IDB WORKING PAPER SERIES Nº IDB-WP-671

Labor Earnings, Misallocation, and the Returns to Education in Mexico

Santiago Levy
Luis Felipe López-Calva

Inter-American Development Bank
Department of Research and Chief Economist

February 2016

Labor Earnings, Misallocation, and the Returns to Education in Mexico

Santiago Levy*
Luis Felipe López-Calva**

* Inter-American Development Bank
** World Bank

February 2016

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Levy, Santiago.
Labor earnings, misallocation, and the returns to education in Mexico / Santiago Levy, Luis Felipe López-Calva.
p. cm. — (IDB Working Paper Series ; 671)
Includes bibliographic references.
1. Wages-Mexico. 2. Resource allocation-Mexico. 3. Education-Economic aspects-Mexico. I. López-Calva, Luis Felipe. II. Inter-American Development Bank. Department of Research and Chief Economist. III. Title. IV. Series.
IDB-WP-671

http://www.iadb.org

Abstract[*]

Over the last two decades Mexico has had an open trade regime, experienced macroeconomic stability, and made substantial progress in education. However, average workers' earnings have stagnated and earnings for workers with more schooling have declined, compressing the earnings distribution and lowering the returns to education. We hypothesize that these developments are explained by large and persistent of distortions that misallocate resources towards less productive firms, since these firms are substantially less intensive in educated workers than more productive ones. We show that at the same time that 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 and demand for educated workers. We decompose worker earnings into observable and unobservable firm and individual worker characteristics, and simulate a counterfactual earnings distribution in the absence of misallocation. Under the counterfactual, earnings differentials across schooling levels would increase, as would the returns to education. In parallel, earnings differentials, rather than narrowing overtime, would widen. We conclude arguing that the persistence of distortions that misallocate resources toward lower-productivity firms impedes Mexico from taking full advantage of its investments in the human capital of its workers.

JEL classifications: J24, J23, O17, L11
Keywords: Earnings, Misallocation, Returns to education, Human capital

[*] Santiago Levy is with the Inter-American Development Bank and Luis Felipe López-Calva with the World Bank. Their opinions do not necessarily coincide with those of the institutions they are affiliated with. The authors are very grateful to Matias Morales for excellent research assistance and to Luca Flabbi, Rafael de Hoyos, Samuel Freije, Julian Messina, Hugo Nopo and Norbert Schady for very useful comments on an earlier draft. They also thank, without implicating, Rodrigo Negrete and Tomas Ramirez from Mexico's Instituto Nacional de Estadística, Geografía e Informática for their help interpreting the employment surveys.

Contents

Figures

Tables

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 will lead to higher earnings and better jobs, covered by labor and social insurance regulations. Indeed, there has been a significant increase in schooling levels: in 1996, working-age Mexicans (18 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 high school degrees.

But despite these achievements and the fact that these two decades were characterized by macroeconomic stability and a large opening to international trade, hopes for higher earnings and better jobs have not materialized. As documented in this paper, the share of jobs covered by labor and social insurance regulations has remained essentially constant. And average hourly earnings, after recuperating from the 1995 financial crisis, have in fact fallen slightly, as a result of an absolute decline in the earnings of workers with more years of schooling. This has by-and-large offset the expected increase in average earnings associated with the change in the schooling composition of the labor force.

This paper argues that misallocation of resources—as evidenced by large differences in the productivity of resources across firms—explains these phenomena. The paper presents a preliminary exploration of the impact of misallocation on labor earnings and the returns to education. We have two basic hypotheses. First, in the particular case of Mexico, distortions result in too many resources allocated toward low-productivity firms that demand less-educated workers, depressing the earnings of workers with more education.[1] Second, over the last two decades these distortions have persisted even while the composition of the labor supply has changed toward more educated workers, implying a growing mismatch between the supply and demand for workers with more years of schooling.

