

Persistent Misallocation and the Returns to Education in Mexico

Santiago Levy
Luis Felipe López-Calva



WORLD BANK GROUP

Development Economics Vice Presidency
Strategy and Operations Team
January 2019

Abstract

Over the last two decades, Mexico has experienced macroeconomic stability, an open trade regime, and substantial progress in education. Yet average workers' earnings have stagnated, and earnings of those with higher schooling have fallen, compressing the earnings distribution and lowering the returns to education. This paper argues that distortions that misallocate resources toward less-productive firms explain these phenomena, because these firms are less intensive in well-educated workers compared with more-productive ones. It shows that while the relative supply of workers with more years of schooling has increased, misallocation of resources toward less productive firms has persisted. These two trends have generated a widening mismatch between the supply of, and the demand

for, educated workers. The paper breaks down worker earnings into observable and unobservable firm and individual worker characteristics, and computes a counterfactual earnings distribution in the absence of misallocation. The main finding is that in the absence of misallocation average earnings would be higher, and that earnings differentials across schooling levels would widen, raising the returns to education. A no-misallocation path is constructed for the wage premium. Depending on parameter values, this path is found to be rising or constant, in contrast to the observed downward path. The paper concludes arguing that the persistence of misallocation impedes Mexico from taking full advantage of its investments in the education of its workforce.

This paper is a product of the Strategy and Operations Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at calva@undp.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Persistent Misallocation and the
Returns to Education in Mexico

Santiago Levy*

Luis Felipe López-Calva*

JEL classifications: J24, J23, O17, L11[O]

Keywords: earnings inequality, returns to schooling, misallocation, Mexico.

*Santiago Levy is the Vice President for Sectors and Knowledge at the Inter-American Development Bank; his email address is slevy@iadb.org. * Luis Felipe López-Calva (corresponding author) is the Regional Director for Latin America and the Caribbean at the United Nations Development Programme in New York; his email is luis.lopez-calva@undp.org. The opinions of the authors do not necessarily coincide with those of the institutions with which they are affiliated. The authors are grateful to Matias Morales for excellent research assistance and to Luca Flabbi, Samuel Freije, Rafael de Hoyos, Julian Messina, Hugo Nopo, Norbert Schady, and Miguel Székely for useful comments on an earlier draft. Comments by three anonymous referees and the journal editor were helpful to improve the presentation and content of this version. Participants in presentations held at the World Bank, the Inter-American Development Bank, the Bank of Spain, Universidad Iberoamericana and Centro de Investigación y Docencia Económica in Mexico; and the Latin American and Caribbean Economic Association's Conference in Medellín, Colombia, also provided useful comments. The authors also thank, without implicating, Rodrigo Negrete and Tomas Ramirez of Mexico's Instituto Nacional de Estadística, Geografía e Informática, for their help interpreting the employment surveys.

1. Introduction

Over the last two decades, Mexico has made notable efforts to increase the schooling of its workers in the hope that accumulating human capital would lead to higher earnings and a greater number of jobs covered by labor and social insurance. Indeed, there has been a significant increase in schooling levels. In 1996, working-age Mexicans (18 years or older) had on average 4.7 years of education; by 2015, that figure had almost doubled, to 9.2 years. Similarly, in 1996 less than 19 percent of working-age Mexicans had completed high school; by 2015, 33 percent had done so. However, despite these achievements, and the fact that the two decades were characterized by macroeconomic stability and a wide opening to international trade, hopes for higher earnings and better jobs have not materialized. The share of jobs covered by labor and social insurance regulations has remained essentially constant, and average hourly earnings, after recovering from the 1995 financial crisis, have fallen slightly. This fall in earnings since the mid 1990s is the result of an absolute decline in the earnings of workers with more years of schooling that has, by and large, offset the expected increase in average earnings associated with the change in the schooling composition of the labor force.

The literature on the returns to education has focused on understanding the relative importance of supply and demand factors in determining the distribution of earnings across educational levels. In Mexico's case, attention has focused on the fact that the earnings differential between workers with more and fewer years of education—the wage premium—has narrowed over the last decade, if not before.¹ This finding is puzzling because, on the one hand, human capital is thought to be a

¹ López-Calva and Lustig (2010) find that there has been a steady decline in the wage premium between skilled and unskilled workers at least since 2002. Robertson (2007) suggests that the decline started at the end of the 1990s. In parallel, Campos-Vázquez, Esquivel, and Lustig (2012) find that returns to schooling started to decline after 1994. See Levy and López-Calva (2016) for a brief review.

constraint on growth in Mexico; and, on the other hand, because it is the opposite of the trend found in the United States (Mexico's largest trading partner by far), where the wage premium has increased (Goldin and Katz 2007; Autor, Katz, and Kearney 2008).

In an immediate sense, of course, the fact that earnings of workers with more years of education have fallen as their supply has increased suggests a normal market adjustment. But this explanation is almost tautological, begging the question as to why the demand for workers with more education has lagged.

This paper makes a bridge between the literature on the returns to education and the literature on misallocation. It argues that large and persistent misallocation explains why the demand for more educated workers has lagged. The main point is that because of misallocation, the number, type, and size of firms participating in the demand side of the labor market is strongly distorted toward low productivity firms that are intensive in less educated workers. The perspective here is that, given workers' observable and unobservable characteristics, their earnings partly depend on the nature of the firms that employ them. In this sense, the paper explores how the size distribution of firms (measured by the total number of workers) and the type distribution of firms (measured by the contractual composition of their workforce) impact the distribution of employee earnings and the returns to schooling.

This line of inquiry is relevant because of three empirical regularities documented below. First, controlling for firm type, larger firms are more intensive in educated workers than smaller ones; second, controlling for firm size, firms that offer their workers contracts with coverage of labor and social insurance regulations are more intensive in educated workers than other firms; and, third, there is a strong positive correlation between firms that are large and firms that offer their

workers contracts with coverage of labor and social insurance regulations. As a result, if too few resources are allocated to these firms, the schooling composition of the demand for labor will tilt in the direction of workers with fewer years of education.²

A large literature has analyzed the role of taxation, social insurance and labor regulations, credit frictions, market failures, and other factors, such as registration costs, in generating misallocation (IDB 2010, LaPorta and Schleifer 2008). In the case of Mexico, Levy (2008) and Busso, Fazio, and Levy (2012) emphasize the role of labor and social insurance regulations; Leal (2014), the role of taxation; and López-Martin (2015), the role of credit. In all these cases, misallocation results in too many low-productivity firms employing too many workers without coverage of labor and social insurance regulations, that is, in a large informal sector.

This paper does not model the frictions, market, or regulatory failures that generate misallocation. Rather, to test our hypothesis, a three-step approach is followed. First, a model of individual worker earnings that controls for all observable worker characteristics is estimated, but focuses on estimating the coefficients associated with observable firm characteristics.

Second, the implications for individual workers of eliminating misallocation are also considered, interpreted here as eliminating firm informality. To do so, a counterfactual earnings distribution is constructed keeping constant the unobservable and observable characteristics of individual workers, such as years of schooling, age, gender, and location, but assuming that the size and type distribution of firms mimics that of formal firms. The purpose is to measure how worker earnings are affected if, because of misallocation, there are too many informal firms in the demand side of

² Bobba, Flabbi, and Levy (2017) develop a search and bargaining model where firms and workers form matches (jobs) that can be formal or informal. Modeling some of the distortions present in Mexico's labor market, they show that the returns to education follow a path consistent with the findings in this paper, with the additional effect of a disincentive for individuals to invest in higher levels of education.

the labor market, independently of worker education and abilities. In this context, the analysis shows that eliminating firm informality changes the schooling composition of the demand for labor, increasing average earnings and augmenting the demand for more educated workers relative to those with fewer years of schooling, thus increasing the returns to schooling. The mean of earnings for all educational levels is higher, while the distribution across educational levels is widened. In other words, misallocation acts like a penalty on earnings that is paid by all workers, but proportionately more by the more educated.

Finally, the aggregate implications of eliminating misallocation are considered. The analysis shows that, given the supply of workers at each educational level, if the schooling composition of the demand for labor in the economy equaled that of formal firms, there would be an excess supply of workers with few years of education. The size of the excess supply is measured and, critically, it is assessed that it would be increasing over time. Next, for given reasonable values of the elasticity of substitution between workers of different schooling levels, the changes in earnings required to absorb the excess supply are computed. Then the observed path of the wage premium is compared with alternative paths: where there is no firm informality and earnings adjust in each period to clear the market. The analysis finds that, in the absence of firm informality, the wage premium is higher and that the difference relative to the observed premium increases over time.

The paper follows Hsieh and Klenow (2009) to calculate the measures of misallocation. Firms within narrowly defined sectors (at the six-digit level in our calculations) are assumed to produce goods that are imperfect substitutes for each other. Firms also differ in their production technology.³ Misallocation is reflected in the fact that because of idiosyncratic distortions, and

³ Hsieh and Klenow assume that the production technology is $Q=AK^\alpha L^{1-\alpha}$, with firms differing in their value for A; firms with higher values for A relative to the sector's average should in principle be larger than those with lower values for A. However, if distortions are present, this may not be the case.

given the demand for their product and their production technology, firms demand the wrong amount of resources. Some firms are larger than they should be, given their underlying technology; and, concomitantly, some are smaller. Because the schooling composition of firms' demand for labor differs, misallocation then affects the schooling composition of the demand for labor.

Section 2 shows that misallocation shifts too many resources towards low productivity firms, and it documents that these firms are less intensive in educated workers. Section 3 constructs a counterfactual earnings distribution in the absence of misallocation, and shows that the returns to education would be higher. Section 4 extends the analysis to consider a counterfactual path for the returns to education. Section 5 concludes.

2. Misallocation Favors Low-Productivity Firms Intensive in Unskilled Workers

In the absence of distortions, individuals efficiently distribute themselves between entrepreneurs, employees, and self-employment. In turn, entrepreneurs hire an efficient number of employees given their abilities. The resulting distribution of individuals across occupations and the distribution of firms across sizes, maximize the productivity of the economy and the returns to the factors of production. Conversely, in the presence of distortions, the distributions of individuals across occupations and of firms across sizes are suboptimal. Individuals who should be employees are entrepreneurs (or vice versa), while firms are larger (or smaller) than they should be, given their underlying productivity. The implied misallocation of capital and labor lowers aggregate productivity and distorts the returns to factors (Restuccia and Rogerson 2008; Hsieh and Klenow 2009).

The size and Type Distribution of Firms

Given the focus on firm behavior as a key determinant of workers' earnings, we begin classifying them by size and type. The size classification is straightforward, given by the number of workers in each firm, regardless of their educational level. Very small firms are arbitrarily labeled as those having 1 to 5 workers, small firms as those having 6 to 10, medium as those employing 11 to 50, and large firms as those having 51 or more employees.

To consider the type classification, it is essential to refer to Mexico's Constitution and laws, which make a central distinction between salaried and nonsalaried workers. The former are dependent employees of a firm, performing a task in exchange for a fixed payment per unit of time, a wage. The latter are associated with firms, but not subordinated to them, and their remuneration can be per unit of output produced or sold, or result from profit-sharing agreements, but it is not a wage. This distinction is critical, because firms that hire salaried workers are obligated to pay for their workers' social insurance benefits, and are also subject to regulations on dismissal, minimum wages, and unions. Firms that associate with nonsalaried workers, on the other hand, are not bound by these regulations; see Appendix S1 for more details and Levy (2008).

If the Law in Mexico was fully observed, there would be three types of firms, depending on their contractual structure: firms hiring only salaried workers, firms mixing salaried and nonsalaried workers, and firms associating only with nonsalaried workers. But as documented elsewhere, this is not the case, because many firms hire salaried workers but fail to comply with applicable labor and social insurance regulations (Levy 2008; Busso, Fazio and Levy 2012). This results in a fourth type: firms hiring salaried workers illegally.

Following standard practice, we associate the formal and informal labels to firms based on the observance of labor and social insurance regulations: formal and legal firms (hiring only salaried

workers and complying with applicable laws); mixed firms (with salaried and non-salaried workers partly complying with these laws); informal and legal firms (associating only with nonsalaried workers); and informal and illegal firms (hiring salaried workers, but violating applicable laws). In parallel, formal workers are salaried employees covered by labor and social insurance regulations; all other workers, whether salaried or nonsalaried, are informal.⁴

While the formal-informal nomenclature is used throughout the paper, we are concerned here with how firms behave, not with how firms are labeled. Specifically, the focus is on whether firms demand workers with more or less educational attainment, and whether they offer jobs with coverage of labor and social insurance regulations.

It is useful to make a few observations on the distribution of firms and employment in 2008. First, 90 percent of all firms are small (up to five workers), and 90.3 percent are informal (legal and illegal). Second, while informal firms have, on average, 2.8 workers, they account for more than half of all employment captured in the census (57.3 percent). Third, larger firms have proportionately more salaried workers than smaller ones: firms with only salaried workers, formal and informal, are only 25.6 percent of all firms, but employ 38.7 percent of all workers. Finally, within firms that only have salaried contracts, formal ones are larger than informal ones: on average, 27.8 vs. 4.0 workers⁵. Table 1 summarizes these observations.

⁴ The existence of mixed and illegal informal firms indicates that some firms are exploiting two different margins of informality: an intensive margin (hiring some workers illegally) and an extensive margin (not registering the firm at all and thus hiring all workers illegally). This is consistent with the findings of Ulyssea (2018) for Brazil, who also finds evidence of firms' informal behavior along these margins. Additionally, however, our paper documents one aspect of firm informality not associated with these two margins, where firms employ nonsalaried workers, and are informal but legal.

⁵ Using the Economic Census for 1993, 2003, 2008 and 2013, Levy (2018) shows that the size distribution of firms in Mexico has been practically constant over this 15-year period. Unfortunately, the 1993 Economic Census is not available, which would help to assess the impact that the North American Free Trade Agreement (NAFTA), which started in 1994, had on the size distribution. Other analysis have shown that trade liberalization has changed this distribution. For instance, Nataraj (2011) shows that India's trade liberalization in the 1990s increased average firm size. Since our data begin in 1998, it already incorporates the effects of Mexico's main liberalization episode.

Table 1. Size and Type Distribution of Firms and Employment, 2008 (shares)

	<i>Formal and legal</i>	<i>Mixed</i>	<i>Informal and legal</i>	<i>Informal and illegal</i>	<i>Total</i>
<i>Firms</i>					
1–5	1.22	3.65	65.84	19.29	90.00
6–10	0.69	1.54	1.10	2.43	5.77
11–50	0.84	1.20	0.55	0.87	3.46
50+	0.25	0.29	0.18	0.05	0.77
Total	3.00	6.69	67.67	22.64	100.00
<i>Workers</i>					
1–5	0.83	2.65	25.49	11.80	40.77
6–10	1.17	2.56	1.73	3.90	9.36
11–50	3.98	5.45	2.50	3.49	15.42
50+	12.56	13.47	7.41	13.47	34.44
Total	18.54	24.14	37.12	20.20	100.00

Source: Author’s computations based on Economic Census data.

Misallocation is High and Persistent in Mexico

Misallocation in Mexico is significant, as evidenced by the fact that the marginal revenue product of capital and the labor of firms that produce similar goods are different. (See appendix S1 for definitions and for descriptions of the data.) In fact, misallocation is higher in Mexico than in the United States or in other Latin American countries with comparable data (Busso, Madrigal, and Pagés 2013). While there is debate over the exact nature of the distortions that induce this phenomenon, three results are robust: distortions result in large productivity losses; they operate in the direction of allocating too much capital and labor to low-productivity firms that are less intensive in more educated workers; and they are persistent through time.

Following Hsieh and Klenow (2009), total revenue productivity, $TFPR_{is}$, is defined as the value produced by firm i in sector s with one peso of capital and labor (a weighted average of the marginal revenue products of the labor and capital used by the firm). In the absence of any

distortions, revenue productivity would be the same for all firms in a sector and across all sectors. This implies that the greater the dispersion of TFPR, the greater the degree of misallocation. Table 2 presents three measures of dispersion of TFPR.⁶

Table 2. Measures of Dispersion of TFPR, 1998–2013

	<i>1998</i>	<i>2003</i>	<i>2008</i>	<i>2013</i>
Standard deviation	0.95	0.98	1.08	1.11
p75–p25	1.23	1.25	1.39	1.39
p90–p10	2.39	2.44	2.73	2.82

Source: Author’s computations based on Economic Census data.

