..., J-1$

Part b. Use definition of $wedge_j$ and apply Part a.

Part c. and Part d. Result is implied by Part b. ‖

Remark:

1. Prop B.1 also implies that $\ln wedge_j = \ln taxwedge_j + \ln markup$, where $\ln markup = \ln \frac{Rev(1-T_{J+1})(1-\alpha_J)}{\sum_{j=1}^{J-1} w_j x_j(1+T_j)}$. Thus, the markup component from the cost-share approach is the same for both inputs as is true in a solution to P1.

To implement the cost share approach using Mexican data, proceed as follows. First, assume that capital is preset or quasi fixed as is commonly assumed in the applied literature. Thus, $\alpha_j(x) = \frac{w_j x_j(1+T_j)}{\sum_{j=1}^{2} w_j x_j(1+T_j)}(1 - \alpha_3(x))$ and labor and intermediate input are variable inputs. Second, set labor cost $w_1 x_1(1+T_1)$ and intermediate input cost $w_2 x_2(1+T_2)$ using empirical expenditure measures and marginal rates. Third, assume that $\alpha_3(x) = \alpha_K$ is common within an industry and equals the average industry capital cost share. Absent some restriction, the system of equations is singular and cannot be usefully solved.

Implementation: let i denote a plant from a specific industry

1. $\alpha_j^i = \frac{(w_j x_j(1+T_j))_i}{\sum_{j=1}^{2}(w_j x_j(1+T_j))_i}(1 - \alpha_3^i)$
2. Set $\alpha_3^i = K_share$ for each industry based on Appendix B.3
3. Set $(w_1 x_1(1+T_1))_i = L4_i(1+T_1)_i$ and $(w_2 x_2(1+T_2))_i = I_i(1+T_2)_i$, where $(L4_i, I_i)$ are defined in Appendix B.1
4. Marginal rates are constructed in Appendix B.2. For year $t = 2009, 2014$, they are restated below. For $t = 2019$ see Appendix B.2.

$$(1+T_1)_i = \begin{cases} 1 + [\tau_{it}^{ss} + \tau_{it}^{sp} - \tau_{it}^{c}(1 + \tau_{it}^{ss} + \tau_{it}^{sp})] & \text{if case 1: } \hat{Rev}_{it} > 2M \text{ and } \tau_{it}^{va} Rev_{it} \geq \tau_{it}^{va,m} I_{it} \\ 1 + [\tau_{it}^{ss} + \tau_{it}^{sp} - \tau_{it}^{c}(1 + \tau_{it}^{ss} + \tau_{it}^{sp})] & \text{if case 2: } \hat{Rev}_{it} > 2M \text{ and } \tau_{it}^{va} Rev_{it} < \tau_{it}^{va,m} I_{it} \\ 1 + [\tau_{it}^{ss} + \tau_{it}^{sp}] & \text{if case 3: } \hat{Rev}_{it} < 2M \text{ and } \tau_{it}^{va} Rev_{it} \geq \tau_{it}^{va,m} I_{it} \\ 1 + [\tau_{it}^{ss} + \tau_{it}^{sp}] & \text{if case 4: } \hat{Rev}_{it} < 2M \text{ and } \tau_{it}^{va} Rev_{it} < \tau_{it}^{va,m} I_{it} \end{cases} \quad \text{(B.13)}$$

$$(1+T_2)_i = \begin{cases} 1+[-\tau_{it}^{c}] & \text{if case 1} \\ 1+[\tau_{it}^{va,m}-\tau_{it}^{c}] & \text{if case 2} \\ 1+[0] & \text{if case 3} \\ 1+[\tau_{it}^{va,m}] & \text{if case 4} \end{cases} \tag{B.14}$$

B.4.2 Estimate a Translog Production Function

Consider the translog production function which generalizes the Cobb-Douglas specification:

$$\ln y = \ln z_1 + \sum_{j=1}^{J} \beta_j \ln x_j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk} \ln x_j \ln x_k \tag{B.15}$$

The elasticity of output with respect to input j is:

$$\alpha_j(x) \equiv \frac{\partial \ln y}{\partial \ln x_j} = \beta_j + \sum_{k=1}^{J} \beta_{jk} \ln x_k \tag{B.16}$$

Estimation method Assume the output of firm (establishment) i at time t is such that

$$y_{it} = X_{it}'\beta + \omega_{it} + \varepsilon_{it},$$

where y_{it} is the log of output and X_{it} is a vector containing the log of labor, log of intermediate inputs, and log of capital, along with their interactions as specified by a translog function.[16] The term ω_{it} is the firm's unobserved productivity, which is assumed to follow an AR(1) process: $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$, where ξ_{it} is the innovation to productivity. β and ρ are parameters to be estimated, and ε_{it} is an i.i.d. shock unanticipated by the firm or measurement error.

We follow the estimation method in Ackerberg, Caves, and Frazer (2015), and assume capital is pre-determined whereas labor and intermediate inputs are freely chosen after observing the current productivity ω_{it}. The estimation proceeds in two stages. In the first stage, we estimate a non-parametric function of all inputs to obtain a consistent estimate of the unanticipated shock ε_{it}. This involves regressing output on a flexible polynomial (for which we set to be a full third-order polynomial) of all inputs,

$$y_{it} = \Phi(k_{it}, l_{it}, m_{it}) + \varepsilon_{it},$$

where $\Phi(\cdot)$ subsumes the production function itself and the productivity term, such that $\Phi(X_{it}) = X_{it}'\beta + \omega_{it}$. The predicted value from this regression, $\hat{\Phi}_{it}$, provides a "cleaned" measure of output net of the idiosyncratic error while the residual provides a consistent estimate of ε_{it}.

[16] We use Busso, Fentanes, and Levy (2018) concordance tables to identify establishments in the 1999, and 2004 census waves and be able to link the five waves and conduct our longitudinal data analysis.

In the second stage, we use the moment condition that the productivity innovation ξ_{it} is unpredictable given past information,

$$\mathbb{E}[\xi_{it}|I_{it-1}] = 0.$$

Since ε_{it} is also uncorrelated with past information, this implies the full moment condition $\mathbb{E}[\varepsilon_{it} + \xi_{it}|I_{it-1}] = 0$. From here, we can express the productivity innovation in terms of observables and parameters. For any given parameter vector β, productivity can be computed from the first-stage results as $\hat{\omega}_{it}(\beta) = \hat{\Phi}_{it} - X'_{it}\beta$. The innovation is then the residual from the AR(1) process, $\hat{\xi}_{it}(\beta) = \hat{\omega}_{it}(\beta) - \rho\hat{\omega}_{it-1}(\beta)$, in which ρ is fitted based on the sample given the value of β. Let Z_{it-1} be a vector of valid instruments from the information set I_{it-1}, which include log current capital, log lagged variable inputs and their interactions. The GMM estimator finds the parameter vector $\hat{\beta}$ that minimizes the sample of the moment conditions by solving a minimization problem. The number of moment conditions equals the number of parameters and the parameters are exactly identified. Practically, we use the inverse of the sample covariance of the instrument variables to weight each moment condition residual, which stabilizes the minimizer. That is, we solve

$$\min_{\beta} \Big(\frac{1}{N}Z'\xi(\beta)\Big)' \big(\frac{1}{N}Z'Z\big)^{-1} \Big(\frac{1}{N}Z'\xi(\beta)\Big),$$

where N is the number of samples; Z is the instrument variable data matrix which is of size $N \times N_Z$, where N_Z is the number of instrument variables; $\xi(\beta)$ is the the AR(1) residual data matrix which is of size $N \times 1$. All optimization problems are solved with a residual smaller than 1e-6.

Data and Measurement We conduct this estimation for each 3-digit industry, pooling all years. Since we do not have an exact panel, but only observations in 1999, 2004, 2009, 2014, 2019—with 5 year gaps—, our $t-1$ and t are actually 5-years apart.

We calculate output as firm revenue divided by the output deflator at the 3-digit industry level for the corresponding year. We calculate labor inputs as labor expenditures ($L3$ or $L4$) divided by the Consumer Price Index (CPI) for the corresponding year. We calculate intermediate inputs as intermediate inputs expenditures (I) divided by the *input* deflator at the 3-digit industry level for the corresponding year. We calculate capital as the nominal value of capital stock (defined in Table B.3) divided by the capital price for the corresponding year. These procedures require the corresponding deflators for all sample years.[17] For capital price and CPI, we have observations for all sample years. For the output deflator, we have observations at the 3-digit level for 2009, 2014 and 2019, and use the 2-digit level (manufacturing) inflation to construct deflators for 1999 and 2004. For the input deflator, we have observations at the 3-digit level for 2009, 2014 and 2019, and use the *output* inflation of the manufacturing sector to construct deflators for 1999 and 2004 since input deflators for the manufacturing sector are not available for these years.

[17]We take care of the fact that the actual year referred to is one year ahead of the survey year. For example, for the survey 2019, we use the deflators for 2018.

B.4.3 Estimation result

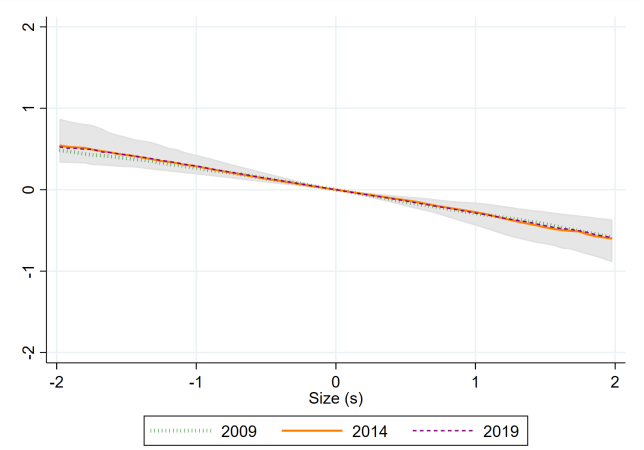
Table B.5: Summary Statistics of Output Elasticity