Our paper can be seen as a bridge between the literature on misallocation and the literature on returns to schooling in the particular context of Mexico, where misallocation looms

[1] In this paper the word "distortions" is to be interpreted very broadly, as any market or regulatory failure or frictions in output, labor, and credit markets that cause wedges between the marginal revenue products of labor and capital across firms. Distortions can result from the interaction of many policies related to, among other things, taxation, credit, labor and social insurance regulations (including the nature of their enforcement). Distortions can also come from the absence of policies to correct for market failures, artificial barriers to entry or special subsidies to firms or sectors, or high registration or transaction costs.

large. The literature on misallocation starts from the premise that in the absence of distortions, individuals efficiently distribute themselves between entrepreneurs, employees, and self-employment. In turn, entrepreneurs hire the efficient number of employees given their abilities. The resulting distribution of individuals across occupations, and of firms across sizes, maximizes the productivity of the economy and the returns to factors. But if distortions are present, the distribution of individuals across occupations and the size distribution of firms are suboptimal: some individuals who should be employees are entrepreneurs (or vice versa), while some firms are larger (or smaller) than they should be given their underlying productivity. In parallel, firms may change the nature of the contracts offered to their workers. The implied misallocation of capital and labor lowers aggregate productivity and distorts the returns to factors (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008).

The empirical evidence for Mexico summarized below shows that misallocation is very relevant. While there is an important debate as to the exact nature of the distortions that induce this phenomenon, three results are robust: distortions result in large productivity losses; operate in the direction of allocating too much capital and labor to low-productivity firms that are less intensive in educated workers; and are persistent through time.

On the other hand, 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, at times called the wage premium, has narrowed over the last decade, if not before.[2] This finding is puzzling because, on the one hand, human capital is thought to be a constraint on growth in Mexico; and, on the other, the finding is the opposite of the trend found in the United States—by far Mexico's largest trading partner—where the wage premium has actually increased (Autor, Katz, and Kearney, 2008; Goldin and Katz, 2007).

In an immediate sense, of course, the fact that given the composition of the demand for labor, the earnings of workers with more years of education fall as their relative supply increases suggests a normal market adjustment. But this explanation is almost tautological, and begs the

[2] Lustig and López-Calva (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. Campos-Vázquez et al. (2010) and Campos, Esquivel and Lustig (2012) find that returns to schooling started to decline after 1994.

question as to why demand for workers with more education has lagged. Moreover, this explanation does not square with the empirical evidence for Mexico. As documented below, during the last two decades the earnings of workers who completed junior high school have increased relative to workers with a university education, despite the fact that the supply of the former has increased faster than the latter. This suggests that, in addition to supply-side considerations, the determinants of the schooling composition of the demand for labor have played a critical role.

Our bridge between the literature on misallocation, and that on returns to education, involves focusing on firms as an important observable determinant of workers' earnings, in a context where as a result of misallocation, the number, type, and size of firms participating in the demand side of the labor market is strongly distorted in the direction of low-productivity firms that are intensive in less-educated workers. The perspective taken here is that given workers' observable and unobservable characteristics, their earnings partly depend on the nature of the firms that employ them. In this context, we explore 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) affect the distribution of employees' 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 labor and social insurance regulation coverage are more intensive in educated workers than other firms. Third, there is a strong positive correlation between firms that are large and firms that offer their workers contracts with labor and social insurance regulation coverage. When, as a result of distortions, too few resources are allocated to these firms—as we document to be the case in Mexico—the schooling composition of the demand for labor will tilt in the direction of workers with fewer years of education.

Firms that offer contracts to their workers with coverage of labor and social insurance regulations are typically referred to as *formal* firms, while those that do not offer such coverage are known as *informal* firms. In turn, workers are referred to as formal or informal depending on whether they are employed by the former or the latter set of firms. Using that terminology, one can state that in Mexico distortions result in too many resources allocated to informal firms, and

in too many informal workers. In short, distortions generate large informality. Critically, however, what matters for our analysis is how firms behave, and not how they are labelled. Informality is a manifestation of distortions that result in a market equilibrium with too many low-productivity firms, low demand for educated workers, and jobs not covered by labor and social insurance regulation. Firm labels could change without changing the underlying distortions that determine firm behavior, in general, and the schooling composition of their demand for labor, in particular.[3] That said, and following standard practice, we will refer in this paper to firms and workers as formal and informal, but with the understanding that the focus is on the underlying phenomenon of misallocation, and not on the formal-informal labels.