Note: pX refers to the X th percentile on the TFPR distribution. TFPR refers to the Total Revenue Productivity

Two facts follow from the analysis of TFPR. First, there is large dispersion in revenue productivity across firms, implying substantial misallocation of resources. For example, in 2013, a firm in the 75th percentile of the revenue productivity distribution was 39 percent more productive than one in the 25th percentile.⁷ Second, over the 15 years considered, this dispersion has persisted (in fact, increased slightly), indicating persistence in the misallocation of capital and labor.

Misallocation Has Specific Patterns Across Firms

Misallocation in Mexico shows specific patterns. To identify these patterns, revenue productivity is compared across firm types. Table 3 shows the results of OLS regressions of the *log* of firm i TFPR in sector s over the *log* of the average TFPR of all firms in that sector. Regressions include

⁶ Computations are done at the six-digit sector level and include firms of all sizes in manufacturing, services, and commerce. There are 672 sectors in 1998, 679 in 2003, 687 in 2008, and 691 in 2013. Comparisons are only made for firms within the same sector. The numbers in the table are averages across all sectors.

⁷ The difference between firms in the 90th and 10th percentile is 182 percent. This compares with a difference of 92 percent in the manufacturing sector of the United States (Syverson 2004). Syverson’s computations are carried out at the four-digit level, and one would expect smaller differences at the six-digit level. The Inter-American Development Bank (IDB 2010) compares the dispersion of TFPR between manufacturing firms in the United States and Mexico, and finds that dispersion is substantially higher in Mexico. Busso, Madrigal, and Pagés (2013) find that dispersion is also higher in Mexico compared with other Latin American countries.

controls for firm size, location, and age; and the excluded category includes formal and legal firms; see Levy and López-Calva (2016) for further details.

With one exception, in all years, formal and legal firms have higher revenue productivity than all other firms implying that, on average, resources are more valuable in these firms. Informal legal firms, which table 1 shows are the majority, are always the least productive of all, and productivity differences between these firms and illegal informal ones are significant.

The fact that informal firms, legal or illegal, have systematically lower revenue productivity than formal ones indicates that the effect of distortions in Mexico is to allocate too many resources to informal firms. And the fact that this result is observed in all periods considered shows that distortions operate systematically in the same direction.

Table 3. Productivity Differences by Firm Type, 1998–2013
(percent, relative to formal legal firms)

	1998	2003	2008	2013
	TFPR	TFPR	TFPR	TFPR
Mixed	0.001 (0.002)	-0.029 (0.002)	-0.032 (0.002)	-0.177 (0.002)
Legal and informal	-0.414 (0.002)	-0.360 (0.002)	-0.401 (0.002)	-0.633 (0.001)
Illegal and informal	-0.120 (0.002)	0.023 (0.002)	-0.162 (0.002)	-0.184 (0.002)
Observations	2,368,471	2,537,348	2,655,551	3,371,272
R-squared	0.091	0.056	0.072	0.072

Source: Author's computations based on Economic Census data.

Note: Standard errors are in brackets. All coefficients are significant at the 99 percent level.

2.4. Formal and Informal Firms Co-exist in the Same Sectors

At times the reference to formal and informal sectors suggests that firms in each sector produce very different goods. This is not the case in Mexico. Formal and informal firms co-exist in the same six-digit sector, and in fact the overlap between firms in the same sector has been increasing, in parallel to the increase in misallocation. In 1998 informal firms (legal and illegal) were the

majority of firms in 34 percent of the 278 six-digit sectors in manufacturing, in 62 percent of the 142 six-digit sectors in commerce, and in 66 percent of the 252 six-digit sectors in services. By 2013 these percentages had increased to 51 in manufacturing, 81 in commerce and 88 in services.

The Schooling Composition of the Demand for Labor Differs Between Formal and Informal Firms

Transportation services in Mexico may be provided by 100 self-employed workers driving their own trucks or by a single firm with 100 salaried employees. In both cases, there will be 100 trucks and 100 drivers, but, in the latter, there will be a need for a logistics engineer doing dispatches and a sales manager. Tortillas can be produced with simple technologies in small establishments with unskilled labor or in large plants that require engineers; the same holds for apparel and food processing, among many manufacturing activities. This is also true for retail commerce, which can take place in small stores employing workers with basic literacy and numeracy or through large supermarket chains employing industrial designers. In general, the complexity of tasks and the division of labor increase with firm size and generate the need for more educated workers. But complexity is not only an issue of size: a small informal firm producing jeans to be sold in a street market will be less likely to need an accountant than a formal firm of the same size selling jeans to a large retailer.

Table 4 uses data from the employment survey to show the distribution of employment by schooling level and firm size and type.⁸ In each cell, the upper number is the share of workers of a given schooling level in the total number of workers in firms of that size; thus, columns add to 100

⁸ If there were no mixed firms, all informal (formal) workers would be employed by informal (formal) firms. Mixed firms complicate this picture. Unfortunately, we cannot use ENOE (Encuesta Nacional de Ocupación y Empleo) data to correct for this problem, and table 4 is the best approximation to differences in the schooling composition of the workforce between formal and informal firms.

percent and reflect the schooling composition of the workforce of firms by firm size and firm type. The lower number is the share of workers of a given educational level in firms of a given size and sector in the total number of workers of that educational level in all firms in that sector; thus, each row adds to 100 percent.

Table 4. Distribution of Employees by Education and Firm Size, 2006

	<i>Informal firms</i>					<i>Formal firms</i>				
	<i>1-5</i>	<i>6-10</i>	<i>11-50</i>	<i>51+</i>	<i>Total</i>	<i>1-5</i>	<i>6-10</i>	<i>11-50</i>	<i>51+</i>	<i>Total</i>
<i>Incomplete Primary</i>										
Column	11	9.3	7.01	4.52	9.58	6.79	4.26	3.95	2.58	3.35
Row	72.1	12	11.9	3.98	100	11.8	9.14	34.8	44.2	100
<i>Complete Primary</i>										
Column	23.6	16.4	13.2	8.65	19.8	11	11.1	11.2	10.9	11
Row	75.2	10.3	10.9	3.69	100	5.8	7.3	30.2	56.7	100
<i>Incomplete Junior High</i>										
<i>High</i>										
Column	6.02	6.08	6.13	3.08	5.79	2.43	3.18	2.92	2.64	2.75
Row	65.3	13	17.2	4.48	100	5.15	8.33	31.4	55.1	100
<i>Complete Junior High</i>										
Column	31	25.8	25	19.8	28.4	22.1	21.8	20.8	24.9	23.3
Row	68.6	11.3	14.3	5.86	100	5.52	6.74	26.4	61.3	100
<i>Incomplete Senior High</i>										
<i>High</i>										
Column	11.6	12.6	13	14.3	12.2	19.1	18.6	16.1	17.1	17
Row	60	12.8	17.3	9.87	100	6.55	7.87	28	57.6	100
<i>Complete Senior High</i>										
Column	9.27	11.2	11.4	14.5	10.3	17.1	16.1	17.5	14.2	15.5
Row	56.7	13.5	17.9	11.9	100	6.45	7.48	33.5	52.6	100
<i>University</i>										
Column	7.52	18.5	24.3	35.2	13.9	21.4	24.9	27.5	27.7	27.1
Row	33.9	16.5	28.3	21.3	100	4.6	6.62	30	58.8	100
<i>Total</i>										
Column	100	100	100	100	100	100	100	100	100	100
Row	62.9	12.4	16.2	8.43	100	5.82	7.2	29.6	57.4	100

Source: Authors' computations based on Economic Census and Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo data.

Note: *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Three facts are worth highlighting. First, considering the totals for each sector, formal firms are found to be more intensive in educated workers: 42.6 percent of their workforce has at least completed high school, while 14.3 percent have, at most, completed primary school, in contrast to 24.1 percent and 29.4 percent for informal firms, respectively. Second, these patterns hold controlling for firm size. Thus, for instance, in informal firms with up to five workers, an average

16.8 percent of their workforce has at least a high school education, while formal ones have more than double this share (38.5 percent). Third, for any educational level, the distribution of workers across firm size differs, reflecting that formal firms are on average larger. Thus, 72.1 percent of all informal workers with incomplete primary are employed in firms of up to five workers vs. 11.8 percent of all formal workers at the same schooling level. For workers with university education, the corresponding numbers are 33.9 percent and 4.6 percent.

3. Misallocation Lowers the Returns to Education

Our hypothesis is that, in the absence of misallocation, the wages of workers with more education would increase relative to those with less education. To test it, descriptive statistics on trends in educational attainment, labor informality and the returns to education are provided first. Next, the sample of workers on which the econometric analysis focuses is described. The impact of observable firm characteristics on workers earnings is then identified. Finally, these results are used to construct a counterfactual earnings distribution in the absence of misallocation.

The Schooling of the Labor Force Has Increased

For the purpose of this analysis the working-age population (WAP) is defined as all persons 18 years of age or older, and the economically active population (EAP) is defined as the subset of the WAP who participate in the labor market. The WAP is taken as the potential supply of labor, determined by demographics and schooling investments, and the EAP as the observed supply (=demand), given the participation rate of each schooling group. For our econometric analysis we focus on a sample of workers consisting of private sector employees between 18 and 65 years of age, living in localities of 100,000 inhabitants or more, and working between 30 and 48 hours a week. This group represents approximately 20 percent of the EAP and is the most urbanized and

educated segment of Mexico’s labor force. Table 5 presents growth rates by schooling levels over the 1996-2015 period.

Table 5. Growth Rates of the Labor Force by Schooling Levels, 1996–2015

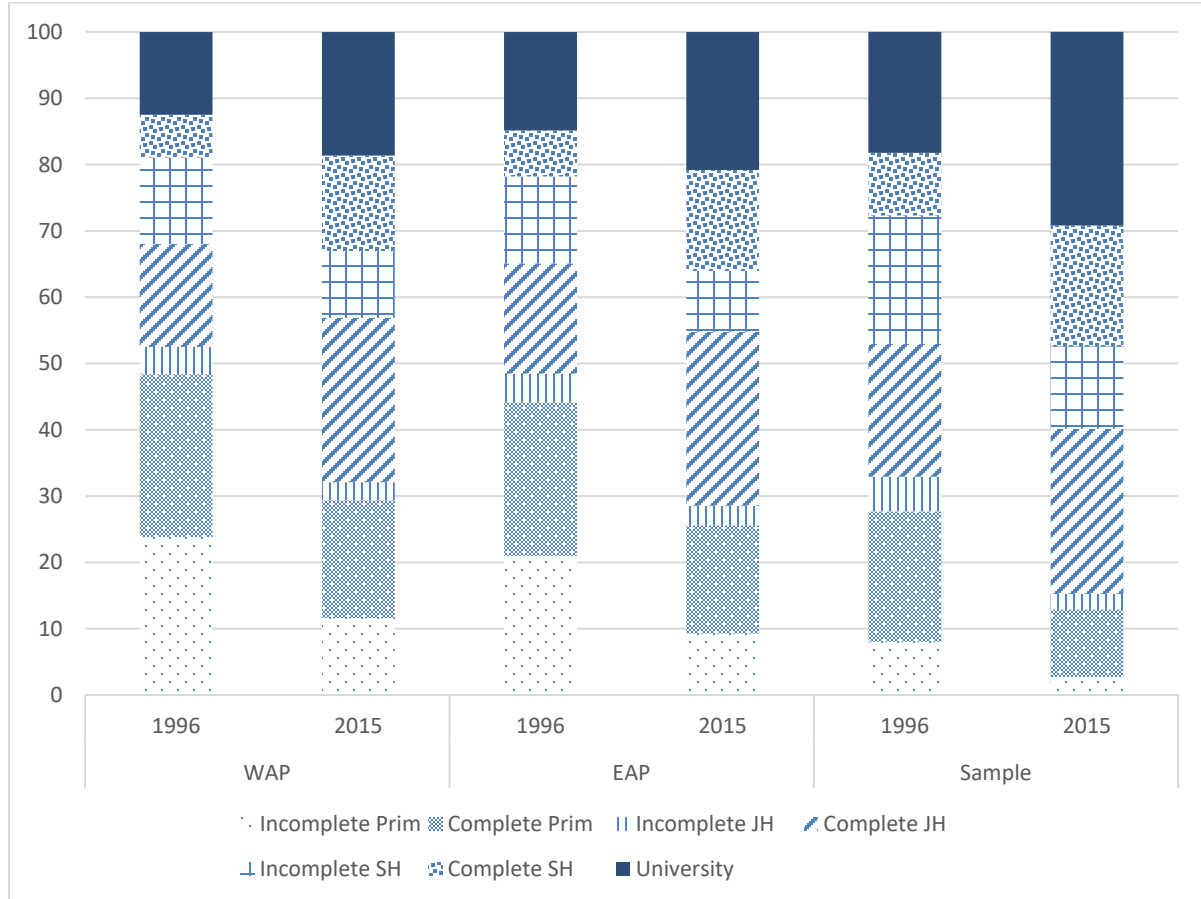
	<i>Annual average growth rate</i>		
	<i>WAP</i>	<i>EAP</i>	<i>Sample</i>
Incomplete Primary	-1.22	-1.68	-3.39
Complete Primary	0.85	0.81	-1.12
Incomplete Junior High	0.44	0.45	-1.99
Complete Junior High	5.03	5.03	3.45
Incomplete Senior High	1.19	0.75	0.00
Complete Senior High	6.18	6.16	5.77
University	4.67	4.39	4.79
All	2.20	2.31	2.32
Years of schooling	—	—	—

Source: Authors’ computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *WAP* (Working-age Population) includes all persons 18 years of age or older. *EAP* (Economically Active Population) is a subset of *WAP* who participate in the labor market. *Sample* population consists of private sector employees between 18 and 65 years of age, living in localities of 100,000 inhabitants or more, and working between 30 and 48 hours a week.

Four facts are highlighted from the resulting changes in the schooling composition of the labor force over the same period. First, there is a substantive change in the schooling composition of the WAP: in 1996, 48.4 percent had at most completed primary school, and only 18.9 percent had at least completed high school; by 2015, these shares changed to 29.3 and 33.0 percent, respectively. The average years of schooling are 4.7, 7.8, and 9.8 years respectively for WAP, EAP, and the sample in 1996. By 2015, the years have increased to 9.2, 9.8, and 11.5. (See Figure 1).

Figure 1. Schooling Composition of the Labor Force, 1996-2015



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Prim*, *JH*, and *SH* refer to Primary, Junior High, and Senior High, respectively. *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Second, the growth rates of the EAP largely mimic those of the WAP, in contrast with the sample, where employment of workers with incomplete junior high or less schooling falls in absolute terms. Third, as a result, the schooling composition of the sample changes rapidly: in 1996, the share with at most primary education equaled that with at least high school (27.7 percent); by 2015, the first share had fallen to 12.9 percent, and the second risen to 47.4 percent. Finally, the sample has more years of schooling: in 2015, workers who had, at most, completed primary school

represented 25.6 percent of the EAP, but only 12.9 percent of our sample; at the other extreme, workers who had at least completed high school were 36.0 percent of the EAP, but 47.3 percent of our sample.

Descriptive statistics for the sample of workers are presented. Average age increases by about four years over the 20-year span. Formal employees are older than informal ones, but the differences are minor, and they narrow over time; they also work slightly more hours than informal ones (but the differences are small). The share of women in the sample increases, reflecting higher female participation rates; over time, the share of women in informal employment increases. Finally, formal employees have more years of schooling than informal ones (Table 6).