Cobb Douglas									
	Labor				Intermediate Inputs				Sum elas
Industry	p10	p50	p90	% < 0	p10	p50	p90	% < 0	p50
311	0.27	0.27	0.27	0.00	0.80	0.80	0.80	0.00	1.10
312	0.17	0.17	0.17	0.00	0.83	0.83	0.83	0.00	1.02
314	0.20	0.20	0.20	0.00	0.80	0.80	0.80	0.00	1.06
315	0.37	0.37	0.37	0.00	0.71	0.71	0.71	0.00	1.09
316	0.28	0.28	0.28	0.00	0.70	0.70	0.70	0.00	1.03
321	0.10	0.10	0.10	0.00	0.92	0.92	0.92	0.00	1.02
323	-0.07	-0.07	-0.07	1.00	1.07	1.07	1.07	0.00	0.97
327	0.04	0.04	0.04	0.00	0.93	0.93	0.93	0.00	0.99
332	0.19	0.19	0.19	0.00	0.86	0.86	0.86	0.00	1.08
337	0.14	0.14	0.14	0.00	0.88	0.88	0.88	0.00	1.04
Translog									
	Labor				Intermediate Inputs				Sum elas
Industry	p10	p50	p90	% < 0	p10	p50	p90	% < 0	p50
311	0.28	0.33	0.38	0.00	0.76	0.79	0.85	0.00	1.16
312	0.20	0.26	0.33	0.00	0.65	0.75	0.85	0.00	1.06
314	0.26	0.35	0.47	0.00	0.71	0.79	0.89	0.00	1.23
315	0.27	0.50	0.81	0.01	0.42	0.64	0.83	0.00	1.21
316	-0.19	0.23	0.91	0.22	0.16	0.76	1.16	0.06	1.06
321	0.24	0.30	0.36	0.00	0.79	0.84	0.89	0.00	1.16
323	0.18	0.26	0.33	0.00	0.78	0.86	0.94	0.00	1.15
327	0.10	0.17	0.21	0.00	0.85	0.89	0.97	0.00	1.09
332	0.25	0.31	0.37	0.00	0.80	0.85	0.89	0.00	1.19
337	0.19	0.26	0.34	0.00	0.77	0.88	0.92	0.00	1.15

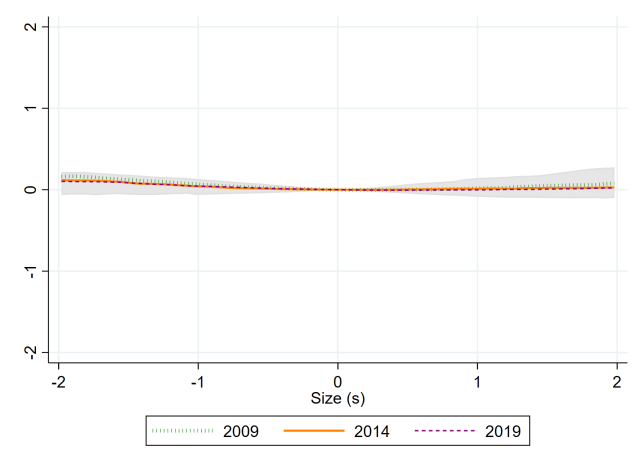
Note: Output elasticities by establishments are constructed based on the production function estimation by industry. Only industries with full panel (5 waves) are kept. % < 0 column means fraction of negative elasticities for labor or intermediate inputs.

Table B.5 tabulates the output elasticities of labor, intermediate inputs, capital and their sum, as implied by the estimated production function by 3-digit industry. The percentiles are first constructed across establishments within each 3-digit industry reporting establishments in all Census (1999 to 2019), and the max, median and min industry-level statistics are reported. As shown, the implied labor and intermediate inputs elasticities are positive for most establishments, and the median sum of elasticities is around 1.1, implying an economy of scale slightly greater than than 1.

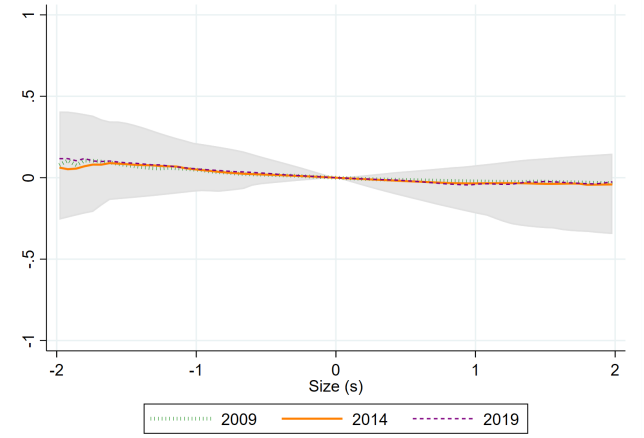
Figure B.2 plots size prfiles of output elasticities based on the translog and cost share methods. The results show that output elasticities of labor fall with size for both methods, whereas the output elasticities of the intermediate input show very little size variation. Figure B.3 examines size profiles



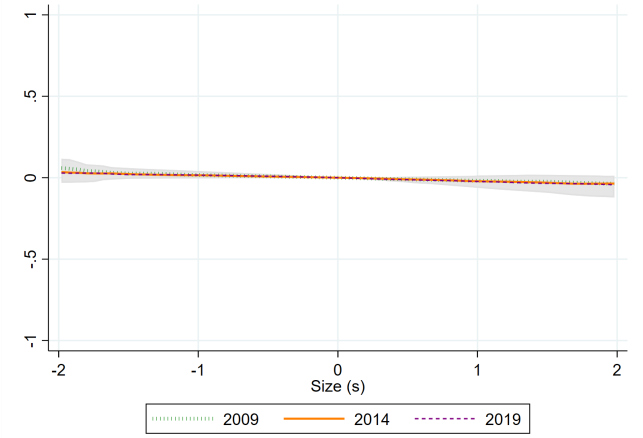
(a) Labor Output Elasticity, Cost Share



(b) Intermediate Input Output Elasticity, Cost Share



(c) Labor Output Elasticity, Translog

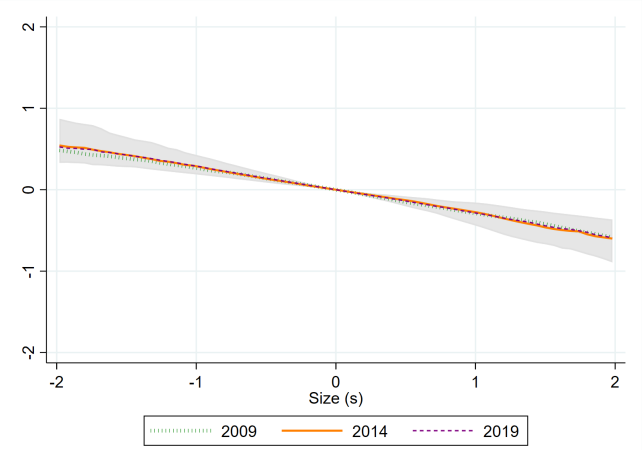


(d) Intermediate Input Output Elasticity, Translog

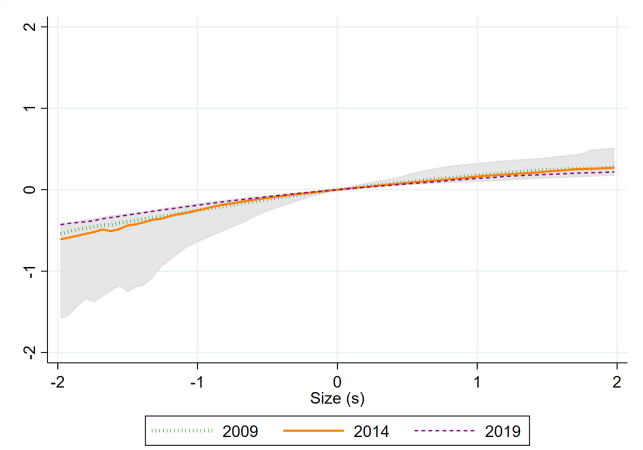
Note: The plotted line shows the median value of $E[\ln \alpha_j|s] - E[\ln \alpha_j|s = 0]$ across industries in a given year. Size s is measured by demeaned and normalized log intermediate input expenditure for the left panel and demeaned and normalized log labor expenditure for the right panel. The top two panels are elasticities based on the cost share approach; the bottom two panels are elasticities implied by estimated translog production functions. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Conditional mean profiles are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth.

Figure B.2: Output Elasticity - Size Profiles

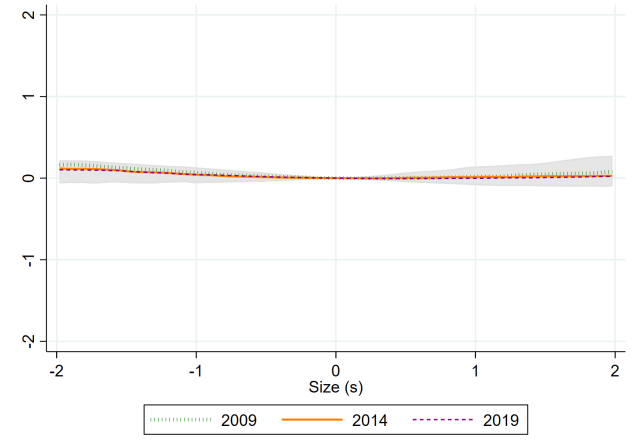
using different size measures.



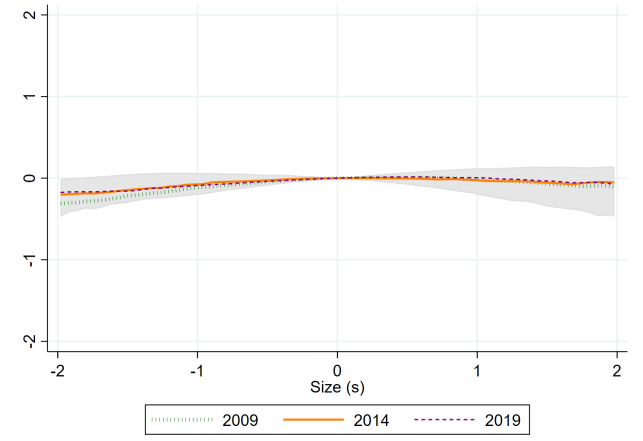
(a) Labor Output Elasticity, Over ln ii



(b) Intermediate Input Output Elasticity, Over ln ii



(c) Labor Output Elasticity, Over ln L4



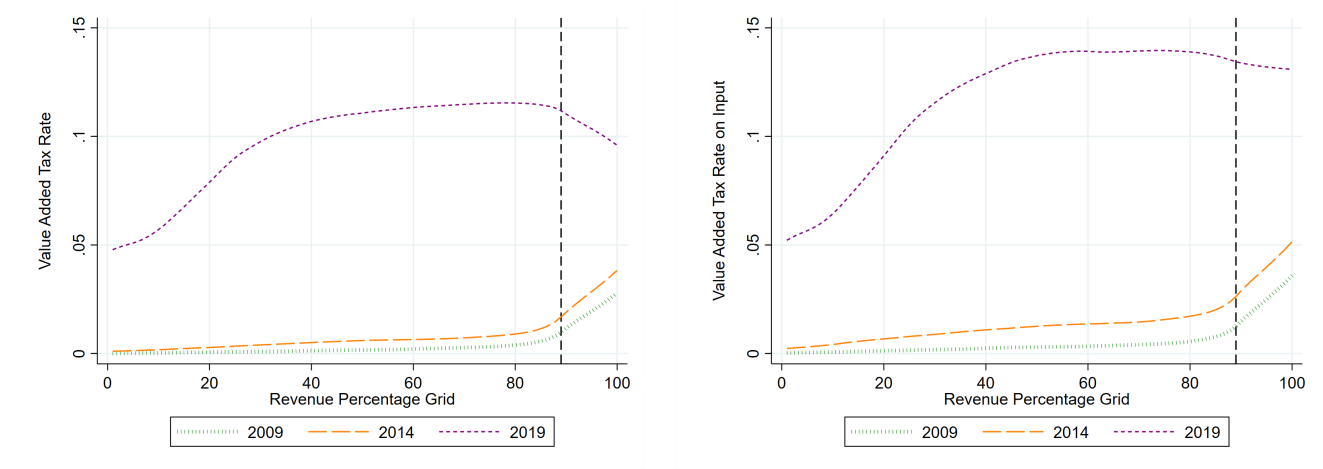
(d) Intermediate Input Output Elasticity, Over ln L4