A large literature has focused in understanding the role of taxation, social insurance and labor regulations, credit frictions, market failures, and other factors like registration and transaction costs in generating misallocation; see IDB (2010) for a summary. In the case of Mexico, Busso, Fazio and Levy (2012) and Levy (2008) have emphasized 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—in other words, a large informal sector. In parallel, an emerging literature is focusing on the links between misallocation and human capital, using models where the size distribution of firms is endogenous. Torres-Coronado (2015) analyzes the impact of size-dependent firm taxes on the returns to skills, and Busso, Neumayer, and Spector (2015) study the interaction between firm size and the distribution of skills. Finally, Bobba, Flabbi, and Levy (2016) focus on the impact of labor market distortions on the returns to education in a search-bargaining model.

In this paper we do not model the frictions or market or regulatory failures that generate misallocation and distort the size and type distribution of firms. Rather, we follow a three-step approach to test our hypothesis. In the first step, we estimate a model of individual workers' earnings that controls for all observable worker characteristics but focuses on estimating the coefficients associated with observable firm characteristics.

[3] This observation is relevant because policy can change the formal-informal firm labeling without affecting firm behavior. An example would be so-called "formalization programs" that offer subsidies to firms to register with social security authorities, but do not change underlying distortions in output, credit, tax, and labor markets faced by these firms, and therefore do not change their demand for labor. In this case, firm formality would increase while misallocation would persist. Similarly, self-employed workers may be offered subsidies to induce them to formalize, in which case labor informality could decline, but again without any changes in firm behavior.

In the second step, we consider the implications of eliminating misallocation, interpreted here as eliminating firm informality, only from the point of view of individual workers. To do this, we construct a counter-factual earnings distribution keeping constant individual workers' unobservable characteristics, as well as observable characteristics like years of schooling, age, gender, and location, but assuming that the size and type distribution of firms mimics that of formal firms. Our purpose is to measure how workers' earnings are affected when as a result of misallocation there are too many informal firms in the demand side of the labor market, independently of workers' education and abilities. In this context, we show that eliminating firm informality increases average earnings and changes the composition of the demand for labor, augmenting the demand for more educated workers relative to those with fewer years of schooling; and thus increasing the returns to schooling. The mean of earnings across all educational levels is higher and the distribution widens. Put differently, the distortions that misallocate resources toward informal firms act like a penalty on earnings that is paid by all workers but proportionately more by the more educated. Misallocation matters more to educated workers than to workers with little schooling.

In the third and final step we consider the aggregate implications of eliminating misallocation. We show that, given the supply of workers from each educational level, if the schooling composition of the demand for labor in the economy were that of formal firms, there would be an excess supply of workers with few years of education. We measure the size of excess supply and, critically, show that it would increase overtime. Next, for given values of the elasticity of substitution between workers of different schooling levels, we compute the changes in earnings required to absorb excess supply. We then compare the observed path of the ratio of earnings of workers with more versus less years of schooling (i.e., the wage premium) with alternative paths where there is no firm informality and where earnings adjust in each period to clear the market. We find that in the absence of firm informality the wage premium is substantially higher, and that the difference vis-à-vis the observed premium increases overtime. These results, in turn, suggest that the returns to education would be even higher than what our counter-factual simulations indicate.

Of course, if the distortions that misallocate resources toward the informal sector were removed, there would be further changes in the economy that our approach fails to capture. As the earnings of employees relative to those of the self-employed increase, there would be

changes in the distribution of workers between these two occupational categories. Participation rates would also change, as would the equilibrium level of unemployment. Further, even the size distribution of formal firms would change, probably in the direction of larger firms. Clearly, a model with more structure than what we present here is needed to fully measure these general equilibrium effects. Nonetheless, our preliminary results do show that by distorting the schooling composition of the demand for labor, misallocation is lowering the returns to education in Mexico, and that the persistence of misallocation in the face of increased schooling of the labor force goes a long way toward explaining the observed downward trend of the wage premium.

Section 2 of this paper briefly reviews the literature on the behavior of the wage premium and the returns to education in Mexico. Section 3 defines firm and worker informality and describes our data. Section 4 documents misallocation and describes the size and type distribution of firms, while Section 5 presents stylized facts on workers' earnings and schooling. Section 6 analyzes the impact of observable firm characteristics on workers' earnings and constructs a counterfactual scenario where firm informality is absent. Section 7 sheds light on the evolution of the wage premium over the 1996-2015 period. Section 8 presents concluding remarks.