Table 6. Characteristics of the Sample of Workers, 1996–2015

	1996			2006			2015		
	All	Formal	Informal	All	Formal	Informal	All	Formal	Informal
Mean age, years	32.17	32.69	30.96	34.58	35.13	33.29	36.46	36.62	36.04
Median age	30	31	28	33	34	31	35	36	34
Hours worked	42.24	42.41	41.85	42.48	42.83	41.66	42.79	43.16	41.84
Years of schooling	9.85	10.30	8.82	10.76	11.39	9.31	11.51	12.14	9.93
Share of women	39.68	39.79	39.39	44.82	44.14	46.41	45.11	43.50	49.03
Formality rate		70.1			70.0			71.6	
<i>Education shares</i>									
Incomplete primary	8.02	5.66	13.55	5.22	3.35	9.59	2.70	1.33	6.18
Complete primary	19.67	18.10	23.35	13.62	11.00	19.77	10.22	7.07	18.18
Incomplete junior high	5.23	4.53	6.85	3.65	2.73	5.79	2.30	1.55	4.21
Complete junior high	20.06	19.80	20.67	24.85	23.34	28.39	24.90	23.41	28.67
Incomplete senior high	19.33	21.44	14.38	15.57	17.04	12.13	12.44	13.06	10.88
Complete senior high	9.45	9.86	8.50	13.93	15.48	10.30	18.23	19.32	15.48
University	18.23	20.60	12.69	23.16	27.06	14.02	29.20	34.26	16.40

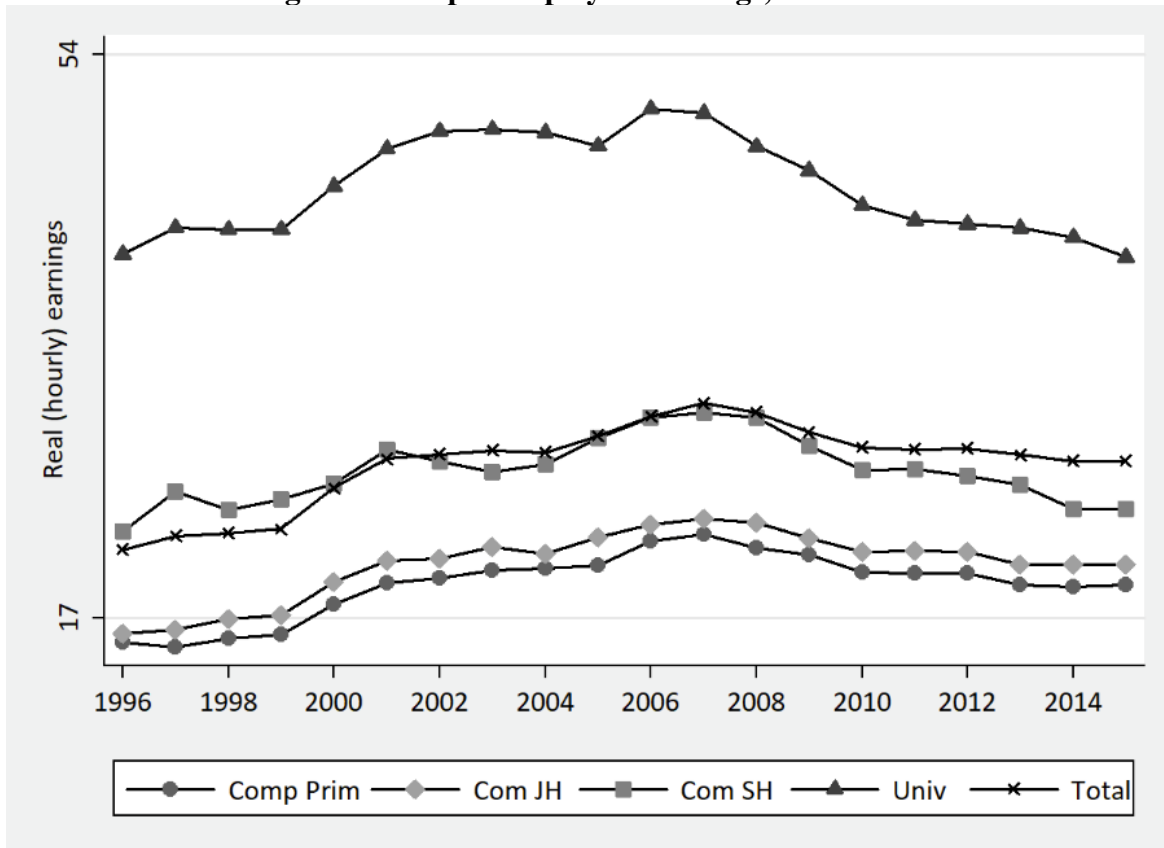
Source: Author's computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Earnings and Labor Informality Are Basically Constant

The analysis shows that earnings for all groups in our sample increase up to 2003, reflecting the recovery from the sharp fall observed in the 1995 financial crisis (Figure 2).⁹ After that, they stagnate for employees who have completed primary and junior high, and fall for those who have completed high school or university education. In fact, by 2015, earnings for the latter group were 7 percent below the 1996 level, and, for the former, the same as in 1996. The contrast with employees who have completed junior high is revealing: as seen in table 5, the supply of persons of this educational group grows more quickly than those with university education. Yet, earnings of employees with completed junior high in 2015 are 26 percent higher relative to 1996. These asymmetries clearly indicate that there are other forces aside from changes in supply determining the behavior of earnings. The earnings path for each educational group (together with changes in their relative shares) results in a decrease in average employee earnings (labeled total in Figure 2), which, by 2015, are the same as in 2000.

⁹ There are no employment surveys for 1992–94, and the available ones for earlier years cannot be compared with ENE-ENOE (Encuesta Nacional de Empleo-Encuesta Nacional de Ocupación y Empleo). To corroborate our statements, in appendix S2 we use data from household surveys to show (a) a sharp fall in earnings in 1995 following that year’s financial crisis; (b) that by 2012 earnings had yet to reach the levels observed in 1994; and (c) a similar trend in earnings after 1996 in the two surveys (contrast Figure 2 with figure S2.3 in appendix S2). Appendix S2 also describes the procedure used to correct for missing observations on earnings.

Figure 2. Sample Employee Earnings, 1996–2015



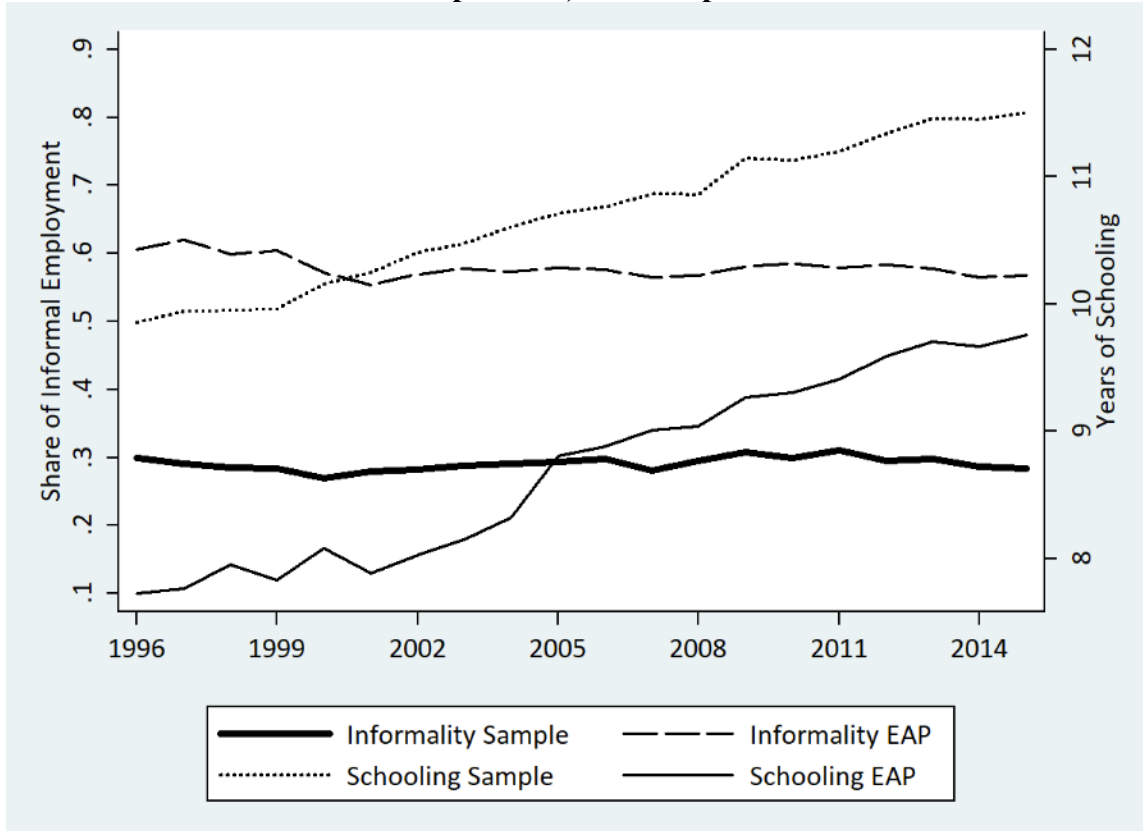
Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Comp prim* refers to complete primary, *Com JH* refers to complete Junior High, and *Com SH* refers to complete Senior High.

Figure 3 depicts the paths of the rate of informal employment (on the left axis) and average years of schooling (on the right axis) for our sample and the whole EAP. As expected, the rate is lower for our sample because the self-employed, rural workers, and urban employees working less than 30 hours a week or living in relatively less urbanized localities are excluded. Both rates are practically constant through time, even though, over this period, the average years of schooling of the EAP increased from 7.8 to 9.8, and those of our sample from 9.8 to 11.5.¹⁰

¹⁰ Levy and Székely (2016) use data from Mexico's household surveys to follow separate cohorts of workers between 1989 and 2012. They find that younger cohorts have more years of schooling than older ones, but that their rates of informality are the same; differently put, the paths in Figure 3 are not a result of changes in the age composition of the labor force.

Figure 3. Years of schooling and share of informal employment, Economically Active Population, and Sample



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Informality* refers to the rate of informal employment within each population. *Schooling* refers to the average years of schooling for each population. *Sample* population consists of private sector employees between 18 and 65 years of age, living in localities of 100,000 inhabitants or more, and working between 30 and 48 hours a week. *EAP*, the economically active population, consists of all persons 18 years of age or older who participate in the labor market.

Labor Mobility Is Large and the Returns to Education Have Been Falling

The large mobility of workers of all educational levels across firm sizes and types is a central feature of Mexico's labor market (table 7). A panel of all workers in our sample surveyed in the second quarter of 2006 (2006q2) is constructed, following them through five consecutive quarters. In one extreme, some workers were observed in the four previous quarters and are last surveyed in 2006q2; at the other, some first entered the survey in 2006q2 and were observed for the next four quarters; the rest correspond to intermediate cases with a mix of up to three quarters before

or after 2006q2. Altogether the period spans from 2005q3 to 2007q1, and the number of times that each individual worker changed firm type (from formal to informal or vice versa) or size (from any one of the four firm sizes considered to any of the other three) is recorded.

Table 7: Mobility of individual workers across firm size and type, 2005q3 – 2007q1

	<i>(shares)</i>	
	<i>Firm type</i>	<i>Firm size</i>
<i>Education</i>	<i>Change*</i>	<i>Change*</i>
<i>Incomplete Primary</i>	15.26	31.86
<i>Complete Primary</i>	16.41	33.93
<i>Incomplete Junior High</i>	20.09	36.84
<i>Complete Junior High</i>	19.33	38.98
<i>Incomplete Senior High</i>	21.02	44.83
<i>Complete Senior High</i>	20.3	45.99
<i>University</i>	21.15	47.67

Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

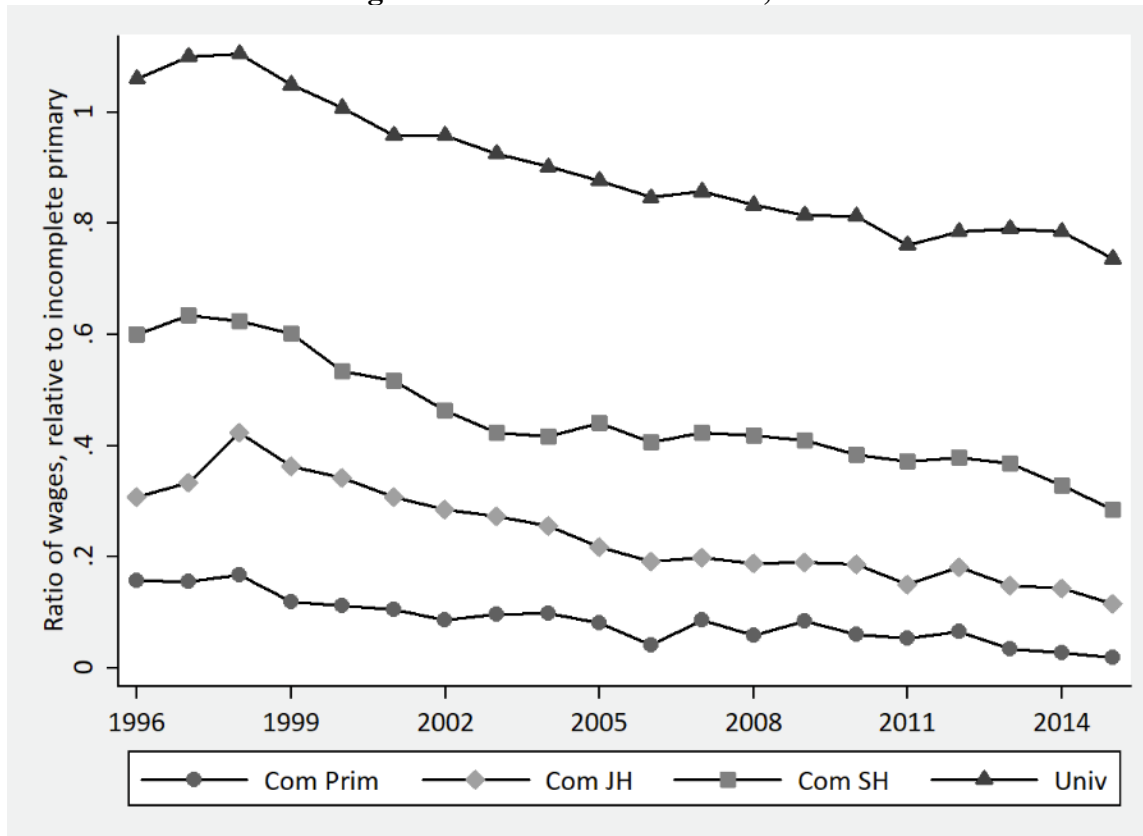
Note: *Change refers to at least one movement over one year. *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Mobility is very high even for such a brief period of time, and is in fact slightly increasing with educational levels. During one year, anywhere between 15 to 21 per cent of workers change firm type. Mobility across firm sizes is even higher, and again is increasing with educational level. This level of mobility, consistent with analysis presented elsewhere for Mexico (e.g., Levy 2008, Maloney 2004), allows us to have more robust estimates of the changes in the size and type distribution of firms on workers' earnings obtained from the counterfactual exercise. The

transitions between formal and informal status are not unique to Mexico or even Latin America. For instance, McCaig and Pavcnik (2015) document that, in Vietnam, there are also significant transitions, depending on workers age, education, gender, and rural urban locations.

Figure 4 depicts the returns to education over 1996–2015. The points are the coefficients for each year from an ordinary least squares regression where the excluded category is incomplete primary education, with controls for age, experience, and municipality of residence. Over the 20-year period, returns are falling for all groups, although the declines are more pronounced for employees with more years of education. Clearly, for each year, the returns to education are a weighted average of the returns to education of workers employed in formal and informal firms. In table 10 (below), we separate these returns for 2006 (the same year on which we perform our counterfactual exercise) and show that they are higher when workers are formally employed.