Note: The plotted line shows the median value of $E[\ln \alpha_j|s] - E[\ln \alpha_j|s = 0]$ across industries in a given year. Size s is measured by demeaned and normalized log intermediate input expenditure for the upper panel and demeaned and normalized log labor expenditure for the lower panel. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Conditional mean profiles are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth.

Figure B.3: Output Elasticity, Cost Share Approach - **Projected Over the Same Size Measure**

B.5 Properties of Tax Rates, Tax Wedges and Wedges

Figure B.4 plots value added tax rates on output and on inputs by revenue bins. Figure B.5 plots the mean value of tax wedges for different revenue bins. Figure B.6 and B.7 plot the variance of wedges conditioned on size.

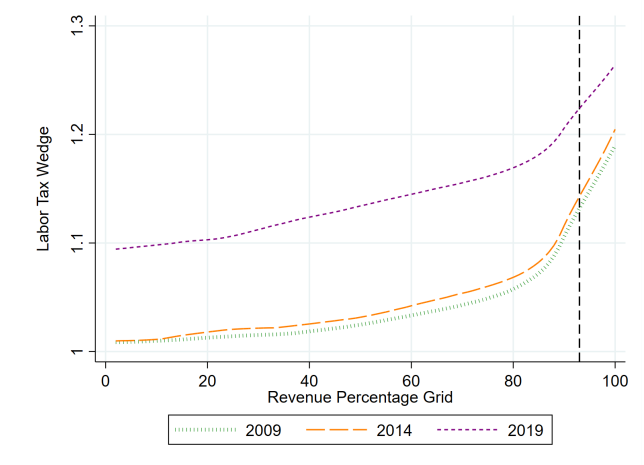


(a) Output Tax Rate (b) Input Tax Rate

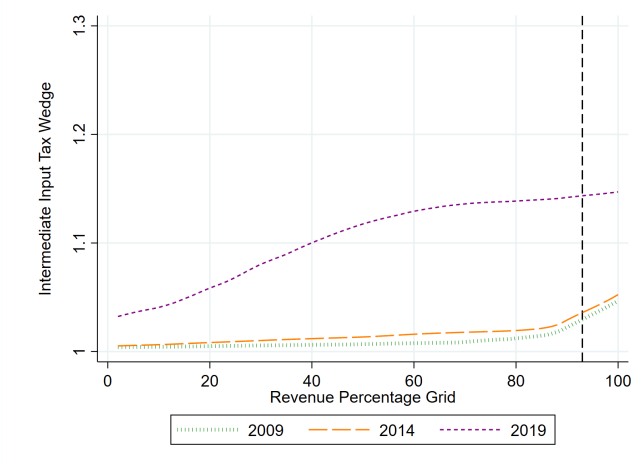
Note: Plotted are the value added tax rates on output and input. Rates are averages of taxes paid divided by revenue and input expenditure, respectively, for each revenue bin.

Figure B.4: Value Added Tax Rates by Size

Table B.6 documents the extent to which the labor wedge gradient exceeds the labor tax wedge gradient across industries. Figure B.8 plots how the log ratio of labor expenditure to intermediate input expenditure varies with size, where size is measured by normalized log revenue. The ratio decreases robustly with size within industries and is a key fact driving the decrease in the output elasticity of labor inferred from the cost-share approach.



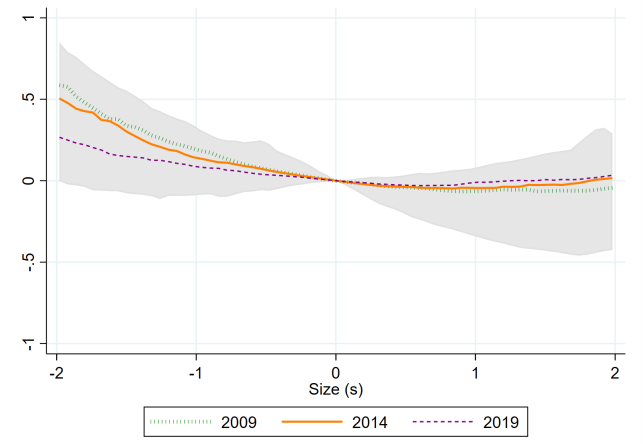
(a) Labor Tax Wedge



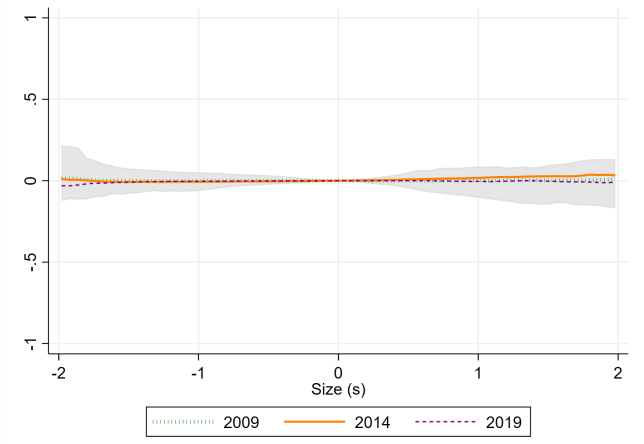
(b) Intermediate Input Tax Wedge

Note: Average wedges within a revenue bin are plotted. The calculation of labor tax wedges and intermediate input tax wedges is explained in Appendix B.

Figure B.5: Tax Wedges By Size



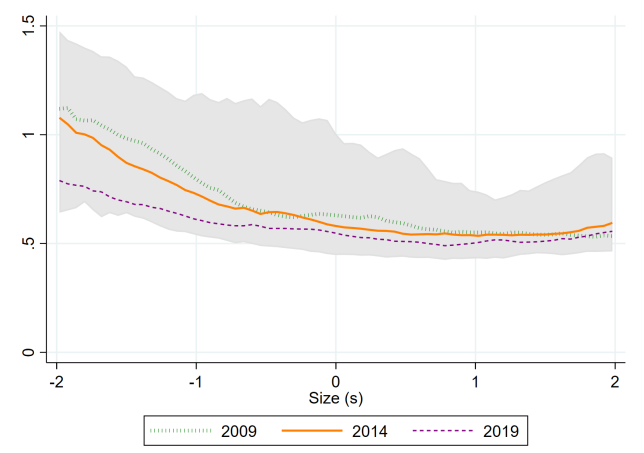
(a) Labor Wedge



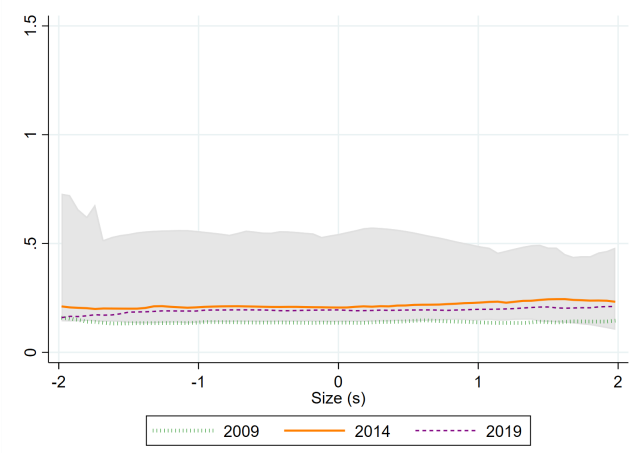
(b) Intermediate Input Wedge

Note: The vertical axis plots $Var[\ln wedge_j|s] - Var[\ln wedge_j|s=0]$. Size s is measured by (demeaned) log intermediate input expenditure, normalized to standard deviations from the industry mean, for the left panel and demeaned and normalized log labor expenditure for the right panel. $Var[\ln wedge_j|s]$ is constructed from $E[(\ln wedge_j)^2|s] - (E[\ln wedge_j|s])^2$, where $\ln wedge_j$ is the log wedge under the constant output elasticity assumption. The plotted line shows the median value of $Var[\ln wedge_j|s] - Var[\ln wedge_j|s=0]$ across industries in a given year. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Industry conditional moments are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth.