Figure 4. Returns to Education, 1996–2015



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Com* refers to completed. *Prim*, *JH*, *SH*, and *Univ* refer to primary, junior high, senior high, and university, respectively. *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Worker Earnings and Observable Firm Characteristics

To measure the importance of observable firm characteristics on worker earnings, a within-educational group earnings function is estimated. Firm characteristics and the observable characteristics of individuals for each educational level are controlled for. Because the regressions are carried out by educational group, the potential bias induced by unobserved individual characteristics associated with educational choices is controlled for. However, the characteristics

of returns to firm may be biased upward if there is unobserved ability among workers that is correlated with worker selection into different firms (say, more able workers sorted into larger firms). To address this problem, we take advantage of the fact that the observed mobility of workers across firm sizes and types is sufficiently high to allow identification using individual fixed effects.¹¹ The panel regression with worker fixed effects, with all earnings measured per hour, is as follows:

$$\text{Log } y_{it,e} = \mathbf{D}_{i,e} + \beta_e \mathbf{X}_{it,e} + \gamma_e \mathbf{Z}_{it,e} + \partial_e \mathbf{FS}_{it,e} + \epsilon_{it,e} \quad (1)$$

where $y_{it,e}$ is earnings for worker i of educational level e at quarter t ; $\mathbf{D}_{i,e}$ is an individual fixed effect for individual i with education level e ; $\mathbf{X}_{it,e}$ is a vector of characteristics of individual i with educational level e at quarter t (age, experience, gender, location), with associated coefficients β_e ; $\mathbf{Z}_{it,e}$ is a vector of firm characteristics (such as size and written contract) for individual i with education level e at quarter t ; and γ_e is a vector of associated coefficients.¹² A dummy for whether the worker is formal or informal and a set of interactions between formality and observable firm characteristics ($\mathbf{FS}_{it,e}$) are included. Finally, there is an error term, $\epsilon_{it,e}$, which is assumed to be uncorrelated with $\mathbf{Z}_{it,e}$ and $\mathbf{FS}_{it,e}$, that implies that firm size and firm characteristics are not correlated with unobserved time-varying individual characteristics.¹³ Results are shown in table 8.

¹¹ To test for the existence of this bias, we ran the same regressions in (1) without workers' fixed effects, and, indeed, the returns to firm size are about 15 percent to 20 percent higher, consistent with the hypothesis of unobserved time-invariant characteristics associated with workers sorting into firms of different sizes.

¹² An important firm characteristic is payment of an end-of-the-year bonus to workers (*aguinaldo*). Levy and López-Calva (2016) provide data on these payments. In general, rates of bonus payments are higher in formal firms than in informal firms; controlling for type, they are higher in larger vs. smaller firms; and, controlling for firm size and type, they are higher for more educated workers. Bonus payments are excluded from both sides of regression (1), but are then added to worker earnings in proportion to the rates observed for each educational level and firm size.

¹³ This assumption may still imply a bias if there were time-varying individual characteristics associated with the selection of workers into firms (such as training or investment in skills). The best way to deal with this problem would

Table 8. Earnings Regressions by Education Level (anchor 2006q2, panel 2005q3–2007q1)
excluded firm size [0-5]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Inc Prim</i>	<i>Com Prim</i>	<i>Inc JH</i>	<i>Com JH</i>	<i>Inc SH</i>	<i>Com SH</i>	<i>Univ</i>
Age	0.0368*** (52.19)	0.0382*** (89.27)	0.103*** (62.00)	0.0143*** (47.15)	0.0688*** (127.57)	0.000694 (1.21)	0.0431*** (101.47)
Age * age	-0.000539*** (-66.80)	-0.000496*** (-94.28)	-0.00131*** (-54.43)	-0.000137*** (-32.66)	-0.000890*** (-127.02)	0.000164*** (21.20)	-0.000480*** (-91.35)
6 - 10 Workers	0.108*** (85.39)	0.0482*** (53.04)	-0.0593*** (-42.03)	0.0197*** (26.54)	0.0884*** (71.11)	0.0870*** (62.12)	0.0695*** (48.52)
11 - 50 Workers	0.130*** (99.33)	0.0569*** (58.84)	0.0883*** (49.80)	-0.00508*** (-6.54)	0.108*** (89.70)	0.0217*** (15.15)	0.156*** (112.40)
51+ Workers	0.0432*** (19.62)	-0.0169*** (-10.64)	0.198*** (66.22)	0.0122*** (10.36)	0.0918*** (58.02)	0.00625*** (3.34)	0.212*** (134.81)
Written Contract	-0.0884*** (-43.03)	0.00702*** (6.41)	-0.0516*** (-28.32)	0.0467*** (57.39)	0.0503*** (46.10)	0.0992*** (74.71)	0.0844*** (79.84)
Formal	0.00577*** (4.11)	0.0128*** (13.92)	-0.00382*** (-2.05)	0.0477*** (61.16)	0.0857*** (72.43)	0.0371*** (26.41)	0.126*** (84.73)
Formal * Contract	0.117*** (51.68)	-0.0124*** (-10.08)	0.0765*** (35.33)	-0.0422*** (-44.97)	-0.0319*** (-25.03)	-0.0707*** (-45.62)	-0.0143*** (-11.02)
Formal * [6–10]	-0.0310*** (-15.71)	-0.0610*** (-46.88)	0.0550*** (24.57)	-0.0216*** (-20.30)	-0.0455*** (-29.67)	-0.0256*** (-14.60)	0.00335 (1.88)
Formal * [11–50]	-0.0756*** (-40.64)	-0.0666*** (-54.07)	-0.0674*** (-28.29)	0.0530*** (52.39)	-0.0300*** (-20.92)	0.0704*** (41.86)	-0.0215*** (-12.98)
Formal * [51+]	0.0378*** (14.81)	0.0492*** (27.87)	-0.156*** (-45.48)	0.0432*** (32.27)	-0.0158*** (-9.17)	0.125*** (60.57)	-0.00584*** (-3.27)
Observations	1453795	3991436	1090511	7354974	5005826	4181318	7057757
Adjusted R^2	0.746	0.718	0.755	0.708	0.733	0.746	0.704
Controls	Munic	Munic	Munic	Munic	Munic	Munic	Munic

Source: Authors' analysis based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: t statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The contrasts across educational levels are sharp. For workers with incomplete primary education, the effect on earnings of being employed by larger firms compared with small ones [0–5] is small; the difference with being employed in a small [6–10], medium [11–50] or large [50+] one is 7.7 percent, 5.4 percent, and 8.1 percent, respectively (and, in the last case, not statistically significant). This is in contrast with workers with university education: relative to small firms, the difference of being employed by a small, medium, or large firm is 7.3 percent, 13.4 percent, and 20.6 percent, respectively. Comparable results hold contrasting workers with completed primary education with those that have completed high school: for the former, the difference between working for a large firm relative to a small one is 3.2 percent; for the latter, 13.1. More generally, there is a premium

be to match employer-employee data over time, which is not available for Mexico. The magnitude of this bias, if any -particularly in terms of its relative importance across educational levels- should not affect the counterfactual exercise.

for working with larger firms and that premium increases with educational level (net of interactions between firm size and formality) (See Table 8).

Having a written contract also matters for earnings, also differentially by educational level. In the case of workers with incomplete primary, the net effect is a 2.8 percent increase in earnings; for those with university education, 7 percent. A similar result holds comparing workers with complete primary vs. complete high school: 0 percent vs. 7.3 percent.

Finally, recall that formal firms have higher productivity than informal ones (table 3). Because firm type cannot be identified in the Encuesta Nacional de Empleo/Encuesta Nacional de Ocupación y Empleo (ENE-ENOE) data, we add a dummy for formality, which is positive and significant for all educational levels (except incomplete junior high). Thus, controlling for firm attributes such as size and written contract, other dimensions of firm formality impact positively on worker earnings, increasing them by between 0.5 percent and 12.6 percent.

Earnings Distributions in the Absence of Misallocation

Now counterfactual exercises to identify the impact of misallocation on worker earnings are performed. In principle, we would like to identify the size and type distribution of firms in the absence of misallocation. Leal (2014) calibrates a dynamic general equilibrium model for Mexico whereby distortions induce firms to separate into formal and informal, each with different size distributions and with larger average size for formal firms (as in table 1). He then finds that, in the absence of distortions, there would be no informal firms, while the average size of formal firms would increase.

We use these results and operationalize the no-misallocation scenario by constructing a hypothetical earnings distribution for informal workers if they were employed by formal firms,

but assuming there is no change in the size distribution of formal firms. More precisely, the question is this: What would be the earnings of informal workers if, given their age, education, gender, location, and individual unobservable characteristics (as captured by the error term in equation [1]), they were distributed across firm sizes in the same proportions as formal workers of the same educational level and with the same proportions of written contracts and bonus payments? No changes are made to the earnings of formally employed workers or to the size distribution of formal firms. Thus, our interpretation of the effects of misallocation is fairly narrow in scope and most likely underestimates the effects on earnings of eliminating the misallocation, as it is limited only to measuring the effects on earnings of having informal workers employed by the same firms that employ formal ones.

We proceed in three stages. First, for each educational level, workers in informal firms are redistributed to replicate the proportions across the sizes in which workers of the same educational level are employed in formal firms. Consider, for example, workers with incomplete primary education. In the formal sector, 44.2 percent are in large firms, and 11.8 percent in small ones, in contrast to 3.9 percent and 72.1 percent, respectively, in the informal sector; furthermore, 34.8 percent of all formally employed are in medium firms, and 9.1 percent in small ones, versus 11.9 percent and 12.0 percent, respectively, of all informally employed (Table 6). To replicate the proportions observed in the formal sector, we randomly subtract 60.3 percent ($= 72.1 \text{ percent} - 11.8 \text{ percent}$) of informally employed workers in small firms and 2.9 percent ($= 12.0 \text{ percent} - 9.1 \text{ percent}$) from small firms and randomly assign them to medium and large firms to reach the desired proportions (34.8 percent and 44.2 percent, respectively). The redistribution is only done at the margin, that is, thus only excess workers from firm sizes that employ proportionately more than

formal firms of the same sizes are removed and placed in firm sizes that employ proportionately less. This exercise is performed for each educational level separately.

Second, for each individual worker of a given educational level that is moved, the regression coefficients from table 7 are used to impute the implied change in earnings derived from working for a different size firm. To capture the effects of changing firm type, the value of the dummy for formality is imputed to informal workers, and the effect of a written contract is randomly assigned to those that did not have one to reach the same proportions as formal workers of that educational level and firm size. All other determinants of earnings are left intact, including the unobservable characteristics of workers, as captured by the error term.

The final stage consists of adding the value of the pro rata hourly amount of the yearly bonus to formal workers as observed in the data for each individual case and of randomly adding it to informal workers who did not receive one. This is to achieve—by firm size and educational level—the proportions observed in the formal sector. This exercise results in two earnings distributions: one for what we now label as already formal workers (which did not change from the one observed in the data) and one for newly formal workers (which did change).

Table 9 reports the observed and counterfactual mean and standard deviation of the distribution of worker earnings by educational level and the observed share of formal and informal workers. By construction, columns 2 and 5 are the same, while columns 3 and 6 show the observed and counterfactual earnings of informal workers and newly formal workers, respectively. Consider first the results for newly formal workers of all educational levels, shown in the last row. Mean earnings increase by 17 percent, a number that synthesizes the cost to workers of being employed in the informal rather than in the formal sector; this cost is the misallocation penalty. Critically,

this penalty does not result from workers lacking education or abilities; it results from distortions that allocate too many resources to low-productivity informal firms. Second and in line with our hypothesis, the penalty is not evenly distributed across educational levels: for those with incomplete primary, it is 9 percent, but, for those with university education, it is significantly higher, 29 percent. Similarly, for those with complete primary education, the penalty is 3 percent, while, for those who have completed high school, it is 17 percent.

The counterfactual mean of earnings for all workers by educational level is shown in column 4. Because this is simply the weighted average of earnings for newly formal workers and already formal workers, the increase is lower than that of newly formal workers by themselves, given that earnings of already formal workers did not change. Overall, the mean increase is 4 percent and is similar across educational levels. This result follows from the fact that the proportion of informally to formally employed workers falls as educational levels increase. Thus, the 9.0 percent increase in earnings for informally employed workers with incomplete primary benefits 56.7 percent of all workers with that level of education, while the 29.0 percent increase in earnings for informally employed workers with university education applies to only 19.9 percent of these workers. Recall that the share of formal workers in our sample is 70 percent, while for the occupied labor force as a whole it is only 42.3 percent, so we can speculate that, in the informality-free scenario, the mean increase in earnings for all the occupied labor force would be higher than the one obtained here for our sample.

Table 9. Observed and Counterfactual Earnings (2006q2 anchor)

	<i>Observed</i>			<i>Counterfactual</i>			<i>(6)/(3)</i>	<i>(4)/(1)</i>	<i>(2)/(3)</i>	<i>(5)/(6)</i>
	<i>All</i> <i>(1)</i>	<i>Formal</i> <i>(2)</i>	<i>Informal</i> <i>(3)</i>	<i>All</i> <i>(4)</i>	<i>Formal</i> <i>(5)</i>	<i>Informal</i> <i>(6)</i>				
Incomplete Primary										
Mean	21.4	23.7	19.6	22.4	23.7	21.4	1.09	1.05	1.21	1.11
SD	9.99	11.2	8.5	10.2	11.2	9.22
Shares	100	43.2	56.7			
Complete Primary										
Mean	22.6	24.6	20	22.8	24.6	20.6	1.03	1.01	1.23	1.2
SD	11.5	12.4	9.73	11.6	12.4	10
Shares	100	55.5	44.5			
Incomplete Junior High										
Mean	23.1	25.2	20.7	24.6	25.2	23.9	1.16	1.06	1.22	1.05
SD	11.8	12.2	10.7	12.4	12.2	12.7
Shares	100	53.9	46.1			
Complete Junior High										
Mean	24.1	25.9	20.4	25.3	25.9	24	1.18	1.05	1.27	1.08
SD	12.4	13	10.4	12.8	13	12.4
Shares	100	67.0	33.0			
Incomplete Senior High										
Mean	29.1	31.6	21.2	30	31.6	25	1.18	1.03	1.49	1.26
SD	17.5	18.1	13	17.6	18.1	15.2
Shares	100	75.3	24.6			
Complete Senior High										
Mean	31.5	33.9	23.8	32.5	33.9	27.9	1.17	1.03	1.42	1.22
SD	20	20.4	16.6	20.4	20.4	19.7
Shares	100	76.4	23.5			
University										
Mean	51.1	55.2	34	53	55.2	43.9	1.29	1.04	1.62	1.26
SD	31.5	31.7	23.5	31.7	31.7	29.9
Shares	100	80.1	19.9			
Total										
Mean	32	35.9	22.8	33.1	35.9	26.6	1.17	1.04	1.58	1.35
SD	23	24.7	14.8	23.4	24.7	18.5
Shares	100	69.7	30.3			

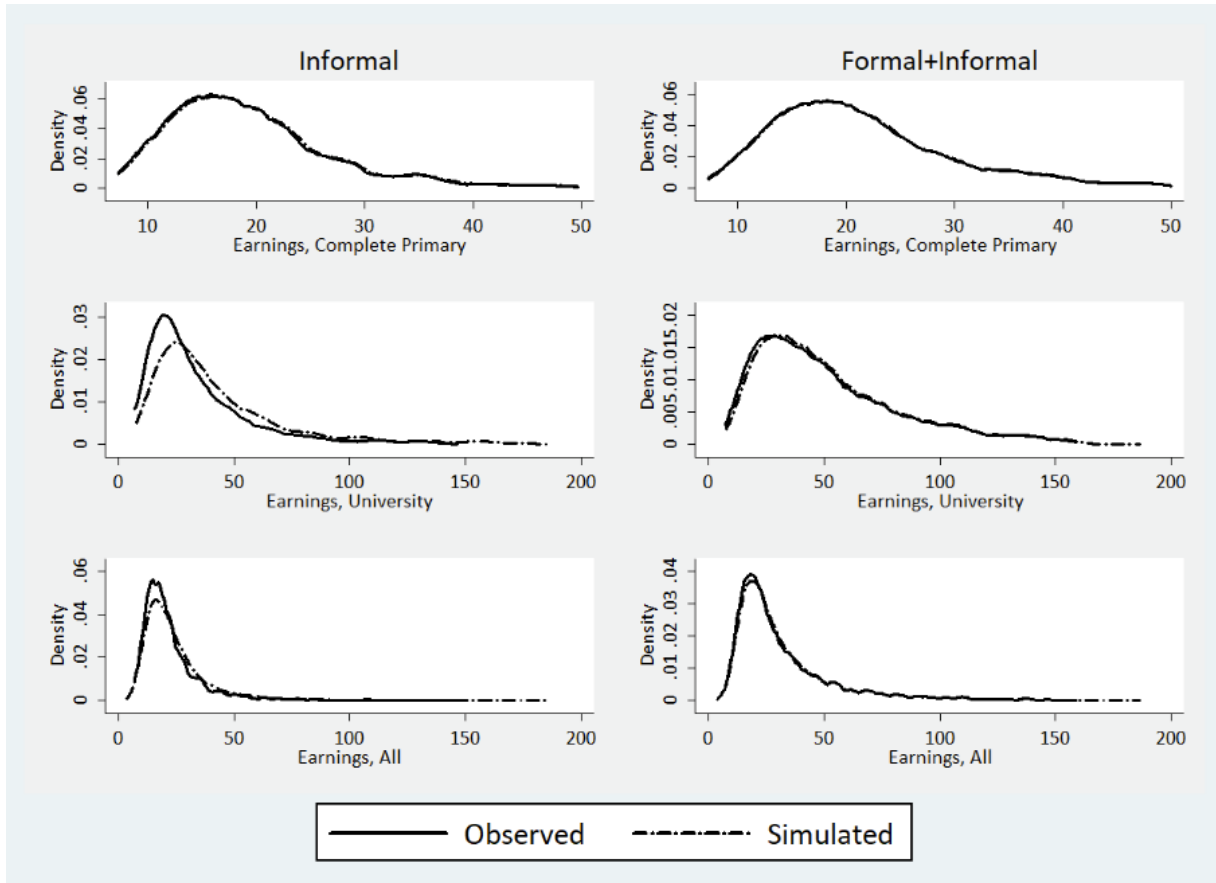
Source: Authors' analysis based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

The last two columns in table 9 show the formal-informal earnings differentials. The observed one is positive and increasing in educational levels. The differential narrows in the counterfactual scenario and proportionately more for workers with more education: for workers with incomplete primary education, it falls from 1.21 to 1.11 percent vs. 1.62 to 1.26 percent for those with university education (and from 1.58 percent to 1.35 percent for the full sample). The counterfactual earnings differentials are interpreted as a result of other differences in firm characteristics that are, in principle, observable (such as unionization or job tenure), but that are not considered in our earnings regressions because they are not available for every year in the ENE-ENOE data.

Figure 5 depicts the observed and counterfactual earnings distributions of employees with complete primary, university education, and all educational categories together (note that the scales vary). Panels on the left refer to informal employees, while panels on the right refer to formal and informal ones. Considering only the former, it is clear that the mean and the dispersion of earnings increase in the counterfactual scenario; furthermore, the contrast between the two educational levels is notable. These contrasts are reduced considering the sum of formal and informal employees, in the right-hand panels, because the share of informal employees with complete primary in all employees of that educational level is higher than that of employees with university education. Still, the mean and the dispersion are larger in the counterfactual scenario for both groups. This holds for all educational groups combined, as can be seen in the bottom graphs (and verified in table 8).

Figure 5. Observed and Counterfactual Earnings Distributions, 2006



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Table 10 presents the estimates of the returns to education in the observed and counterfactual earnings distribution. The most significant result, of course, is the significant increase in the returns to education for informal workers, which again highlights the point that, if workers are employed by informal firms, accumulating more years of education is substantially less valuable than if they are formally employed. The returns to education for informal workers are still below those for formal workers. This simply refers to the other side of the coin, whereby the earnings differences with formal workers were narrowed by our counterfactual exercise, but not eliminated.

Table 10. Observed and Counterfactual Returns to Education, 2006q2 anchor

	<i>Observed</i>			<i>Counterfactual</i>	
	<i>All</i>	<i>Formal</i>	<i>Informal</i>	<i>All</i>	<i>Informal, Sim</i>
Age	0.0472*** (32.71)	0.0456*** (25.88)	0.0312*** (13.86)	0.0430*** (30.57)	0.0311*** (13.69)
Age * age	-0.000445*** (-23.16)	-0.000400*** (-17.05)	-0.000342*** (-11.38)	-0.000401*** (-21.39)	-0.000345*** (-11.36)
Complete Primary	0.0732*** (5.63)	0.0691*** (3.71)	0.0213 (1.27)	0.0350** (2.76)	-0.0406* (-2.40)
Incomplete Junior High	0.186*** (10.79)	0.203*** (8.48)	0.0787*** (3.45)	0.196*** (11.63)	0.130*** (5.64)
Complete Junior High	0.219*** (17.74)	0.224*** (12.71)	0.0758*** (4.62)	0.217*** (18.01)	0.146*** (8.81)
Incomplete Senior High	0.362*** (28.16)	0.384*** (21.38)	0.0969*** (5.26)	0.349*** (27.80)	0.175*** (9.40)
Complete Senior High	0.450*** (34.48)	0.459*** (25.32)	0.194*** (10.12)	0.431*** (33.79)	0.263*** (13.62)
University	0.863*** (69.82)	0.880*** (50.49)	0.500*** (27.65)	0.859*** (71.16)	0.668*** (36.56)
Observations	35059	24440	10619	35059	10619
Adjusted R ²	0.340	0.362	0.188	0.346	0.249
Controls	Municipality	Municipality	Municipality	Municipality	Municipality

Source: Authors' analysis based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

General Equilibrium Effects

A counterfactual earnings distribution of informal employees has been constructed assuming they are employed by formal firms. From the point of view of these employees, the exercise ensures that the effects of observable firm characteristics on earnings are the same for them as for formal employees. But it is critical to highlight that, from the point of view of firms, the counterfactual earnings distribution is not an equilibrium outcome. This is because formal and informal firms are not the same: even controlling for size, underlying differences in their production functions result

in differences in the schooling composition of their demand for labor and in their productivity.

Table 4 helps illustrate this point. The redistribution of informally employed workers across firm sizes to reproduce the distribution of those formally employed implies that, for each educational level, the rows in that table in the formal and informal sector are the same. However, the redistribution kept constant the total number of informal workers at each educational level. Thus, despite the redistribution, it is still the case that, of all newly formalized workers, 9.5 percent have incomplete primary, 19.8 complete primary, and 13.9 percent university education. This composition of the workforce differs substantially from the one observed in the formal sector, where only 3.3 percent of workers have incomplete primary, 11 percent complete primary, and 27.1 university education.

The point here is this: if formal firms were to employ informal workers, they would do so in the same proportions in which they employ formal ones and not in the proportions in which informal workers are available. This implies that, relative to the schooling composition of the supply of informal workers, there are either not enough workers with higher education, or too many workers with low education; whatever the viewpoint, in the counterfactual earnings distribution, there would be a mismatch between the labor demand of firms and the supply of workers by educational levels.

To sharpen this point, let N_{fe} and N_{ie} be the number of formal and informal workers of educational level e , and N_f , N_i the total number of formal and informal workers. From the point of view of firms, the issue is that:

$$N_{ie} / N_i \neq N_{fe} / N_f = \alpha_e \quad (2)$$

What would be the adjustment required to the supply of informally employed workers of each schooling category to match the proportions observed in the formal sector? This can be answered taking as given α_e and finding the values of δ_e that solve

$$\delta_e N_{ie} / \sum_e \delta_e N_{ie} = \alpha_e \quad (3)$$

When expanded, (3) is a system of seven linear homogeneous equations of the form $A\delta = 0$, where A is a square matrix of coefficients, and δ and 0 are vectors.¹⁴ To obtain a nontrivial solution, we normalize $\delta_7 = 1$, implying that we measure any excess supply or demand of workers of educational level e relative to workers with university education. The interpretation is straightforward: if $\delta_{e \neq 7} \leq 1$, the number of informal workers of educational level e needs to be reduced so that their share is the same as the one observed in the formal sector (or increased if the inequality is reversed). Differently put, if $\delta_{e \neq 7} \leq 1$, there is an excess supply of workers of that educational level, while $(1 - \delta_e)N_{ie}$ measures the absolute size of the excess supply, and $(1 - \delta_e)N_{ie} / (N_{ie} + N_{je})$ the excess supply relative to the total number of workers of that educational level.

Two results are of interest in the solution to equation (3) (table 11 column 2). First, given the schooling composition of the demand for labor in the formal sector, if all firms in the economy were formal, there would be a 14.4 percent excess supply of workers of all other schooling levels.

¹⁴ The coefficients along the main diagonal of matrix A are $(1 - \alpha_e)N_{ie}$ and along the rows $-\alpha_e N_{ik, (k \neq e)}$ for $k = 1, 2, \dots, 6$.

Second, in relative terms, the excess supply is largest for workers with the least amount of education.¹⁵

Table 11. Excess Supply of Workers by Schooling Category, 2006

	δ_s	N_{is}	$(1-\delta_s)N_{is}$	$N_{is} + N_{fs}$	$(1-\delta_s)N_{is} / (N_{is} + N_{fs})$
Incomplete Primary	0.180	227171	186260	413389	0.451
Complete Primary	0.288	469525	334510	1080167	0.310
Incomplete Junior High	0.243	137191	103819	289111	0.359
Complete Junior High	0.424	673273	387796	1968788	0.197
Incomplete Senior High	0.723	289234	80233	1235346	0.064
Complete Senior High	0.776	243601	54632	1102977	0.049
University	1.000	328109	0	1834499	0
Total		2368104	1147250	7924277	0.144

Source: Authors' analysis based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree. N_{is} refers to the number of informal workers of the educational levels specified in each row. N_{fs} refers to the number of formal workers in each of the educational levels. $(1-\delta_s)N_{is}$ denotes the absolute size of the excess supply of workers with that level of education, and $(1-\delta_s)N_{is}/(N_{is}+N_{fs})$ denotes the excess supply of workers relative to the total number of workers of that educational level.

These results can be interpreted as follows: the counterfactual earnings distribution presented captures the effects that, from the perspective of workers, would be observed if they were all employed by formal firms (table 9 column 4). On the other hand, formal firms would not be willing to employ all of them in the counterfactual distribution (table 11). More precisely, firms would be least willing to employ workers with fewer years of schooling: 45 percent of workers with

¹⁵ Of course, if we had normalized equation (3) with, say, $\delta_1 = 1$, the results would show excess demand for workers of all other educational levels except, by construction, those with incomplete primary education, with excess demand larger for workers with more education. The point here is that what matters are the proportions, which reflect the fact that the isoquants derived from the production functions of formal firms are different from those derived from the production functions of informal firms.

incomplete primary education would be in excess supply, 31 percent of those with completed primary, and so on.

This excess supply would trigger an adjustment, which can take many forms. On one extreme, all would be in quantities, with workers leaving employee employment in the magnitudes shown in the last column of table 11 (through an exogenous drop in participation rates, say), which would imply that the counterfactual earnings distribution is the equilibrium distribution. On the other extreme, it would occur through prices, with earnings of workers of all educational levels falling as necessary, relative to those with university education, to clear excess supply.

This approach does not allow us to quantify the changes in earnings and employment levels required to establish the equilibrium between the supply and demand of workers of each educational level if all firms behaved as formal firms. Nonetheless, there is one clear and critical implication: unless all adjustment occurred through quantities, the earnings of workers with fewer years of education would fall relative to those with more years of education. In turn, this implies that the counterfactual earnings and returns to education shown in tables 9 and 10 would be different. In particular, earnings differentials across schooling levels would be wider, and, correspondingly, the returns to schooling higher. Differently put, the estimates of the changes in the distribution of earnings and the returns to education presented in tables 8 and 9 are a lower bound of the effects of eliminating misallocation.

4. The Path of the Wage Premium, 1996-2015

The previous results are now used to shed some light on the evolution of the wage premium between 1996 and 2015.

The Supply of Skilled Workers Increased, but So Did Misallocation

To set the stage, table 12 presents census data to track the evolution of the size and type distribution of firms and employment between 1998 and 2013. Notice first that average firm size *fell*. Second, the number of small and very small firms increased more than the number of medium and large firms; in parallel, the number of formal firms *fell* (by almost 8 percent), while that of informal firms *rose* quite substantially, by 64 percent. In parallel, recall from section 2.4 that the share of six-digit sectors where informal firms are a majority increased during this period in manufacturing, commerce and services.

Table 12. Size and Type Distribution of Firms and Employment, 1998–2013

	<i>thousands</i>		
	1998	2013	% change
Total			
Firms	2,693.5	4,099.1	52.2
Employment	11,805.0	17,394.3	47.3
Average firm size	4.38	4.24	0.0
Firms by size			
1–5	2,460.77	3,754.0	52.5
6–10	118.7	186.4	57.0
11–50	91.5	127.9	39.7
51+	22.8	30.6	34.2
Firms by type			
Formal*	443.2	408.0	–7.9
Informal**	2,249.0	3,691.0	64.1
Employment by firm size			
1–5	4,309.8	6,903.4	60.1
6–10	884.9	1,374.2	55.2
11–50	1,879.7	2,663.8	41.7
51+	4,730.5	6,452.7	36.4
Employment by firm type			
Formal*	7,291.2	7,710.2	5.7
Informal**	4,513.8	9,684.0	114.5

Source: Authors' analysis based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: * Sum of legal and formal and mixed. ** Sum of legal and illegal; see table S1.1 in appendix S1.

Employment reflects these changes, and grows more in smaller than in larger ones (that is, in firms that are less intensive in more highly educated workers). These differences are magnified if we focus on firm types: employment in informal firms increases by 114.5 percent vs. 5.7 percent in formal ones. These trends are noteworthy given that, during this period, there were large improvements in the schooling composition of the labor force; so one cannot argue that the absence of a growing supply of educated workers is responsible for them.

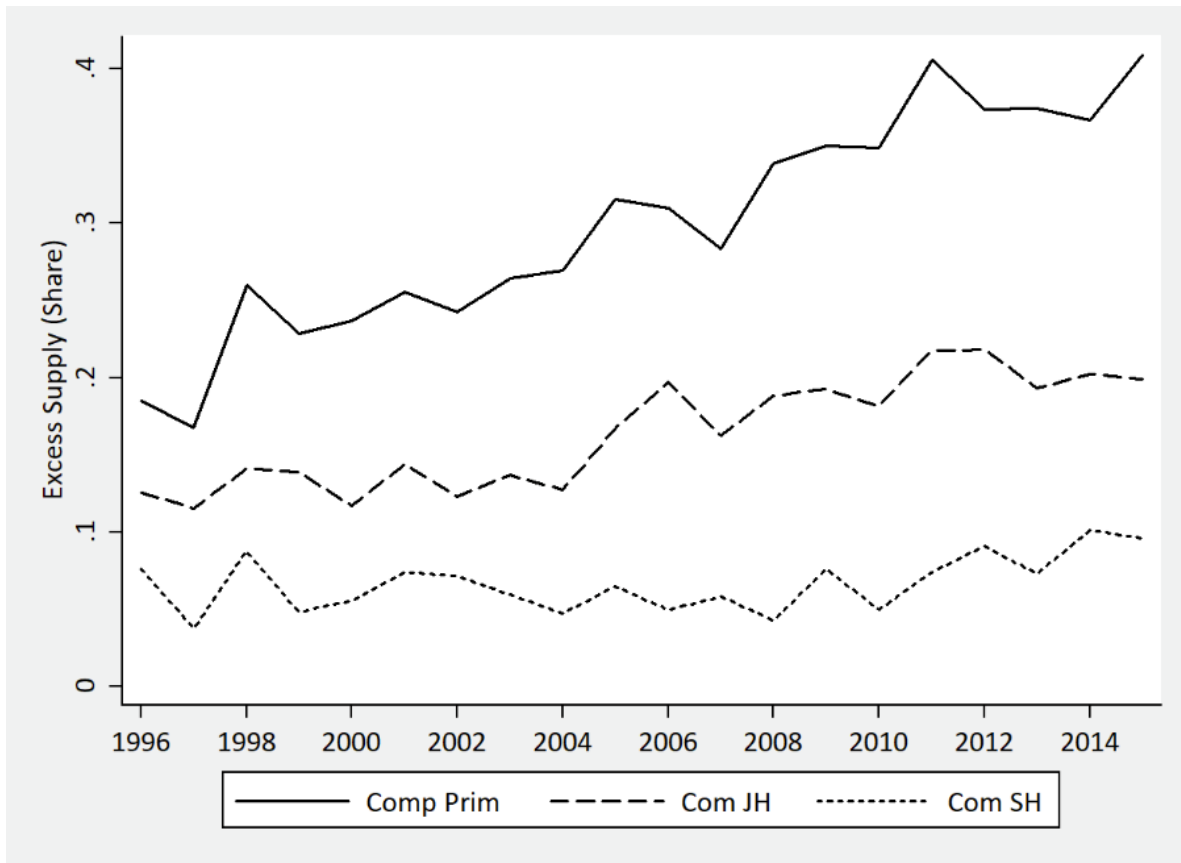
Absent Increasing Misallocation, There Would Be Increasing Excess Demand for Skilled Workers

The trends in the size and type distribution of firms can be contrasted with the trends in the supply of workers by schooling levels as shown in table 5. While the periods do not match exactly, the overlap is substantial. The key point is that, over practically the same period, the educational composition of the working-age population (potential supply) and the economically active population (observed supply) changed importantly as the number of educated persons and workers increased more rapidly than the average of all persons and workers (and, indeed, the number of those with the least education fell in absolute terms). In parallel, however, the degree of the misallocation of resources increased, and always in the direction of firms that are less intensive in educated workers. So, the broad picture that emerges is that of an economy within which, on one hand, the supply of educated workers is increasing relatively quickly but, on the other, given that misallocation is also increasing, firm and employment growth is faster in the informal sector.