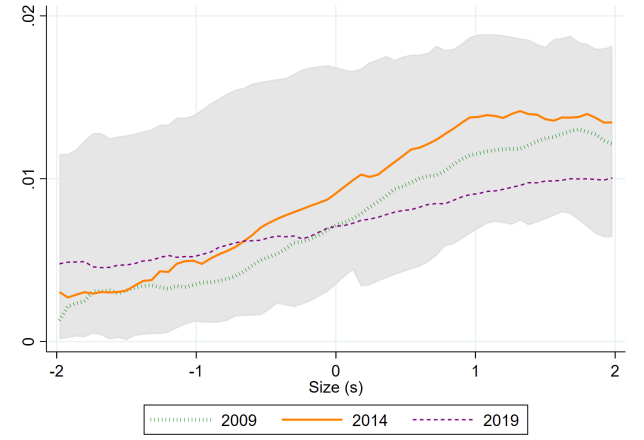
Figure B.6: Variance of Wedge- Size Profiles I



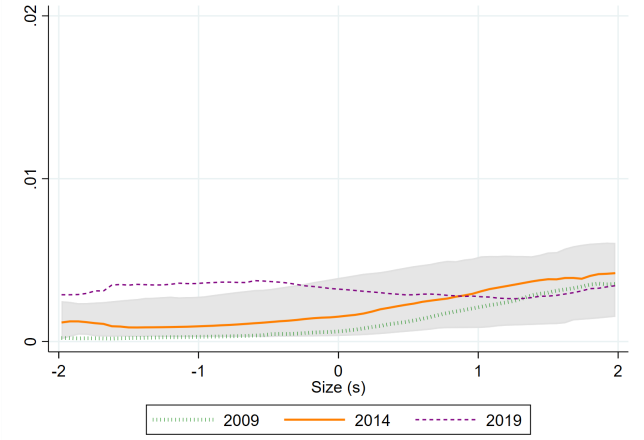
(a) Labor Wedge



(b) Intermediate Input Wedge



(c) Labor Tax Wedge



(b) Intermediate Input Tax Wedge

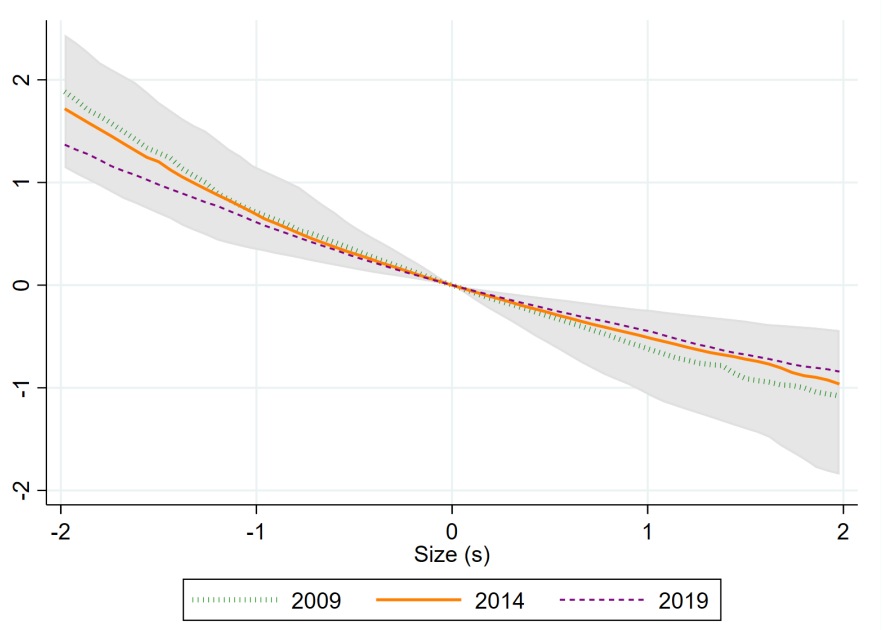
Note: The vertical axis plots $Var[\ln wedge_j|s]$ (top panels) or $Var[\ln taxwedge_j|s]$ (bottom panels). Size s is measured by (demeaned) log intermediate input expenditure, normalized to standard deviations from the industry mean, for the left two panels and demeaned and normalized log labor expenditure for the right two panels. $Var[\ln wedge_j|s]$ is constructed from $E[(\ln wedge_j)^2|s] - (E[\ln wedge_j|s])^2$, where $\ln wedge_j$ is the log wedge under the constant output elasticity assumption. $Var[\ln taxwedge_j|s]$ is constructed from $E[(\ln taxwedge_j)^2|s] - (E[\ln taxwedge_j|s])^2$ The plotted line shows the median value across industries in a given year. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Industry conditional moments are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth.

Figure B.7: Variance of Wedge- Size Profiles II

Table B.6: Fraction of Industries: Wedge Gradient Exceeds Tax Wedge Gradient

Method	2009	2014	2019
CD	0.993	0.997	0.991
TL	1.000	0.995	0.977
CS	0.926	0.634	0.432

Note: Each entry reports the average fraction of industries for which the labor wedge gradient exceeds the labor tax wedge gradient. Define the gradient of a conditional mean profile at size $s \neq 0$ as $(E[\ln w|s]-E[\ln w|s=0])/s$ where w is the wedge or tax wedge. For each industry i, year t and method, we evaluate whether the gradient difference $\Delta_{i,s,t} = (E[\ln wedge_1|s] - E[\ln wedge_1|s = 0])/s - (E[\ln taxwedge_1|s] - E[\ln taxwedge_1|s = 0])/s$ is positive at size levels $s = -2, -1, 1, 2$, compute the fraction of positive cases across these size levels, and report the average of this fraction across all industries. Size s is normalized log intermediate input expenditure. Conditional means are estimated nonparametrically with an Epanechnikov kernel and 0.5 bandwidth. Labor wedges are calculated using the Cobb-Douglas (CD), translog (TL) and cost share (CS) methods.



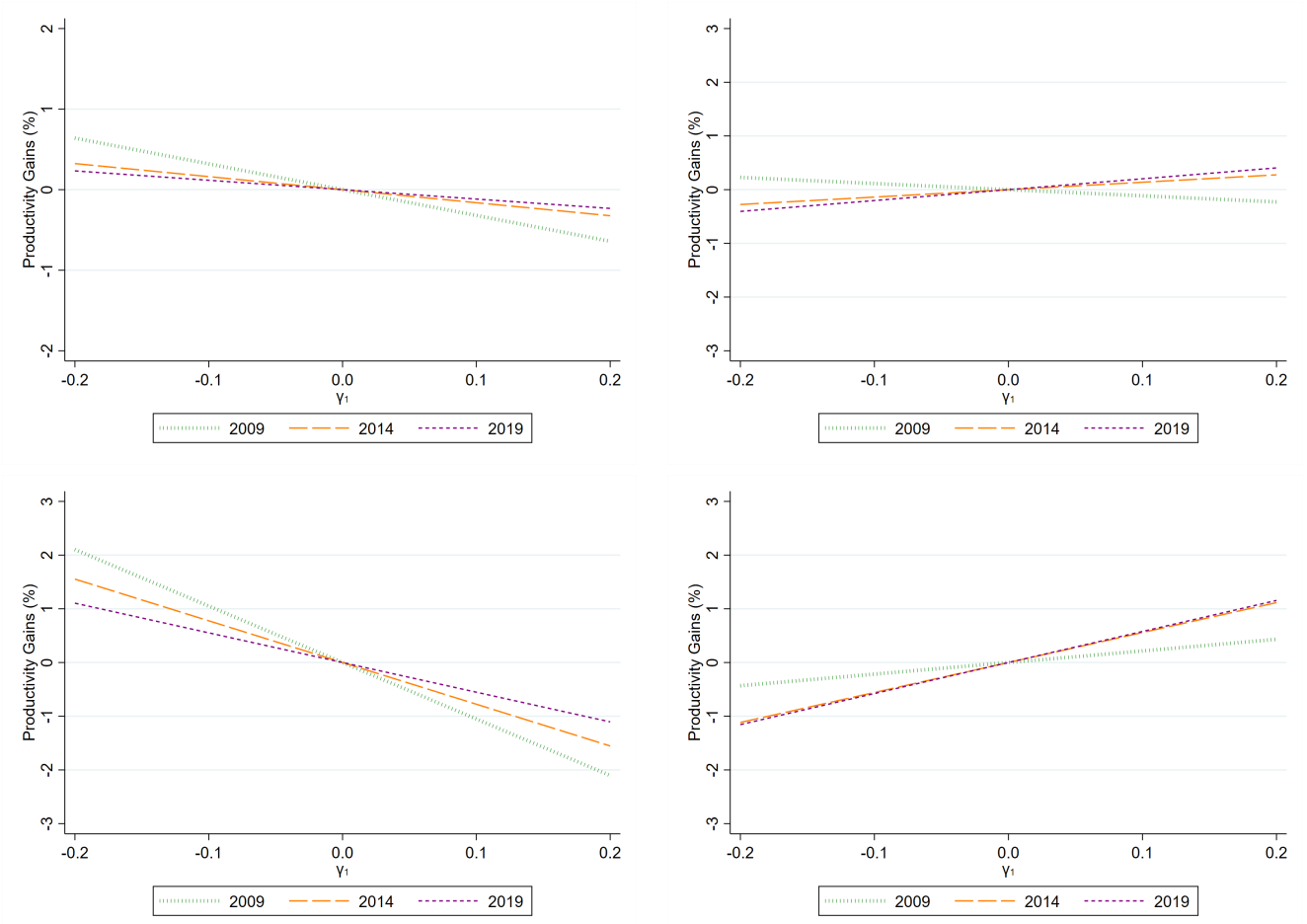
Note: The vertical axis plots $E[\ln(w_1x_1/w_2x_2)|s] - E[\ln(w_1x_1/w_2x_2)|s = 0]$, where w_1x_1 is labor expenditure and w_2x_2 is intermediate input expenditure as measured in Table B.1. Size s is measured by log revenue, demeaned and normalized to standard deviations from the industry mean. The plotted line shows the median value across industries in a given year. The shaded area represents the 90th to 10th percentile range across industries for year 2014. Conditional mean profiles are estimated nonparametrically using an Epanechnikov kernel and a 0.5 bandwidth.

Figure B.8: Log Labor-to-Intermediate-Input Expenditure Ratio by Size

B.6 Reallocation Experiments

We calculate and decompose the reallocation gains (losses), presented in section 5 of the main text, using the (labor reallocation) formula in Prop. C.2.

$$\Omega'(0) = \gamma_1 \underbrace{\frac{\sum_i w_1 x_{1i}}{\sum_i (Rev_i - w_2 x_{2i})}}_{\text{part 1}} \underbrace{\left(\frac{\sum_{i \in N_L} x_{1i}}{\sum_{i \in N_H} x_{1i}}\right)}_{\text{part 2}} \underbrace{\left[\frac{1}{n_L} \sum_{i \in N_L} \left(\frac{x_{1i}}{E[x_1|L]}\right) wedge_{1i} - \frac{1}{n_H} \sum_{i \in N_H} \left(\frac{x_{1i}}{E[x_1|H]}\right) wedge_{1i}\right]}_{\text{part 3}} \tag{B.17}$$



(a) Reallocating Labor

(b) Reallocating Intermediate Input

Figure B.9: Marginal Productivity Gains from Input Reallocation

Note: Plotted are the median reallocation gains across industries for each year. For the top two panels, H and L groups are based on whether labor expenditure exceeds the industry mean. For the bottom two panels, H and L groups are based on whether intermediate input expenditure exceeds the industry mean.