To illustrate the effects of these trends, equation (3) is solved for each year between 1996 and 2015. Figure 6 plots the percentage excess supply of workers of educational categories with

complete school cycles (results for incomplete school cycles are intermediate to those with complete cycles, but not shown here to avoid cluttering). As can be seen, if, in every year, the demand for workers of various schooling levels had been the same as that of formal firms, the excess supply of workers with fewer years of education would have increased over time, particularly for those with complete primary education and, to a lesser extent, with complete junior high education. In the first case, excess supply more than doubles in the 20-year span, from 18.5 percent to 41 percent; for the second, it increases from 12.6 percent to 19.8 percent.

Figure 6. Excess Supply of Workers Relative to Those with University Education



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Comp prim* refers to complete primary, *Com JH* refers to complete Junior High, and *Com SH* refers to complete Senior High.

4.3. Counterfactual Paths of the Returns to Education Are above the Observed One

A simple, but, hopefully, suggestive exercise is performed constructing a counterfactual path for the wage premium. The key assumption is that there are no changes in participation rates so that the adjustment to the elimination of firm informality occurs only through changes in relative earnings. Acemoglu and Autor (2011) derive the following expression linking changes in the wage premium to changes in the relative supplies of workers with high (H) and low (L) schooling levels:

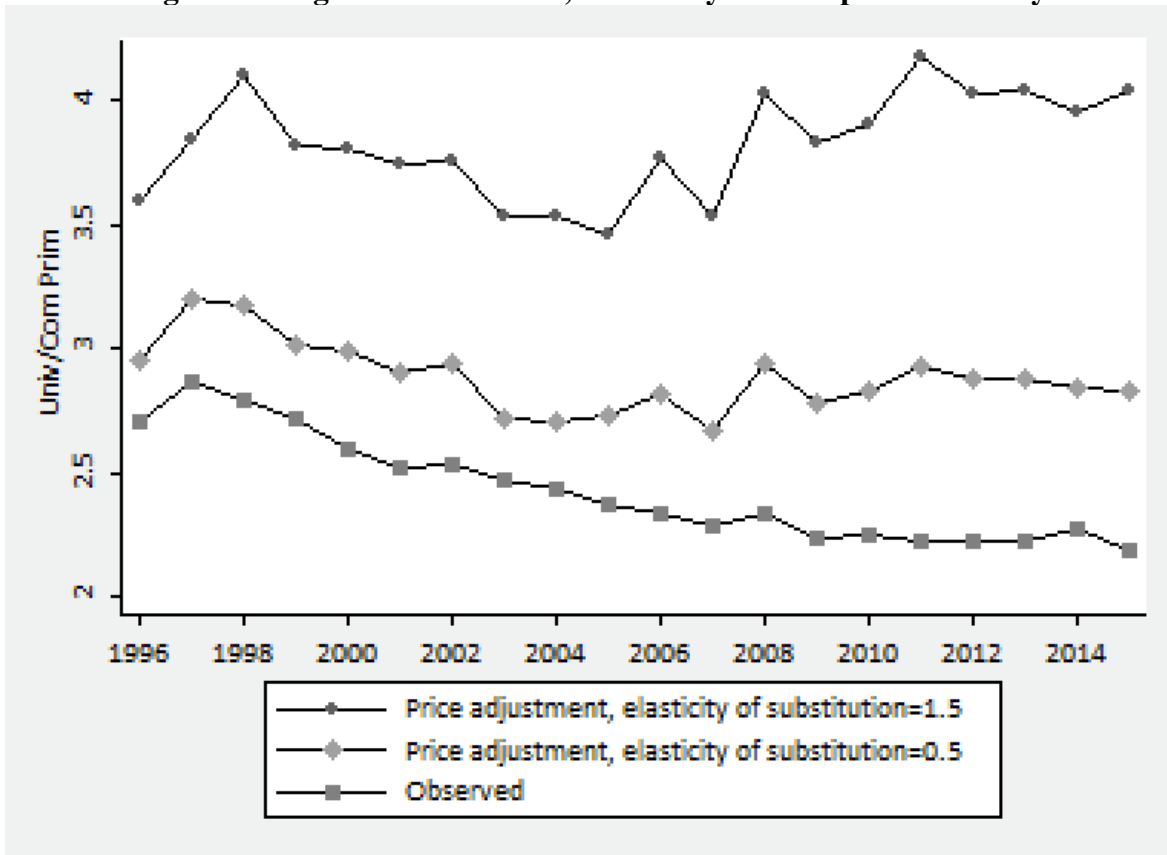
$$\partial \ln(w^H / w^L) = -(1/\sigma) \partial \ln(H / L) \quad (4)$$

where σ is the elasticity of substitution between these two types of labor.

Figure 7 shows three paths of the wage premium between workers with university education and those with complete primary education. The observed path is declining, in line with the earnings trends described in Figure 2. The alternative paths are labeled price adjustment and use equation (4) and the estimates of excess supply of workers with primary education for each year depicted in Figure 6 under two assumed values of the elasticity of substitution, 1.5 and 0.5.¹⁶

¹⁶ Acemoglu and Autor (2011) point out that most estimates of the elasticity of substitution between skilled and unskilled labor in the United States are between 1.4 and 2.0. These estimates correspond to comparisons of workers with university vs. high school education. Benita (2014) provides estimates for Mexico on the order of 3, but again of workers with university vs. primary education. Benita's estimates implicitly reflect Mexico's formal-informal firm composition. In our case, first, we are comparing workers with university vs. complete primary education, and, second, we are considering a scenario where there are only formal firms. For these reasons, the elasticity should be considerably lower. We arbitrarily chose 1.5 and 0.5, which are 50 percent higher and lower than the elasticity of the Cobb-Douglas production function.

Figure 7. Wage Premium Paths, University vs. Completed Primary



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Two results are underlined. First, in the counterfactual scenarios the path of the wage premium is higher than the observed one, even if one assumes a high value for the elasticity of substitution. This implies that the dispersion of earnings across educational levels would be higher than the one we construct, which, as discussed in section 3, does not consider the adjustments needed to bring supply and demand of each schooling level into equilibrium in the absence of firm informality. Second, the trends differ: in the high elasticity scenario, the wage premium is basically constant over time, as opposed to the decreasing trend of the observed premium; in the low value case, the trend is increasing over time.

Our analysis is limited for three reasons. First, only two educational levels are considered, while, if firm informality disappeared, the earnings of workers of all educational levels would be affected. Second, so to speak, the static results of each year are stitched together to compute the path of the wage premium, but there are no intertemporal effects from the equilibrium of one year on the next in the counterfactual scenarios. Third, there is no feedback from changes in earnings to the incentives to invest in education and thus to the supply of workers of different schooling levels. Clearly, if, in any one year, the distortions that give rise to Mexico's large informal sector were removed, the subsequent path of the economy would be different. For these reasons, our results should be considered suggestive. The objective is not to pinpoint precisely the free path of the wage premium in the informal sector, but rather to illustrate that the continuous growth of employment in informal firms due to increasing misallocation, while the schooling of the labor force increases, accounts for part of the explanation for the observed declining path in the premium.

Falling returns to education could be explained by declining school quality. The argument here is that, due to the rapid entry of students into junior high, high school, and university education, the quality of schooling in Mexico has been falling year after year. As a result, the earnings of those with more schooling fall not because of misallocation, as argued above, but because the knowledge and abilities of new workers entering the labor force with more years of schooling are not improving as much as their expanded education would suggest. There is insufficient data on the quality of education in Mexico to test this hypothesis formally, but the few data that there are, described in appendix S3, points in the opposite direction: by both international and national tests, and other available measures, the quality of schooling in Mexico has increased over the period considered here.

A separate concern is whether the counterfactual exercise overestimates the effect of eliminating misallocation on earnings, insofar as it could imply substantial changes in the size distribution of firms. Although it is difficult to identify the appropriate benchmark for comparison, appendix S4 shows that the counterfactual size distribution—which is the observed size distribution of formal firms in Mexico—implies that average firm size in Mexico is substantially smaller than that observed in the United States. Moreover, as shown by Leal (2014), if distortions in Mexico were fully eliminated, the size distribution of formal firms would also change in the direction of larger firms, so that the change in labor earnings and the returns to education would be larger than those estimated here.

5. Conclusions

This paper presents a preliminary assessment of the effects of misallocation on labor earnings and the returns to education in Mexico. Misallocation is viewed here as the outcome of frictions, institutional arrangements, and market and regulatory failures in input and output markets (broadly, distortions) that misallocate too many resources toward firms with non-salaried contracts or with illegal salaried contracts (that is, informal firms). We show that, because the production process in these firms is substantially less intensive in educated workers than in formal ones, misallocation tilts the demand for labor in favor of workers with fewer years of schooling. In doing so, misallocation reduces the mean and the variance of the distribution of labor earnings and lowers the returns to education. We show this by computing the earnings distribution that would be observed in a counterfactual scenario, whereby the schooling composition of the demand for labor from all firms was the same as that of existing formal firms. The basic finding is that misallocation

lowers the average earnings of informally employed workers by 17 percent, but with substantial differences across educational levels. For workers with incomplete primary education, earnings fall by 9 percent; for those with university education, by 29 percent. In other words, misallocation is costly to all, but costlier to workers with more years of schooling.

The analysis is limited in scope because only the direct effects of firm informality on employee earnings are considered, although it is shown that, if general equilibrium effects were considered, the results would be strengthened. The analysis falls short of capturing other channels by which misallocation affects earnings. Clearly, if the distortions that support Mexico's informal sector were removed, there would be changes in the occupational distribution of individuals between those who are employees, those who are self-employed, and those who are entrepreneurs. Furthermore, firm-worker relations would change, with longer job tenures and greater investments in worker training.¹⁷ Firm dynamics would also differ.¹⁸ None of these effects are captured in our paper, though they would quite likely have first-order effects on earnings and the returns to education.

Despite these limitations, we claim to have established that misallocation is a central part of the explanation of why the earnings of more educated workers have fallen in absolute terms and why the wage premium and the returns to education have also fallen. At the end of the day, the story of our paper is simple. Workers are entering the labor market with more years of schooling, but fail

¹⁷ Alaimo et al. (2015) argue that excessive worker rotation is associated with lower investments by firms in worker training and lower returns to experience. Furthermore, they find that, in Mexico, 63 percent of workers will never receive any on-the-job training in their work life and that those who do work have higher education and work mainly in formal firms.

¹⁸ Hsieh and Klenow (2014) compare firm dynamics in Mexico and the United States, and find that, given firm size at birth, over a 40-year span, the average firm in Mexico increases in size by a factor of 2 vs. 7 in the United States. They estimate that these differences lower productivity in Mexican manufacturing relative to the United States by about 25 percent.

to find jobs that fully value their additional education. This is so not because there are not enough firms ready to hire them, but because too many of those firms are informal and do not need more educated workers. This story has a corollary: The distortions that misallocate resources today are costlier now than 20 years ago because there are more workers with higher education now than before (and, if these distortions persist while the schooling of Mexican workers continues to improve, they will be even costlier).

We conclude with four observations. First, the analysis suggests that there is no automatic connection between increases in worker schooling and labor earnings. Rather, the translation from increased schooling to higher earnings is mediated by the degree of misallocation in the economy. If misallocation is large and persistent, earnings can stagnate even as human capital is being accumulated.

Second, our approach may help explain the fall in the returns to education observed in other countries of Latin America beyond Mexico (see Gasparini et al. 2011; Rodríguez-Castelán et al. 2016). In these countries, there has also been a steady increase in the supply of educated workers, in some cases, accompanied by mild declines in labor informality. This paper suggests that it is critical to look beyond worker and firm classifications—which may change without affecting behavior and which are not always comparable across countries. It is essential to concentrate on misallocation. Are there large distortions? Are these distortions de facto taxing firms with certain characteristics and subsidizing others? Are there systematic differences in the schooling composition of taxed and subsidized firms? While not all countries may have sufficient firm-level data to carry out computations such as the ones reported here, those presented here rely on employment surveys and may still shed important light.

The third observation concerns the debate on income inequality in Mexico. López-Calva and Lustig (2010) and Cord et al. (2014) show that a narrowing wage premium was one of the main causes behind the fall in the Gini coefficient, from 0.55 to 0.50, observed between 2002 and 2010. Reduced inequality is clearly welcome in a country as unequal as Mexico, but our analysis suggests that, to the extent that the narrowing wage premium results from persistent misallocation, lower inequality comes partly at the expense of lower mean earnings and lower returns to education. By shifting resources in favor of informal firms that are more intensive in less-educated workers, misallocation in Mexico reduces earnings dispersion. But since misallocation is also reducing productivity, it is lowering worker living standards, particularly among more educated workers. It is far from clear that this source of reduced inequality can be seen in a positive light.

Finally, this analysis highlights the importance of understanding the nature of the distortions that stand behind Mexico's large informal sector. And while research disentangles the role played by various factors, this paper indicates that lack of educated workers is not one of them. The widespread notion that investments in education will gradually eliminate misallocation (or informality) is, in our view, flawed. Rather, this paper suggests the opposite: the persistence of misallocation is impeding Mexico from taking full advantage of its investments in education.

References

- Acemoglu, D., and D. Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings”, in *Handbook of Labor Economics*, Vol. 4, Part B, edited by O. Ashenfelter and D. Card. San Diego: Elsevier.
- Alaimo, V., M. Bosch, D. Kaplan, C. Pages, and L. Ripani. 2015. *Jobs for Growth*. Washington DC: Inter-American Development Bank.
- Andridge, R.R., and R. J. A. Little. 2010. “A Review of Hot Deck Imputation for Survey Non-response”. *International Statistical Review* 78 (1): 40-64.
- Autor, D., L. Katz, and L. Kearney. 2008. “Trends in US Inequality: Revising the Revisionists”, *The Review of Economics and Statistics*, 90(2): 300-323.
- Benita, F. 2014. “A Cohort Analysis of the College Wage Premium in Mexico.” *Latin American Journal of Economics* 51(1): 147-78.
- Bobba, M., L. Flabbi, and S. Levy. 2017. “Labor Market Search, Informality and Schooling Investments.” IZA Discussion Paper No. 11170, Institute of Labor Economics, Bonn, Germany.
- Busso, M., M. V. Fazio, and S. Levy. 2012. “(In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico.” Discussion Paper No. 341, Inter-American

Development Bank, Washington, D.C.

Busso, M., L. Madrigal, and C. Pagés. 2013. "Productivity and Resource Misallocation in Latin America." *B.E. Journal of Macroeconomics* 13(1):1-30.

Campos-Vázquez, R. 2013. "Efectos de los Ingresos no Reportados en el Nivel y Tendencia de la Pobreza Laboral en México." Documento de Discusión No. IV-2013, Centro de Estudios Económicos, El Colegio de México, Mexico City.

Campos-Vázquez, R., G. Esquivel, and N. Lustig. 2012. "The Rise and Fall of Income Inequality in Mexico." WIDER Working Paper No. 2012-10, United Nations University, Helsinki.