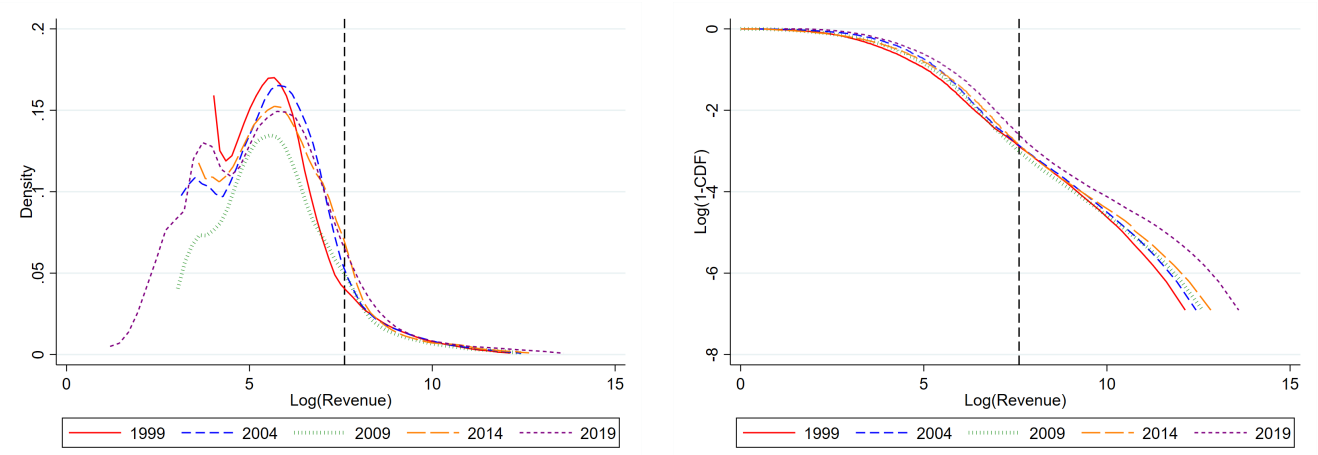
Table B.7: Components of Reallocation Gains

	H and L Based on Labor Expenditure					
	Reallocate Labor			Reallocate Intermediate Input		
Year	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3
2009	0.37	0.28	-0.52	1.33	0.17	-0.04
2014	0.39	0.24	-0.34	1.52	0.17	0.09
2019	0.32	0.21	-0.28	1.30	0.16	0.16
	H and L Based on Intermediate Input Expenditure					
	Reallocate Labor			Reallocate Intermediate Input		
	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3
2009	0.37	0.44	-0.94	1.33	0.13	0.14
2014	0.39	0.38	-0.87	1.52	0.12	0.32
2019	0.32	0.35	-0.84	1.30	0.13	0.40

Note: Reported are the means of the components in equation (B.17). For the top panel, H and L groups are based on whether labor expenditure exceeds the industry mean. For the bottom panel, H and L groups are based on whether intermediate input expenditure exceeds the industry mean.

B.7 The Distribution of Revenue

Figure B.10 shows that there is no evidence for the presence of a bulge or a dip in the revenue distribution near the 2 million Mexican peso Repeco revenue threshold. This is relevant for the issue of how strongly the tax distinction between Repeco firms and non-Repeco firms is enforced near the Repeco threshold.



(a) Densities

(b) Log Tail Probabilities

Note: Panel (a) shows the kernel density estimates of log revenue (exclusive of value added tax) across establishments for selected years, estimated using an Epanechnikov kernel with bandwidth 0.25. Panel (b) displays log(1-CDF) against log revenue, where CDF is the empirical cumulative distribution function. The vertical dashed line indicates the revenue threshold of 2 million pesos.

Figure B.10: Revenue Distribution

Appendix C Additional Theory

Consider the framework employed in Theorem 3 in the main text. Proposition C.1 provides conditions when the marginal productivity gain $\Omega'(0)$ is an upper bound to the total productivity gain $\Omega(1)$.[18] Proposition C.2 restates the formula in Theorem 3 when redistributions take the simple form analyzed in section 5. Proposition C.3 extends Theorem 3 to $S \geq 1$ industries.

Proposition C.1. *Let $(x(v), y(v), Y)$ and $(x^*(v), y^*(v), Y^*)$ be feasible allocations, where $(x(v), y(v), Y)$ is consistent with solutions to the plants' and the bundler's problem. Consider a special case where: $F : R^J \to R$, F is twice differentiable and concave and $\mathcal{Y}(x) = x$. Then (i) $\Omega(\lambda) = \int_0^\lambda \Omega'(\bar{\lambda})d\bar{\lambda}$ and $\Omega''(\lambda) \leq 0$ and (ii) $\Omega'(0) \geq \Omega(1) = \frac{Y^*}{Y} - 1$.*

Proof:

(i) The equality is the Fundamental Theorem of Calculus. The sign of Ω'' and Y'' are the same. Define $H(Y, \lambda) = \int_V \mathcal{Y}(\frac{y(v;\lambda)}{Y})d\mu - 1$, $y(v;\lambda) = z_1 F(x(v;\lambda))$ and $x(v;\lambda) = x(v) + \lambda(x^*(v) - x(v))$. Use the feasibility condition to compute $Y'(\lambda)$ and $Y''(\lambda)$.

$$H(Y, \lambda) = \int \mathcal{Y}(\tfrac{y(v;\lambda)}{Y})d\mu - 1 = 0$$

$$H_1 Y' + H_2 = 0 \Rightarrow Y' = -\tfrac{H_2}{H_1} \text{ and } Y'' = -\tfrac{H_{11}(Y')^2 + 2H_{21}Y' + H_{22}}{H_1}$$

Compute partial derivatives and evaluate Y''.

$$H_1 = -\int \mathcal{Y}'(\cdot)\tfrac{y}{Y^2}d\mu = -\int \tfrac{y}{Y^2}d\mu$$

$$H_2 = \int \mathcal{Y}'(\cdot)\tfrac{y'}{Y}d\mu = \int \tfrac{y'}{Y}d\mu$$

$$H_{11} = \int [\mathcal{Y}''(\cdot)(\tfrac{y}{Y^2})^2 + \mathcal{Y}'(\cdot)\tfrac{2y}{Y^3}]d\mu = \int \tfrac{2y}{Y^3}d\mu = -\tfrac{2}{Y}H_1$$

$$H_{12} = H_{21} = -\int [\mathcal{Y}''(\cdot)(\tfrac{y}{Y^2})\tfrac{y'}{Y} + \mathcal{Y}'(\cdot)\tfrac{y'}{Y^2}]d\mu = -\int \tfrac{y'}{Y^2}d\mu = -\tfrac{1}{Y}H_2$$

$$H_{22} = \int [\mathcal{Y}''(\cdot)(\tfrac{y'}{Y})^2 + \mathcal{Y}'(\cdot)\tfrac{y''}{Y}]d\mu = \int \tfrac{y''}{Y}d\mu$$

$$Y'' = -\tfrac{H_{11}(Y')^2 + 2H_{21}Y' + H_{22}}{H_1} = -\tfrac{H_{11}(-\frac{H_2}{H_1})^2 + 2H_{21}(-\frac{H_2}{H_1}) + H_{22}}{H_1} = -[\tfrac{-\frac{2}{Y}H_1(-\frac{H_2}{H_1})^2 - 2\frac{H_2}{Y}(-\frac{H_2}{H_1}) + H_{22}}{H_1}]$$

$$Y'' = -\tfrac{H_{22}}{H_1} = \tfrac{\int \frac{y''}{Y}d\mu}{\int \frac{y}{Y^2}d\mu} = \tfrac{\int \frac{(x^* - x)\cdot z_1 D^2 F \cdot (x^* - x)}{Y}d\mu}{\int \frac{y}{Y^2}d\mu} \leq 0$$

The inequality above follows from the fact that F is concave if and only if the Hessian D^2F is a negative semidefinite matrix. Thus, $(x^* - x) \cdot D^2F \cdot (x^* - x) \leq 0$ for all $(x^* - x)$ evaluating D^2F at any point (i.e. $x(v) + \lambda(x^*(v) - x(v))$ for $v \in V$ and $\lambda \in [0, 1]$).

[18] The formula bounds any gain or loss. If one views a loss as a positive number, then the formula provides a lower bound $|\Omega'(0)|$ to such a productivity loss.

(ii) $\Omega'(0) \geq \Omega(1) = \frac{Y^*}{Y} - 1$ follows from Proposition C.1(i) and from Ω being concave. $\|$

Proposition C.2 states the formula in Theorem 3 when the reallocation takes a simple form: $x_1^*(v) = x_1(v)(1+\gamma_1)$ if $x_1(v) \leq \bar{x}_1$ and $x_1^*(v) = x_1(v)(1-\gamma_2)$ otherwise. Feasibility implies $\gamma_2 = \gamma_1 \frac{\int_L x_1 \mathrm{d}\mu}{\int_H x_1 d\mu}$, where $L = \{v \in V : x_1(v) \leq \bar{x}_1\}$ and $H = \{v \in V : x_1(v) > \bar{x}_1\}$ for some threshold value $\bar{x}_1$.

Proposition C.2. *Consider the simple reallocation of input $j = 1$ indexed by γ_1. The marginal productivity gain is $\Omega'(0) = \gamma_1(\frac{\int_L x_1 d\mu}{\int_V x_1 d\mu})\omega_1 \left[\hat{E}_1[wedge_1|L] - \hat{E}_1[wedge_1|H]\right]$.*

Proof: Theorem 3 implies that the marginal productivity gain to reallocating input $j = 1$ is given in the first equation below. The second equation applies the simple reallocation and feasibility. The third and fourth equations reorganize terms and express the formula using conditional means, where $E[x_1|L] = \int_L x_1 \frac{d\mu}{\mu(L)}$.

$$\Omega'(0) = \omega_1 \int_V \frac{x_1}{\int_V x_1 d\mu} wedge_1 \frac{(x_1^* - x_1)}{x_1} d\mu$$

$$\Omega'(0) = \gamma_1\omega_1 \left[\int_L \frac{x_1}{\int_V x_1 d\mu} wedge_1 d\mu - (\frac{\int_L x_1 d\mu}{\int_H x_1 d\mu}) \int_H \frac{x_1}{\int_V x_1 d\mu} wedge_1 d\mu\right]$$

$$\Omega'(0) = \gamma_1(\frac{\int_L x_1 d\mu}{\int_V x_1 d\mu})\omega_1 \left[\int_L \frac{x_1}{\int_L x_1 d\mu} wedge_1 d\mu - \int_H \frac{x_1}{\int_H x_1 d\mu} wedge_1 d\mu\right]$$

$$\Omega'(0) = \gamma_1(\frac{\int_L x_1 d\mu}{\int_V x_1 d\mu})\omega_1 \left[E[(\frac{x_1}{E[x_1|L]})wedge_1|L] - E[(\frac{x_1}{E[x_1|H]})wedge_1|H]\right]$$

$$\Omega'(0) = \gamma_1(\frac{\int_L x_1 d\mu}{\int_V x_1 d\mu})\omega_1 \left[\hat{E}_1[wedge_1|L] - \hat{E}_1[wedge_1|H]\right] \ \|$$

Theorem 3 is extended to apply to $S \geq 1$ industries and two bundler levels. In this framework, Y is the total output, Y_s is output for industry s and $y_s(v)$ denotes the output of a monopolistic plant with state variable v in industry s. The primitives describing production are $(F(x;s), \mathcal{Y}_s, \mu_s)_{s=1}^S$ and H.