Campos-Vazquez, R., L.F. López-Calva, and N. Lustig. 2016. "Declining Wages for College-Educated Workers in Mexico: Are Younger or Older Cohorts Hurt the Most?" Working Paper No. 1522, Department of Economics, Tulane University, New Orleans, LA.

Cord, L., L. Barriga, C. Lucchetti, L. Rodríguez-Castelán, D. De Souza, and F. Valderrama. 2014. "Inequality Stagnation in Latin America in the Aftermath of the Global Financial Crisis." Policy Research Working Paper WPS7146, World Bank, Washington DC.

Gasparini, L., S. Galiani, G. Cruces, and P. Acosta. 2011. "Educational Up-Grading and the Returns to Skills in Latin America: Evidence from a Supply-Demand Framework, 1990-2010." Documento de Trabajo No. 127, CEDLAS, Universidad Nacional de la Plata, La Plata,

Argentina.

Goldin, C. and L. Katz. 2007. "The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005." NBER Working Paper No. 12984, National Bureau of Economic Research, Cambridge, MA.

Hsieh, C., and P. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403-1448.

_____. 2014. "The Life-Cycle of Plants in India and Mexico." *Quarterly Journal of Economics* 129 (3): 1035-84.

Inter-American Development Bank (IDB). 2010. *The Age of Productivity: Transforming Economies from the Bottom Up*. Washington, DC: Inter-American Development Bank.

LaPorta, R., and A. Schleifer. 2008. "The Unofficial Economy and Economic Development," *Brookings Papers on Economic Activity* 39 (2): 275-363.

Leal, J. 2014. "Tax Collection, the Informal Sector and Productivity." *Review of Economic Dynamics* 17 (2): 262-86.

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

_____. 2018. *Under-Rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. Washington, DC: Inter-American Development Bank.

Levy, S., and L.F. López-Calva. 2016. “Labor Earnings, Misallocation and the Returns to Schooling in Mexico.” Working Paper Series IDB-WP-671, Inter-American Development Bank, Washington, DC.

Levy, S., and M. Székely. 2016. “Más Escolaridad, Menos Informalidad? Un Análisis de Cohortes para México y América Latina.” *El Trimestre Económico* 83 (332): 449-548.

López-Calva, L.F., and N. Lustig. 2010. *Declining Inequality in Latin America: A Decade of Progress?* Washington, DC: Brookings Institution Press.

López-Martin, B. 2015. “*Informal Sector Misallocation*.” Working Paper No. 2016-09. Research Department, Banco de Mexico, Mexico City.

Maloney, W. 2004. “Informality Revisited”, *World Development* 32(7): 1159-1178.

McCaig, B., and N. Pavcnik. 2015. “Informal Employment in a Growing and Globalizing Low-Income Country.” *American Economic Review Papers and Proceedings* 105 (5): 545-50.

Nataraj, S. 2011. “The Impact of Trade Liberalization on Productivity: Evidence from India’s

Formal and Informal Manufacturing Sectors.” *Journal of International Economics* 85(2): 292-301.

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

Robertson, R. 2007. “Relative prices and wage inequality: Evidence from Mexico.” *Journal of International Economics* 64(2): 387–409.

Rodríguez-Castelán, C., L.F. López-Calva, N. Lustig, and D. Valderrama. 2016. "Understanding the Dynamics of Labor Income Inequality in Latin America." Policy Research Working Paper Series 7795, The World Bank, Washington, DC.

Syverson, C. 2004 . “Product Substitutability and Productivity Dispersion.” *Review of Economics and Statistics* 86(2):534-50.

Ulyssea, G. 2018. “Firms, Informality, and Development: Theory and Evidence from Brazil.” *The American Economic Review* 108 (8): 2015-47.

**Appendix to Persistent Misallocation and the Return to Education in Mexico by Santiago
Levy and Luis Felipe López-Calva**

Appendix S1: Definitions and Data

Definitions

We distinguish between self-employed workers and workers engaged with firms. In turn, the latter—employees—may be employed under salaried or non-salaried contractual relations, a key distinction in Mexico’s institutional context. A salaried worker is a *subordinated* employee. The hiring firm is obligated by Law to pay him or her at least the minimum wage, observe regulations regarding dismissals, and contribute to their social insurance benefits. Non-salaried workers, on the other hand, may be associated with a firm but are not subordinated to it. Examples include door-to-door salespeople, workers on temporary contracts performing a nonrecurring task, and, very importantly in the case of Mexico, workers who are relatives and collaborate in a family firm. Critically, the law does not obligate firms to contribute to the social insurance benefits of nonsalaried workers, nor to observe labor regulations regarding dismissals or minimum wages.

The salaried and non-salaried distinction is central to our definition of formal and informal workers. Formal workers are salaried employees covered by labor regulations regarding minimum wages and dismissal, among other matters, and who benefit from social insurance paid by the firm that hires them. All other workers, including the self-employed, are informal. If the law was fully enforced, all salaried workers would be formal. However, this is not the case in Mexico (see Busso, Fazio, and Levy 2012). This implies that informal workers consist of the self-employed, nonsalaried employees and salaried employees in firms that do not comply with the law.

The formal-informal distinction is not as sharp in the case of firms since firms often mix salaried and nonsalaried employees, and at times violate the Law. Table S1.1 identifies four possible combinations (all observed in the data; see table 1 in the main text).

Table S1.1. Formality status of firms and workers

	<i>Contracts between firms and workers</i>				<i>Self-employed workers</i>
	<i>Only salaried, compliant with law</i>	<i>Salaried and nonsalaried, fully or partly compliant</i>	<i>Only nonsalaried, not obligated by law</i>	<i>Only salaried, not compliant with law</i>	
Firm	Formal and legal	Mixed	Informal and legal	Informal and illegal	Not applicable

Worker	Formal	Salaried compliant, formal; rest, informal	Informal	Informal	Informal
--------	--------	--	----------	----------	----------

Source: Authors' own work.

Data

We use Mexico's employment surveys for the period 1996–2015, after the start of NAFTA in 1994 and the financial crisis of 1995. From 1996 to 2004, the survey was known as the Encuesta Nacional de Empleo (National Employment Survey), ENE; after 2005, as the Encuesta Nacional de Ocupación y Empleo (National Survey on Occupation and Employment), ENOE. We refer to it here as ENE-ENOE, a nationally representative quarterly survey on types of employment (public or private employees, or self-employed); labor status (formal, informal or unemployed); location (municipality); size of firm where workers are employed; workers' age, gender, years of schooling; and other job dimensions of a job such as written contracts and yearly bonus payments. ENE-ENOE also records hours worked and earnings net of taxes and contributions. We use data from the second quarter of each year and measure earnings per hour in prices of May 2008. Importantly, the employment surveys have a panel structure that makes it possible to follow the same worker through five consecutive quarters. Thus, we can measure individual worker transitions in the course of a year across firm size and formal-informal status.

The relevant educational categories for Mexico are primary (six years, usually ages 6–12); junior high (three years, usually ages 13–15); senior high (three years, usually ages 16–18); and university (four years or more). With these categories, we classify workers into seven groups: incomplete primary, complete primary, incomplete junior high, complete junior high, incomplete high school, complete high school, and university.

An issue with ENE-ENOE is that some workers fail to report earnings; see Campos-Vazquez (2013). To correct for this problem, we match workers with and without earnings on observable characteristics like gender, years of education, age, location, size of firm, and formality status. We take a random sample of workers with reported earning in each of these categories, and randomly impute their earnings to workers with the same observable characteristics but without reported earnings.¹⁹ See appendix S2 for more details.

We also use the Economic Census, a firm-based data set published every five years. We have four available, for 1998, 2003, 2008, and 2013, covering the 1996–2015 period for which we have employment data. The census collects information on fixed establishments of all sizes (henceforth,

¹⁹ To verify the robustness of our procedure, we compute earnings with data from Mexico's household survey which, as documented in appendix S2, does not have the same underreporting problem that ENE-ENOE has. While the definitions of earnings do not match perfectly between the two surveys, the trends and structure of earnings, including the decline in the returns to schooling, are confirmed.

firms) in urban areas. Rural areas and economic activity in urban areas carried out in the streets are not captured (street vendors, street markets, and so on); the same holds true for public sector employment. As a result, the census only captures about 45 percent of the occupied population portrayed in ENE-ENOE. We restrict the sample from ENE-ENOE to private sector employees in firms in urban areas, ensuring that they are a subset of the employees captured in the census.

The census contains data on the total number of workers and the aggregate of earnings and social security payments at the firm level, allowing us to classify firms according to the typology described in table S1.1; see Busso, Fazio, and Levy (2012). We also group firms by size into four categories: 0 to 5 workers (henceforth, very small firms), 6 to 10 (small firms), 11 to 50 (medium firms), and 51 or more (large firms).

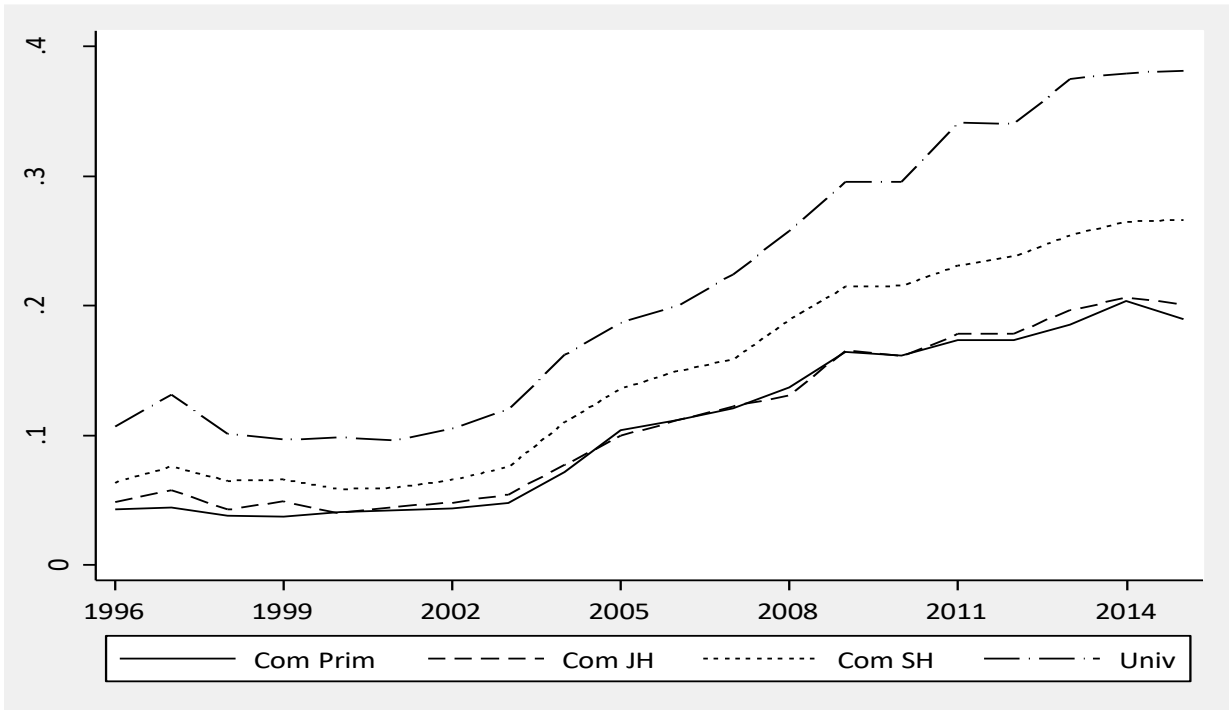
Unfortunately, the census does not include data on individual workers within each firm. Nevertheless, ENE-ENOE identifies the size of the firm where a worker is employed; therefore, we can classify workers by years of schooling, firm size, and formal-informal status. Our data, however, does not allow mapping formal and informal *workers*, as identified in ENE-ENOE, into the formality status of firms as identified in the census. Since we focus on the earnings of individual workers by years of schooling, firm size, and formality status, we mostly rely on ENE-ENOE data. Nonetheless, we use the census to measure misallocation and the size and type distribution of firms.

Appendix S2: Addressing the Problem of Missing Observations

Statistical Approach

A frequent problem with employment surveys is that earnings are not reported by all workers surveyed. This could bias the results of our paper to the extent that those not reporting earnings are not a random sample of the surveyed population. Table S2.1 highlights two facts in the case of ENE-ENOE: first, underreporting rates increase with the level of education: second, for all educational levels underreporting has increased. (See also figure S2.1.)

Figure S2.1. Rates of Under-Reporting of Earnings, Encuesta Nacional de Empleo and Encuesta Nacional de Ocupación y Empleo



Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Empleo and Encuesta Nacional de Ocupación y Empleo.

Note: *Com* refers to complete. *Prim*, *JH*, *SH*, and *Univ* refer to primary, junior high, senior high, and university, respectively. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

To correct for this problem, we follow Campos-Vazquez (2013) who, after discussing several alternative methods, recommends applying the so-called Hot Deck technique to impute earnings to those that fail to report them. This technique is applied to our sample of employees in three steps. First, we define n groups of workers formed by all the cross-combinations of variables that are observed both for individuals that do and do not report earnings. These variables are gender, the seven education levels, the quartile of age in which the individuals are (computed over the economically active population (EAP)), the formality status, and the four categories of firm size. As a result, we end up with 448 groups (7 educational levels*2 gender*4 age categories*2 formality status*4 firm size). Second, within each group we randomly choose a number of individuals who do report equal to the number who do not, but who belong to the same group. Lastly, we randomly impute the earnings of those who do report, to those who do not. Note that this technique does not require assuming any functional form for earnings functions, but it does require assuming that there are no systematic differences in earnings levels, within each group, between those that do and do not report earnings.²⁰

The results of the imputation are displayed in table S2.1 for 2006. We recover a total of 1,069,039 observations corresponding to employees who had failed to report earnings. As can be seen, average earnings for all groups combined increase marginally, from 31.79 to 32.33 pesos, with some heterogeneity across schooling levels, while the standard deviations remain more or less the same.

Table S2.1. Employee Earnings, 2006

	Pre Hot Deck			
Education Level	N	Mean	Sd	Median
Incomplete Primary	391,373	21.02	11.38	19.03
Complete Primary	983,510	21.95	12.75	19.32
Incomplete Junior High	270,262	22.56	14.38	19.09
Complete Junior High	1,796,225	23.30	13.38	20.45
Incomplete Senior High	1,081,256	28.75	19.15	23.64
Complete Senior High	949,475	30.20	19.89	25.14
University	1,524,520	55.70	44.58	43.63
Total	6,996,621	31.79	28.21	23.18
	Post Hot Deck			
Incomplete Primary	422,174	20.87	11.18	18.75
Complete Primary	1,094,546	21.81	12.64	19.09
Incomplete Junior High	295,160	22.47	14.34	19.09
Complete Junior High	1,979,862	23.24	13.24	20.45
Incomplete Senior High	1,244,486	28.61	19.22	22.95
Complete Senior High	1,113,035	30.57	20.19	25.37

²⁰ Further details can be found in Andridge and Little (2010).