1. $Y = H(Y_1, ..., Y_S)$ and $\int_V z_2 \mathcal{Y}_s(\frac{y_s(v)}{Y_s}) d\mu_s = 1$ for $s = 1, \cdots, S$
2. $y_s(v) = z_1 F(x_s(v); s)$, where $x_s(v) = (x_{1s}(v), ..., x_{Js}(v))$, for $s = 1, \cdots, S$

Given measures $\mu = (\mu_1, ..., \mu_S)$, an allocation $(x_s(v), y_s(v), Y_s, Y)$ is feasible provided (i) $y_s(v) = z_1 F(x_s(v); s)$, (ii) $\int_V x_{sj}(v) d\mu_s = X_{sj}$ for $j = 1, ..., J-1$, (iii) $\int_V z_2 \mathcal{Y}_s(y_s(v)/Y_s) d\mu_s = 1$ and $Y = H(Y_1, ..., Y_S)$ and (iv) $\hat{Y} \equiv Y - \sum_s \int_V x_{Js}(v) d\mu_s > 0$. Define the productivity gain $\Omega(\lambda)$ of moving a fraction λ of the way from $(x_s(v), y_s(v), Y_s, Y)$ towards feasible allocation $(x_s^*(v), y_s^*(v), Y_s^*, Y^*)$ as a solution to system $(**)$.

$$(**)\quad 1+\Omega(\lambda)=\frac{Y(\lambda)-\sum_s\int_V(x_{sJ}(v)+\lambda(x^*_{sJ}(v)-x_{sJ}(v)))d\mu_s}{Y-\sum_s\int_V x_{sJ}(v)d\mu_s}$$

$$Y(\lambda)=H(Y_1(\lambda),...,Y_S(\lambda))\ and\ \int_V z_2\mathcal{Y}_s(\frac{z_1F(x_s(v)+\lambda(x^*_s(v)-x_s(v));s)}{Y_s(\lambda)})d\mu_s=1$$

Proposition C.3. *Assume $(x_s(v),y_s(v),Y_s,Y)$ and $(x^*_s(v),y^*_s(v),Y^*_s,Y^*)$ are feasible, where $(x_s(v),y_s(v),Y_s,Y)$ is consistent with solutions to the plants' and the bundler's problem. Then $\Omega(\lambda)=\int_0^\lambda\Omega'(\bar\lambda)d\bar\lambda$ and $\Omega'(0)=\sum_s\sum_j\omega_{js}\hat E_{js}[wedge_{js}\frac{(x^*_{js}-x_{js})}{x_{js}}]-\frac{\sum_s\int_V(x^*_{Js}-x_{Js})d\mu_s}{Y-\sum_s\int_V x_{Js}d\mu_s}$, where $\omega_{js}=\int_V w_jx_{js}d\mu_s/(Y-\sum_s\int_V x_{Js}d\mu_s)$ are the share of final output paid for input j in industry s.*

Proof: $\Omega(\lambda)-\Omega(0)=\int_0^\lambda\Omega'(\bar\lambda)d\bar\lambda$ follows from the Fundamental Theorem of Calculus. The result follows as $\Omega(0)=0$. Define $x_s(v;\lambda)=x_s(v)+\lambda(x^*_s(v)-x_s(v))$ and $y_s(v;\lambda)=z_1F(x_s(v;\lambda))$. Feasibility implies

$$Y'_s(\lambda)=Y_s(\lambda)\int_V\frac{z_2\mathcal{Y}'_s(\frac{y_s(v;\lambda)}{Y_s(\lambda)})}{\int z_2\mathcal{Y}'_s(\frac{y_s(v;\lambda)}{Y_s(\lambda)})y_s(v;\lambda)d\mu_s}\sum_{j=1}^J z_1F_j(x_s(v;\lambda);s)(x^*_{js}(v)-x_{js}(v))d\mu_s$$

$$\Omega'(\lambda)=\frac{\sum_s H_s(Y_1(\lambda),\cdots,Y_S(\lambda))Y'_s(\lambda)-\sum_s\int_V(x^*_{Js}(v)-x_{Js}(v))d\mu_s}{Y-\sum_s\int_V x_{Js}(v)d\mu_s}$$

Restate the marginal productivity gain, evaluated at $\lambda=0$, using the first-order condition of each bundler. The price function of the monopolistic firm's good is then $P_s(y,z_2,Y_s)=\frac{P_sY_sz_2\mathcal{Y}'_s(\frac{y}{Y_s})}{\int_V z_2\mathcal{Y}'_s(\frac{y_s(v)}{Y_s})y_s(v)d\mu_s}=\frac{H_sY_sz_2\mathcal{Y}'_s(\frac{y}{Y_s})}{\int_V z_2\mathcal{Y}'_s(\frac{y_s(v)}{Y_s})y_s(v)d\mu_s}$, where P_s is the price of industry s output. Apply the first-order condition $P_s(y,z_2,Y_s)z_1F_j=w_jwedge_{js}$ to produce the second equation below. Define cost shares of final output $\omega_{js}=\int_V w_jx_{js}d\mu_s/(Y-\sum_s\int_V x_{Js}(v)d\mu_s)$ and employ the normalization $\mu_s(\cdot)/\mu_s(V)$ for the expectations operators to get the third and fourth equation below. The last equation employs input-share weighted expectations operator $\hat E_{js}$ for industry s.

$$\Omega'(0)=\frac{\sum_s\int_V P_s(y,z_2,Y_s)\sum_{j=1}^J z_1F_j(x_s(v);s)(x^*_{js}(v)-x_{js}(v))d\mu_s-\sum_s\int_V(x^*_{Js}(v)-x_{Js}(v))d\mu_s}{Y-\sum_s\int_V x_{Js}(v)d\mu_s}$$

$$\Omega'(0)=\frac{\sum_s\int_V\sum_{j=1}^J w_jwedge_{js}(x_s)(x^*_{js}-x_{js})d\mu_s-\sum_s\int_V(x^*_{Js}-x_{Js})d\mu_s}{Y-\sum_s\int_V x_{Js}d\mu_s}$$

$$\Omega'(0)=\sum_s\sum_j\omega_{js}\int_V\frac{w_jx_{js}}{\int_V w_jx_{js}d\mu_s}wedge_{js}\frac{(x^*_{js}-x_{js})}{x_{js}}d\mu_s-\frac{\sum_s\int_V(x^*_{Js}-x_{Js})d\mu_s}{Y-\sum_s\int_V x_{Js}d\mu_s}$$

$$\Omega'(0) = \sum_s \sum_j \omega_{js} E_s[\frac{x_{js}}{E_s[x_{js}]} wedge_{js} \frac{(x^*_{js} - x_{js})}{x_{js}}] - \frac{\sum_s \int_V (x^*_{Js} - x_{Js}) d\mu_s}{Y - \sum_s \int_V x_{Js} d\mu_s}$$

$$\Omega'(0) = \sum_s \sum_j \omega_{js} \hat{E}_{js}[wedge_{js} \frac{(x^*_{js} - x_{js})}{x_{js}}] - \frac{\sum_s \int_V (x^*_{Js} - x_{Js}) d\mu_s}{Y - \sum_s \int_V x_{Js} d\mu_s} \quad \|$$

Appendix D Levy Narrative

The introduction compressed some of the themes in Levy (2018) into a two sentence narrative that is restated below in bold face. While econometricians celebrate methods for data compression, much is left out. The Appendix documents that what was left in is contained in the work of Levy.

He argues that a combination of tax, labor contracting and enforcement policies imply that some large Mexican establishments face high marginal tax rates on labor costs and many small establishments face low tax rates. He then conjectures that alternative policies that reduce size-dependent taxation are likely to be important for increasing productivity.

Concerning size-dependent taxation, Levy (2018, p. 38) states "Larger firms with salaried contracts are de facto taxed more than smaller ones. This is because enforcement of social insurance contributions by IMSS and Infonavit is concentrated in larger firms." Busso, Fazio and Levy (2012, p. 10) state " large firms will hire salaried workers mostly legally and therefore are taxed, small firms hiring salaried workers will do so mostly illegally and are therefore partly subsidized, and firms with nonsalaried contracts, regardless of size, are subsidized more than small firms hiring salaried workers (since they face no probability of being fined regardless of their number of workers)."

Concerning policy, Levy (2018, p. 2, 10 and 283) states "The book points out that the main policies and institutions impeding growth are those related to taxation, labor and social insurance regulations, and enforcement of contracts." and "Misallocation is the central part of the explanation of why productivity and growth in Mexico have stagnated." and "This book argues that current tax, social insurance, and labor protection policies are flawed, that they are the main reason why growth is slow, and that a policy shift is necessary."

Appendix E Calibrating Model Parameters

Model 1: Monopolistic Competition, Two inputs, CES production, Multiple taxes

1. Primitives: $y = z_1 F(x_1, x_2; z_2) = z_1[a(x_1 z_2)^{\alpha} + (1-a)x_2^{\alpha}]^{1/\alpha}, \mathcal{Y}(x) = x^{\gamma}, (\ln z_1, \ln z_2, \tau^{labor}, \tau^{va}, \tau^{va,m}) \sim N(\mu, \Sigma)$

   where $\sigma = 1/(1-\alpha)$ is the elasticity of substitution, $\gamma \in (0,1)$ and $1/\gamma$ is the markup

2. Parameters: $\Gamma = (\alpha, \gamma, a, \mu, \Sigma)$ (estimated); τ^c (externally set)

3. State Variables: $v = (z_1, z_2, \tau)$ and $\tau = (\tau^{labor}, \tau^{va}, \tau^{va,m})$

4. $T(w_1x_1, w_2x_2, py; \tau) = \tau^{labor} w_1x_1 + \tau^{va,m} w_2x_2 + \tau^{va} py + \tau^c \left[(1-\tau^{va})py - (1+\tau^{labor})w_1x_1 - (1+\tau^{va,m})w_2x_2\right]$

5. Bundler: a solution to the Bundler's problem induces $P(y, Y) = y^{\gamma-1}Y^{1-\gamma}$

The state vector has five dimensions: productivity (z_1, z_2) and tax $(\tau^{labor}, \tau^{va}, \tau^{va,m})$. Productivities are lognormal and tax rates are normal: $(\ln z_1, \ln z_2, \tau^{labor}, \tau^{va}, \tau^{va,m}) \sim N(\mu, \Sigma)$. One additional tax is deterministic: the corporate profit tax τ^c.

Covariance restriction: each tax covaries only with z_1 (neutral productivity). All tax–tax and tax–z_2 covariances are set to zero:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{12} & \sigma_{22} & 0 & 0 & 0 \\ \sigma_{13} & 0 & \sigma_{33} & 0 & 0 \\ \sigma_{14} & 0 & 0 & \sigma_{44} & 0 \\ \sigma_{15} & 0 & 0 & 0 & \sigma_{55} \end{pmatrix}$$

Total estimated parameters: α, γ, a (3) $+\mu$ (5) $+\Sigma$ (9) $= 17$, plus τ^c calibrated externally.

P1: $\max_{x_1,x_2}\ P(y,Y)y - w_1x_1 - w_2x_2 - T(w_1x_1, w_2x_2, P(y,Y)y; \tau)$ s.t. $y = z_1F(x_1, x_2; z_2)$

FOC: $(1-T_3)\gamma P(y,Y) z_1 F_1 = w_1(1+T_1)$ and $(1-T_3)\gamma P(y,Y) z_1 F_2 = (1+T_2)w_2$

Proposition E.1. *Given the assumptions in Model 1, in a solution (x_1, x_2, y) to Problem P1*

(i) $\frac{w_1x_1(1+T_1) + w_2x_2(1+T_2)}{(1-T_3)P(y,Y)y} = \gamma$

(ii) $\ln \frac{w_1x_1}{w_2x_2} = C_0 - \sigma \ln(\frac{1+T_1}{1+T_2}) - (1-\sigma)\ln z_2$, *where* $\sigma = \frac{1}{(1-\alpha)}$, $\frac{1+T_1}{1+T_2} = \frac{1+\tau^{labor}}{1+\tau^{va,m}}$ *and C_0 is a constant.*

Proof:

(1) Multiply FOC by x_1 and x_2, respectively, and sum to produce the first equation below. The second equation follows after rearrangment and applying Euler's Theorem.

$(1-T_3)\gamma P(y,Y)\sum_j z_1 F_j x_j = w_1 x_1(1+T_1) + w_2 x_2(1+T_2)$

$\frac{w_1 x_1(1+T_1)+w_2 x_2(1+T_2)}{(1-T_3)P(y,Y)y} = \gamma$

(2) The ratio of FOC implies the first equation. The second rearranges the first equation. The third multiplies the second by the ratio of factor prices. The last equation takes logs, collects constant terms in C_0 and uses the definition $\sigma = 1/(1-\alpha)$ to express the result in terms of the elasticity of substitution σ.

$\frac{w_1(1+T_1)}{w_2(1+T_2)} = \frac{F_1}{F_2} = \frac{a}{1-a}(\frac{x_1}{x_2})^{\alpha-1} z_2^{\alpha}$

$\frac{x_1}{x_2} = [\frac{w_1(1+T_1)}{w_2(1+T_2)}\frac{1-a}{a} z_2^{-\alpha}]^{1/(\alpha-1)}$

$\frac{w_1 x_1}{w_2 x_2} = [\frac{1-a}{a}]^{1/(\alpha-1)}[\frac{w_1}{w_2}]^{\alpha/(\alpha-1)}[\frac{1+T_1}{1+T_2}]^{1/(\alpha-1)} z_2^{\alpha/(1-\alpha)}$

$\ln \frac{w_1 x_1}{w_2 x_2} = C_0 - \sigma \ln(\frac{1+T_1}{1+T_2}) - (1-\sigma)\ln z_2$, where $\sigma = 1/(1-\alpha)$ $\quad \|$

****

Wenlan: alternative tax system [Case 1 $= \tau^{va,m} w_2 x_2 < \tau^{va} py$, Case 2 $= \tau^{va,m} w_2 x_2 > \tau^{va} py$]:
$T(w_1 x_1, w_2 x_2, py; \tau) = \tau^{labor} w_1 x_1 + \max\{\tau^{va,m} w_2 x_2, \tau^{va} py\} + \tau^c \left[(1-\tau^{va})py - (1+\tau^{labor})w_1 x_1 - w_2 x_2\right]$

$$T_1 = \tau^{labor} - \tau^c(1+\tau^{labor}) \tag{E.1}$$

$$T_2 = \begin{cases} 0 - \tau^c & \text{if case 1} \\ \tau^{va,m} - \tau^c & \text{if case 2} \end{cases} \tag{E.2}$$

$$T_3 = \begin{cases} \tau^{va} + \tau^c(1-\tau^{va}) & \text{if case 1} \\ 0 + \tau^c(1-\tau^{va}) & \text{if case 2} \end{cases} \tag{E.3}$$

$$\frac{1+T_1}{1+T_2} = \begin{cases} \frac{1+\tau^{labor}-\tau^c(1+\tau^{labor})}{1-\tau^c} = 1+\tau^{labor} & \text{if case 1} \\ \frac{1+\tau^{labor}-\tau^c(1+\tau^{labor})}{1+\tau^{va,m}-\tau^c} & \text{if case 2} \end{cases} \tag{E.4}$$

$$\frac{1+T_1}{1-T_3} = \begin{cases} \frac{1+\tau^{labor}-\tau^c(1+\tau^{labor})}{1-[\tau^{va}+\tau^c(1-\tau^{va})]} = \frac{1+\tau^{labor}}{1-\tau^{va}} & \text{if case 1} \\ \frac{1+\tau^{labor}-\tau^c(1+\tau^{labor})}{1-[\tau^c(1-\tau^{va})]} & \text{if case 2} \end{cases} \tag{E.5}$$

$$\frac{1+T_2}{1-T_3} = \begin{cases} \frac{1-\tau^c}{1-[\tau^{va}+\tau^c(1-\tau^{va})]} = \frac{1}{1-\tau^{va}} & \text{if case 1} \\ \frac{1+\tau^{va,m}-\tau^c}{1-[\tau^c(1-\tau^{va})]} & \text{if case 2} \end{cases} \tag{E.6}$$

****

Calibration: find Γ solving $\min_\Gamma \sum_j \sum_k (m_j^k - m_j^k(\Gamma))^2$

Empirical Moments from INEGI:

1. $m_0 = \frac{\sum_i [L4_i + SSTax_i + \tau_i^{sp} L4_i + I_i + VATm_i]}{\sum_i Rev_i}$
2. $m_1^k = E[\ln \text{taxwedge}_{1i} | s^k]$ for $k = 1, ..., K$, where s^k are equally-spaced values of $\ln(I)$.
3. $m_2^k = E[\ln \text{taxwedge}_{2i} | s^k]$ for $k = 1, ..., K$, where s^k are equally-spaced values of $\ln(L4)$.
4. $m_3^k = E[\ln(1 + \tau_i^{va,m}) | s^k]$ for $k = 1, ..., K$, where s^k are equally-spaced values of $\ln(Rev)$
5. $m_4 = (\text{Var}[\tau_i^{labor}],\ \text{Var}[\tau_i^{va}],\ \text{Var}[\tau_i^{va,m}])$
6. $m_5^k = E[\ln(L4_i/I_i) | s^k]$ for $k = 1, ..., K$, where s^k are equally-spaced values of $\ln(Rev)$
7. $(m_6^1, ..., m_6^Q) = (p90 \ln Rev, ..., p10 \ln Rev)$
8. $m_7 =$ #plants/#employ, (i.e. ratio of plants to employment within an industry)

Size profiles (m_1, m_2, m_3, m_5) are local polynomial fits evaluated at K equally spaced sizes. Total: $4K + Q + 5$ moments. Moments m_1, m_2, and m_3 condition on intermediate-input size, whereas m_5 conditions on revenue size. Hence the model uses these profiles jointly rather than comparing them point by point at a common size argument.

(Equilibrium) Model Moments:

1. $m_0(\Gamma) = \gamma$
2. $m_1^k(\Gamma) = E[\ln \text{taxwedge}_1 | \ln(x_2(v)) = s^k]$
3. $m_2^k(\Gamma) = E[\ln \text{taxwedge}_2 | \ln(x_1(v)) = s^k]$
4. $m_3^k(\Gamma) = E[\ln(1 + \tau^{va,m}) | \ln(Rev(v)) = s^k]$
5. $m_4(\Gamma) = (\sigma_{33},\ \sigma_{44},\ \sigma_{55})$
6. $m_5^k(\Gamma) = E[\ln(w_1 x_1(v)/w_2 x_2(v)) | \ln(Rev(v)) = s^k]$
7. $(m_6^1(\Gamma), ..., m_6^Q(\Gamma))$: quantiles of $\ln(Rev(v))$
8. $m_7(\Gamma) = \int_E \mathrm{d}H$

The block structure below describes which moments are most informative about which parameters. The actual estimation is joint (all moments discipline all parameters).

Moment m_0 pins down γ: by Prop. E1, the ratio of total factor cost (inclusive of marginal tax rates) to net-revenue equals γ for each firm. The empirical counterpart uses $L4+SSTax+\tau^{sp}L4+I+VATm$ in the numerator and net-of-output-VAT revenue $\hat{Rev}$ in the denominator.

The tax-rate mean profiles m_1, m_2, m_3 are most informative about the tax means (μ_3, μ_4, μ_5) and the tax–z_1 correlations $(\sigma_{13}, \sigma_{14}, \sigma_{15})$. Since tax–$z_2$ covariances are zero, the profiles' slopes are driven entirely by how taxes covary with neutral productivity z_1. The levels of the profiles pin down the means. The unconditional tax variances m_4 directly identify the diagonal elements $(\sigma_{33}, \sigma_{44}, \sigma_{55})$.

The cost ratio profile m_5 identifies σ (the CES elasticity), a (the distribution parameter), and the z_2 distribution. Based on Prop. E.1:

$$E[\ln(w_1x_1/w_2x_2) \mid s_{rev}] = C_0 - \sigma E[\ln \frac{(1+T_1)}{(1+T_2)} \mid s_{rev}] - (1-\sigma)E[\ln z_2 \mid s_{rev}].$$

$$E[\ln(w_1x_1/w_2x_2) \mid s_{rev}] = C_0 - \sigma E[\ln(1+\tau^{labor}) \mid s_{rev}] + \sigma E[\ln(1+\tau^{va,m}) \mid s_{rev}] - (1-\sigma)E[\ln z_2 \mid s_{rev}].$$

The size distribution m_6 is most informative about the z_1 distribution (μ_1, σ_{11}), as neutral productivity is the primary driver of size dispersion. Revenue depends on particular powers of z_1, z_2 and the tax rates.