University	1,916,397	55.20	44.16	43.60
Total	8,065,660	32.33	28.77	23.55

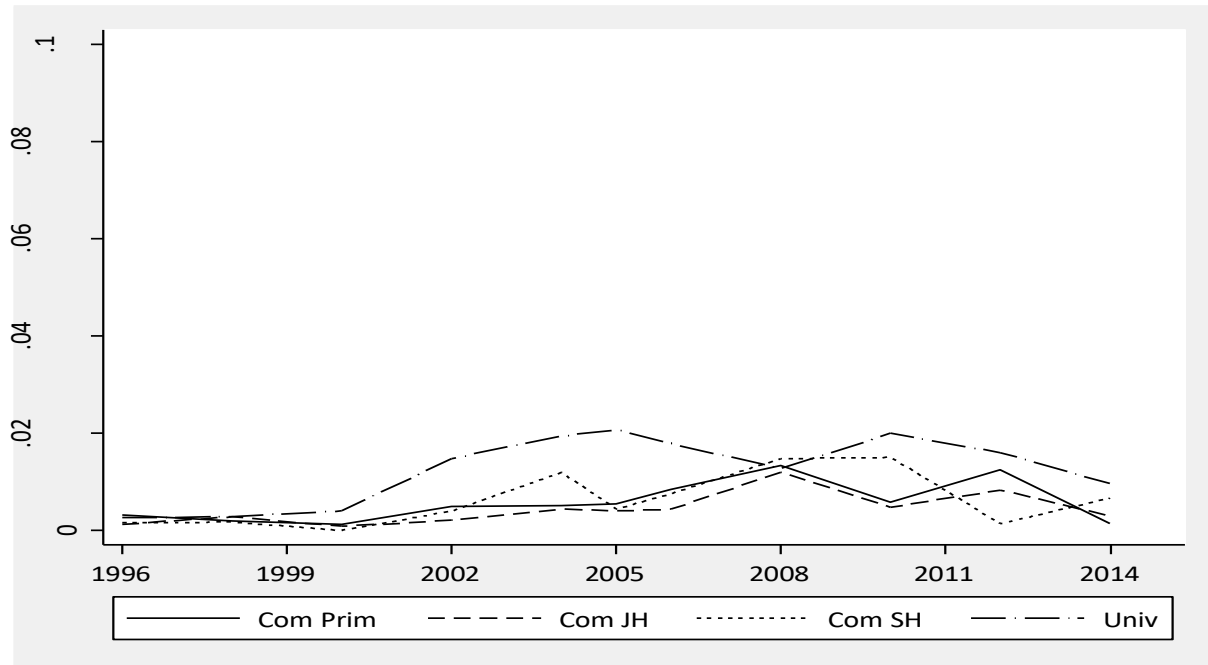
Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Ocupación y Empleo and Encuesta Nacional de Empleo.

Note: *Incomplete Primary* refers to persons who attended primary school but did not complete it. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Incomplete Junior High* refers to persons who attended Junior High but did not complete it. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Incomplete Senior High* refers to persons who attended Senior High but did not complete it. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Corroboration with Data from Mexico's 1994–2012 Household Surveys

The Hot Deck technique implicitly assumes that, within each group, individuals who do not report earnings are a random subsample of all individuals in that group. Since, by definition, we have no information on the earnings of those that do not report, we cannot test for this assumption. Nevertheless, to corroborate our results we also measure the rates of underreporting of workers' earnings using data from Mexico's national household income and expenditure survey (Encuesta Nacional de Ingresos y Gastos de los Hogares, ENIGH), which is collected every two years and is available for the period 1994–2012. As in the paper, we exclude the self-employed and government employees and focus on private sector employees between 18 and 65 years of age, living in towns with over 100,000 inhabitants, who work more than 35 hours a week. Figure S2.2 shows the rates of underreporting in ENIGH.

Figure S2.2. Rates of Under-Reporting of Earnings, ENIGH



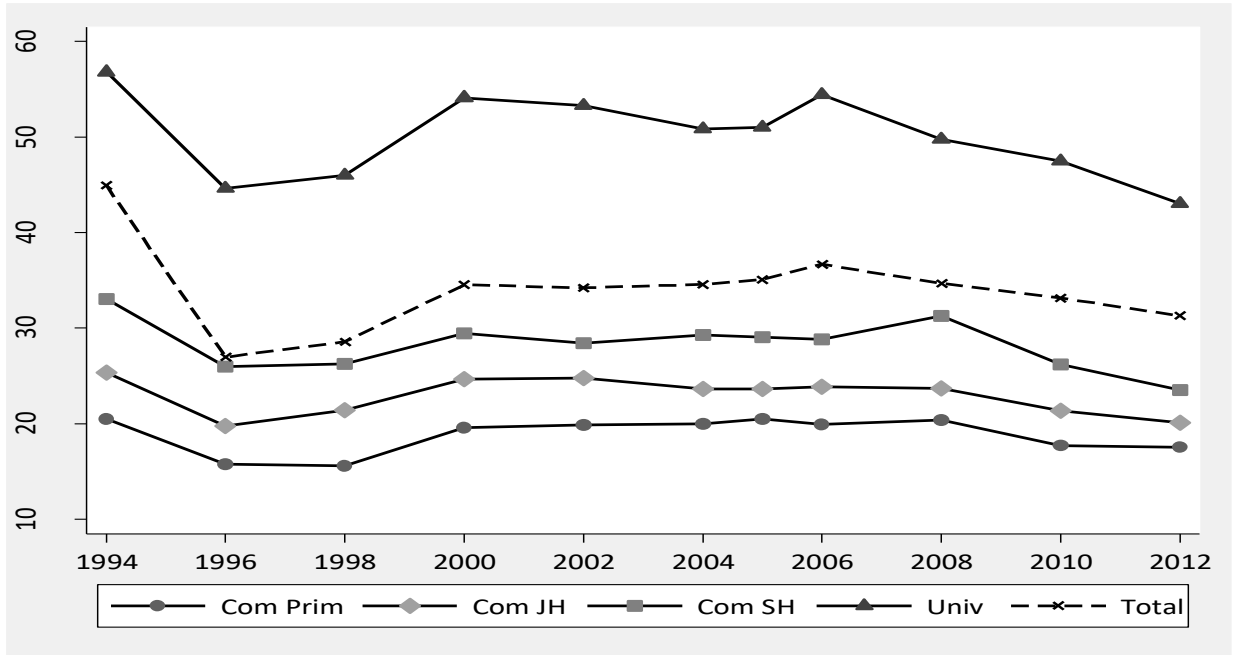
Source: Authors' computations based on Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) data.

Note: *Com* refers to complete. *Prim*, *JH*, *SH*, and *Univ* refer to primary, junior high, senior high, and university, respectively. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Contrasting figures S2.1 and S2.2, it is clear that (1) rates of underreporting are substantially lower in ENIGH compared to ENE-ENOE, (2) differences in underreporting across educational levels are very small, and (3) over time rates of underreporting increased slightly, but then declined.

Figure S2.3 reproduces the same results presented in figure S2.1, which uses ENE-ENOE data. The critical result for our purposes is that the trends in earnings after 1996 in the ENIGH data are similar to those in the ENE-ENOE data after applying the Hot Deck technique, thus providing indirect support for the assumption that workers not reporting earnings in each of the groups formed with the ENE-ENOE data are a random sample of all workers in that group. In addition, one can note in figure S2.3 the sharp drop in earnings immediately following the 1994–95 financial crisis, which could not be seen with the ENE-ENOE data, as well as the fact that, by and large, the upward trend in earnings from 1996 to the early 2000s is basically a recovery from a sharp fall.

Figure S2.3. Earnings by Education Level, ENIGH real hourly wages, November 2014 pesos

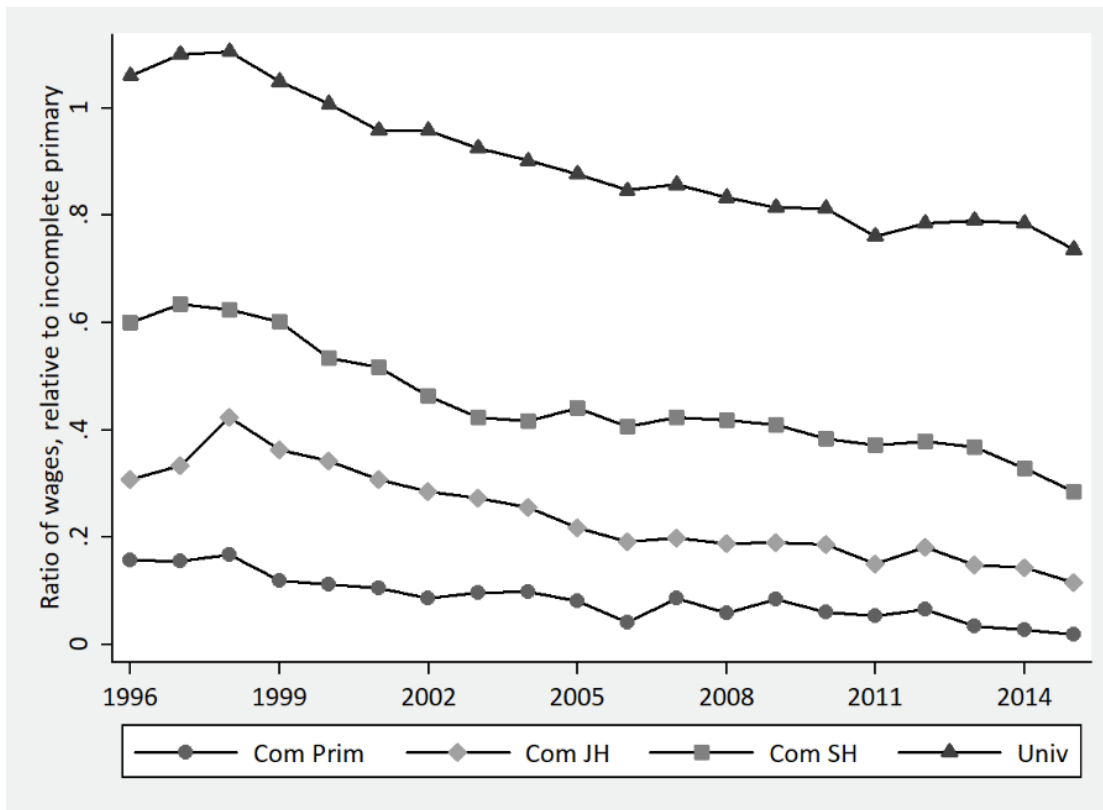


Source: Elaboration based on Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) data. Data exclude self-employed workers.

Note: *Com* refers to complete. *Prim*, *JH*, *SH*, and *Univ* refer to primary, junior high, senior high, and university, respectively. *Complete Primary* refers to persons who attended and completed primary school but did not move on to Junior High. *Complete Junior High* refers to persons who attended and completed Junior High education but did not move on to Senior High. *Complete Senior High* refers to persons who attended and completed Senior High but did not move on to University. *University* refers to persons who attended and obtained a University Degree.

Finally, figure S2.4 presents estimates of the returns to education using ENIGH data, focusing on the same sample of employees described in the text. Again, the results are similar to those presented in figure 3 in the main text using ENE-ENOE data after applying the Hot Deck technique.

Figure S2.4. Evolution of Returns to Schooling by Education Level, ENIGH



Source: Authors' computations based on Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) data.

Note: Com refers to complete. Prim, JH, SH, and Univ refer to primary, junior high, senior high, and university, respectively. Complete Primary refers to persons who attended and completed primary school but did not move on to Junior High. Complete Junior High refers to persons who attended and completed Junior High education but did not move on to Senior High. Complete Senior High refers to persons who attended and completed Senior High but did not move on to University. University refers to persons who attended and obtained a University Degree.

Appendix S3: On the Quality of Education

One potential argument to explain the decline in the returns to education in Mexico is that the quality of schooling has been falling. The facts, however, do not favor that explanation. Even though quality of education is relatively low in Mexico, showing the lowest performance among OECD countries in terms of the Programme for International Student Assessments (PISA) results, what matters for our analysis are the trends, rather than the levels. Looking at those trends, we conclude that the massive increase in coverage and schooling achievement in the last decades has been accompanied certainly by a nondecreasing trend in quality and arguably by a mostly increasing trend.

Tables S3.1 and S3.2 show the results for two types of standardized tests in basic and secondary education for which data are available. First, PISA, which is the internationally comparable test carried out by the OECD, which shows an increase in the average score for Math (from 387 to 413) and Language (from 422 to 424) between 2000 and 2006. A mediocre performance indeed, but nondecreasing over time.

Table S3.1. Mexico, PISA scores

	2000	2003	2006	2009	2012
Math	387	385	406	419	413
Language	422	400	410	425	424

Source: Authors' elaboration based on PISA data.

Table S3.2. Mexico, Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) scores

	Language		Mathematics	
	Primary	Junior High	Primary	Junior High
2006	500.0	500.0	500.0	500.0
2007	513.7	507.8	511.3	509.3
2008	514.1	513.8	519.0	512.8
2009	504.5	520.4	506.0	522.6
2010	488.6	532.2	510.7	529.5

2011	485.6	542.6	513.0	544.1
2012	495.7	550.9	532.2	571.6
2013	494.5	550.7	536.3	583.5

Source: Authors' elaboration based on Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) data.

Additionally, there is national standardized test applied in all schools between 2006 and 2013, the so called ENLACE test. The only slightly decreasing score is for language in primary schooling, which goes from 500 as average score to 494.5, a reduction that would be unlikely to explain an important change in returns. The math core is increasing for primary schooling, from 500 to 536.3 in the same period. For secondary education (junior high) the trends are increasing in both language and math.

There are no comparable results for tertiary education. However, different indicators also point in the direction of increasing quality. First, using the data from ENE-ENOE, the share of workers with university education that completed at least a bachelor's degree (four or five years of education after high school) increased from 66.7 in 1996 to 70.5 in 2015. Second, according to UNESCO data, the gross enrollment ratio of university-age students rose significantly from 15 percent in 1991 to approximately [31.2 percent in 2016](#). Third, even though about 70 percent of tertiary education students are in the public education system, private provision of tertiary education has grown rapidly in the last 15 years. Out of the 3.6 million students in tertiary education in Mexico in 2015, close to one million are in private universities, whose quality is highly heterogeneous (UNESCO 2016). This rapid increase in private supply of tertiary education services has led to the hypothesis that there is a "degraded tertiary" effect, that could explain the reduction in wages and average returns to tertiary education in Mexico and other countries in Latin America facing the same pattern. However, Campos-Vazquez, López-Calva, and Lustig (2016) carry out a cohort analysis and do not find evidence that supports the degraded tertiary hypothesis. Fourth, the accreditation system for tertiary education programs introduced in Mexico in 2000 (COPAES, for its Spanish acronym), also suggests that quality has not declined for the system as a whole. COPAES is a nonprofit civic organization charged by the Ministry of Education to recognize official accrediting bodies in different fields of study. Those accrediting bodies then accredit undergraduate degree programs (licenciado, técnico superior/profesional asociado), designating them to be of "good quality" (buena calidad) if successful. By December 2015, more than 85 percent of tertiary education institutions, public and private, had been accredited, covering about 65 percent of the total number of students (including graduate programs). COPAES reports quarterly on their accreditation process and the information, by location and institution, is available to all prospective students and potential employers.

Appendix S4: Observed vs. Simulated Firm Size Distribution

Table S4.1 compares the observed size distribution of firms in Mexico with the one simulated in the paper and with that of the United States. As can be seen, the differences between the observed distributions between Mexico and the United States are quite substantial, with a significantly larger average firm size in the United States. But note that the simulated distribution for Mexico, while as expected has a larger firm size than the observed one, still implies a smaller firm size than in the United States. Thus, in the simulated scenario 55.7 percent of workers would be in firms with 51 or more workers, compared to 70.1 percent in the United States. At the other end, in the simulated scenario, almost 13 percent of workers would be in firms with up to 10 workers, vs. 11 percent in the United States. This reflects that even the formal size distribution of firms in Mexico is skewed towards smaller firms compared to the United States.

Table S4.1. Distribution of Employment by Firm Size (*shares*)

	<i>Mexico Observed</i>	<i>Mexico Simulated</i>	<i>United States</i>
[1 – 5]	37.8	5.2	5.1
[6 – 10]	8.8	7.7	5.9
[11 – 50]	14.9	31.4	18.1
51+	38.5	55.7	70.1

Source: Authors' computations based on Economic Census and data from Encuesta Nacional de Empleo and Encuesta Nacional de Ocupación y Empleo.

REFERENCES

- Andridge, R. R., and R. J. A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review* 78 (1): 40–64.
- Busso, M., M. V. Fazio, and S. Levy. 2012. "(In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico." Discussion Paper No. 341, Inter-American Development Bank, Washington, DC.
- Campos-Vázquez, R. 2013. "Efectos de los Ingresos no Reportados en el Nivel y Tendencia de la Pobreza Laboral en México." Documento de Discussion No. IV-2013, Centro de Estudios Económicos, El Colegio de México, Mexico City.

Campos-Vázquez, R., L. F. López-Calva, and N. Lustig. 2016. “Declining Wages for College-Educated Workers in Mexico: Are Younger or Older Cohorts Hurt the Most?” Working Paper No. 1522, Department of Economics , Tulane University, New Orleans, LA.

United Nations Educational, Scientific, and Cultural Organization. (2016). Data for the Sustainable Development Goals [Education and Literacy]. Retrieved from <http://uis.unesco.org/en/country/mx?theme=education-and-literacy>