Moment m_7 pins down the fraction of the population acting as plant managers, disciplining the occupational-choice margin. The entry cutoff is implied by (Γ, w_1, Y) in equilibrium, and m_7 helps locate the productivity distribution, especially μ_1, relative to the worker option.

Key intuition: the declining $E[\ln(w_1x_1/w_2x_2)|s]$ arises when $\sigma < 1$ and z_2 increases with size. The separately observed tax wedge profiles (m_1, m_3) help separate the contributions of $\ln(1+\tau^{labor})$ and $\ln(1+\tau^{va,m})$ (which also affect the cost ratio) from z_2.

Parameter(s)	Primary moment(s)
γ	m_0: aggregate cost-to-net-revenue ratio
μ_3, σ_{13}	m_1: labor tax wedge profile
μ_4, σ_{14}	m_2: intermediate input tax wedge profile
μ_5, σ_{15}	m_3: input VA rate mean profile
$\sigma_{33}, \sigma_{44}, \sigma_{55}$	m_4: unconditional tax variances
$\sigma, a, \mu_2, \sigma_{22}$	m_5: cost ratio profile (with m_1, m_3)
$\mu_1, \sigma_{11}, \sigma_{12}$	m_6, m_7: size distribution and plants per person

Estimation Procedure.

For a given Γ: (i) Draw from the joint distribution $(\ln z_1, \ln z_2, \tau^{labor}, \tau^{va}, \tau^{va,m}) \sim N(\mu, \Sigma)$ using Monte-Carlo simulation in 5 dimensions. (ii) Solve the equilibrium (w_1, Y) (Steps 1-4 below). (iii) Compute model moments $m_j(\Gamma)$ by integrating over $H = N(\mu, \Sigma)$ conditional on the relevant size

measure. Size profiles are computed using local polynomials fit, mirroring the data construction. (iv) Evaluate $\sum_j \sum_k (m_j^k - m_j^k(\Gamma))^2$ and update Γ via nonlinear optimization.

Initial guesses: $\gamma = m_0$; (μ_3, μ_4, μ_5) from unconditional means, $(\sigma_{33}, \sigma_{44}, \sigma_{55})$ from unconditional variances of the respective tax rates in data; σ from OLS of $\ln(w_1 x_1 / w_2 x_2)$ on $\ln(1+\tau^{labor})$ and $\ln(1+\tau^{va,m})$; (μ_1, σ_{11}) from revenue dispersion.

Compute an equilibrium, given Γ, τ^c, where $w_2 = 1$ (numeraire).

Step 1: Guess (w_1, Y)

Step 2: Decisions: Compute $x_1(v; w_1, Y), x_2(v; w_1, Y)$ in closed form. Compute $y(v; w_1, Y)$, $\pi(v; w_1, Y)$, and entry set $E(w_1, Y) = \{v : \pi(v; w_1, Y) \geq w_1\}$.

Step 3: Evaluate: $\int_{E(w_1,Y)} x_1(v; w_1, Y) dH = 1 - H(E(w_1, Y))$ and $\int_{E(w_1,Y)} \mathcal{Y}(y(v; w_1, Y)/Y) dH = 1$

Step 4: Stop if Step 3 holds up to tolerance. Otherwise, update (w_1, Y) and repeat Steps 2-3.

Closed-Form Solution to the Firm's Problem.

Step A (input ratio): The ratio of FOC (see the proof of Proposition E.1(ii)) gives the input ratio:

$$r \equiv \frac{x_1}{x_2} = \left[\frac{(1+T_1)\, w_1}{(1+T_2)\, w_2} \cdot \frac{1-a}{a} \cdot z_2^{-\alpha} \right]^{1/(\alpha-1)}$$

Step B (input level): Define $\phi(r) \equiv [a(rz_2)^\alpha + (1-a)]^{1/\alpha}$, so that $F(rx_2, x_2; z_2) = x_2\, \phi(r)$ and $y = z_1\, x_2\, \phi(r)$. The marginal product F_2, evaluated at $x_1 = r x_2$, simplifies as follows:

$$\begin{aligned} F_2 &= [a(x_1 z_2)^\alpha + (1-a) x_2^\alpha]^{1/\alpha - 1} \cdot (1-a)\, x_2^{\alpha-1} \\ &= \underbrace{[x_2^\alpha (a(rz_2)^\alpha + (1-a))]^{(1-\alpha)/\alpha}}_{= x_2^{1-\alpha}\, \phi(r)^{1-\alpha}} \cdot (1-a)\, x_2^{\alpha-1} = (1-a)\, \phi(r)^{1-\alpha} \end{aligned}$$

Thus $z_1 F_2 = z_1 (1-a)\, \phi(r)^{1-\alpha}$ is independent of x_2. Substituting into the FOC for x_2 and using $P(y, Y) y = z_1^\gamma\, x_2^\gamma\, \phi(r)^\gamma\, Y^{1-\gamma}$ yields

$$(1-T_3)\, \gamma\, (1-a)\, z_1^\gamma\, \phi(r)^{\gamma-\alpha}\, Y^{1-\gamma}\, x_2^{\gamma-1} = (1+T_2)\, w_2$$

Since $\gamma - 1 < 0$, this is a power equation with solution

$$x_2 = \left[\frac{(1-T_3)\, \gamma\, (1-a)\, z_1^\gamma\, \phi(r)^{\gamma-\alpha}\, Y^{1-\gamma}}{(1+T_2)\, w_2} \right]^{1/(1-\gamma)}$$

Step C: $x_1 = r\, x_2$, $y = z_1\, x_2\, \phi(r)$, $\pi = (1-T_3)\, y^\gamma Y^{1-\gamma} - (1+T_1)\, w_1\, x_1 - (1+T_2)\, w_2\, x_2$.

Counterfactual 1: Eliminate τ^{labor}, raise output VAT, respecting revenue neutrality.

Keep the distribution H over the state vector v unchanged. Define a new tax function that zeroes out the labor tax on labor expenditures, maintains the existing taxes on revenue and intermediate inputs, and adds a common additional tax τ^{cf} on revenue of all plants:

$$T^{cf}(v) = (\tau^{va} + \tau^{cf})Py + \tau^{va,m} w_2 x_2 + \tau^c \left[(1 - \tau^{va} - \tau^{cf})Py - w_1 x_1 - (1 + \tau^{va,m}) w_2 x_2\right]$$

The firm's counterfactual problem:

$\max_{x_1,x_2}\ (1-\tau^c)\left[(1-\tau^{va}-\tau^{cf})P(y,Y)y - w_1x_1 - (1+\tau^{va,m})w_2x_2\right]$

Revenue neutrality: find τ^{cf} such that $\int_{E^{cf}} T^{cf}(v)\mathrm{d}H = \int_{E^{base}} T^{base}(v)\mathrm{d}H$.

This exercise eliminates the input-mix distortion from the size-dependent τ^{labor}. Since the output VAT does not distort the labor-to-materials input ratio , the replacement of τ^{labor} with a common revenue tax τ^{cf} removes the labor-specific component of the size-dependent wedge between inputs.

Appendix F Additional Descriptive Statistics

F.1 Updated List (updated SSRate)

Updated (10/14/2025): Figure 2,

F.2 Updated List (updated L4)

Updated (11/12/2025): Figure 2, Figure 4, Figure 5, Figure 6, Table 2; Figure B.9, Figure B.10;

Table 1 (added mean estab. employment)

Figure B.2, Figure B.5

Added Figure B.3 for Variance of Wedge- Size Profiles

Updated (11/14/2025): Figure 7, Figure B.10 (on reallocation gains); Table B.6 (reallocation gains decomposition), Table B.4 (industry cost shares), Figure B.5 (new, level variance of wedge/tax wedge-size profiles)

Updated (11/17/2025): Figure 8 (added aggregate gains); Figure 7 (updated Figure (a) with same x-axis scale)

F.3 Updated List (updated ii)

Updated (1/13/2026): Figure 4, Figure 5, Figure 7, Figure 8

F.4 Updated List (1/29/2026)

Updated (1/29/2026):

- Figure 2(a)(b): ssrate and varate by quantile
- Figure 4(a)(b): labor and ii wedge by size
- Figure 6(a)-(f): compare labor and ii wedges 2009/2014/2019
- Figure 8(a)(b): cross-industry distribution of wedge correlations
- Figure 10(a)-(d): productivity gains from input reallocation
- Figure **??**(a)(b): output elasticity by size
- Figure **??**(a)(b): wedge size profiles with variable output elasticity
- Figure B.6(a)(b): variance of wedge by size profiles I

- Figure B.7(a)-(d): variance of wedge by size profiles II
- Figure B.4: value added tax rates and SS subcomponents by size
- Table 1 Sample Selection and Sample Properties
- Table 2 Decomposition of the Labor Wedge Correlation ; Panel B.

References

Ackerberg, D., Caves, K. and G.Frazer (2015), Identification Properties of Recent Production Function Estimators, Econometrica, 83, 2411-51.

Anton, A., Hernandez, F. and S. Levy (2012), The End of Informality in Mexico? Fiscal Reform and Universal Social Insurance, Inter-American Development Bank.

Busso, M., Fazio, M. and S. Levy (2012), (In)Formal and (Un)Productive: the Productivity Costs of Excessive Informality in Mexico, IDB Working Paper 341.

Busso, M., Fentanes, O. and S. Levy (2018), The longitudinal linkage of Mexico's economic census 1999-2014. IDB Technical Note IDBTN-1477.

Chapa, J., Ayala, E., and Cabral, R. (2016), ¿Quien paga el impuesto sobre nominas de los estados de Mexico?: Un analisis de incidencia economica, Premio Nacional de Finanzas Públicas, 6(21), 69–112.

Hsieh, C. and P. Klenow (2009), Misallocation and Manufacturing TFP in China and India, Quarterly Journal of Economics, 124, 1403-48.

Levy, S. (2018), Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico, Inter-American Development Bank.

Raval, D. (2023), Testing the Production Approach to Markup Estimation, Review of Economic Studies, 90, 2592-2611.

Raval, D. (2023a), A Flexible Cost Share Approach to Markup Estimation, Economic Letters, 230, 111262.