2. The Wage Premium and the Returns to Education in Mexico

The general consensus regarding earnings and the returns to schooling in Mexico is that the premium paid to higher-skilled labor increased with the take-off of the North American Free Trade Agreement (NAFTA) in 1994 and then began to decline, with the returns to schooling following a similar trend.

Bouillon (2002) defines the wage premium as the ratio of wages of workers with more than a high school education to those with primary education or less, and finds that the premium rose between 1984 and 1994. Lächler (1998) finds an increase in the dispersion of earnings across different schooling levels from 1984-1994, and also finds that wages rose for workers with a high school education or more but fell for less-educated workers. Esquivel and Rodríguez-López (2003) use a different definition of wage premium, focusing on skilled and non-skilled workers (defined as non-production and production workers), and consider only manufacturing employment. In accordance with Bouillon (2002) and Lächler (1998), they find that the wage premium rose after 1988, but plateaued in the mid-1990s. Robertson (2007) defines the wage

premium as in Esquivel and Rodríguez-López (2003), and finds that it rose steeply prior to 1994, continued to rise slowly until 1999, and declined in the period up to 2005.

With regard to the returns to schooling, Campos, Esquivel, and Lustig (2012) find that relative returns to skilled workers (defined as those holding a high school degree or more versus junior high or less) increased between 1989 and 1994 but declined thereafter. Benita (2014) looks only at a subsequent period, from 2005 to 2012, and finds that the wage premium (defined in this case as the wages of university versus high school-educated male workers) declined for younger workers (ages 25-29), increased for older ones (ages 45-49), and remained constant for the oldest (ages 50-59). His findings suggest a large elasticity of substitution between workers with different levels of education (university and high-school workers appear to be interchangeable to employers).

Lustig, López-Calva, and Ortiz-Juarez (2014) focus on the 1990s and 2000s and define the wage premium as the returns to primary, secondary, and tertiary education versus no schooling or incomplete primary schooling. They find a decline in the returns to education, especially during the 2000s. This is in line with the decline in relative returns for high-skilled workers that Campos, Esquivel, and Lustig (2012) find for the 1994-2010 period. Lustig, López-Calva, and Ortiz-Juarez (2014) also discuss the various explanations given for this behavior: an increase in the supply of workers with higher educational attainment; a decline in the demand for skilled labor; a decline in the quality of higher education; and/or a mismatch in the supply and demand of skills. The authors note that none of these factors has been unambiguously identified.

Three observations are relevant from this brief review. First, studies vary in scope (manufacturing industry, the export sector, the overall economy), data sources, time periods, and groups used to define wage/skill premiums (by schooling levels or by workers' roles in the production process).[4] Second, despite this dispersion, on the whole there is agreement that after an initial widening following the start of NAFTA, the difference in mean earnings between workers with more versus fewer years of education narrowed, and that as a result the returns to education have declined.

[4] In some cases it is unclear whether "workers" refers only to employees or also to the self-employed; to female and male workers or only to the latter; or to formal and informal workers, or only to the former. In terms of data, studies use, inter alia, household surveys, employment surveys, administrative data from social security institutions, and sectoral surveys, particularly of manufacturing.

Third, even though not always explicitly noted, it is clear that average earnings over the last several decades have been strongly influenced by transitory macro shocks. In particular, it is not surprising that earnings rose in the late 1980s as the economy emerged from that decade's debt crisis, but then declined sharply immediately after the 1995 peso crisis. As a result, to isolate the effects of transitory macro shocks, studies have focused on the structure of earnings across educational groups, trying to disentangle the relative importance of changes in the composition of supply (like demographics and educational investments) from changes in the composition of demand (like NAFTA or technical change). In this context, it is somewhat surprising that studies have by-and-large ignored the role of one of the more salient characteristic of the Mexican economy: that as a result of misallocation, most firms, and workers are informal. The remainder of this paper makes a first attempt to assess the effects of misallocation on relative earnings and the returns to education.

3. Definitions and Data

3.1 Definitions

We distinguish between self-employed workers and workers engaged with firms. The latter—employees—can be so under salaried or non-salaried contractual relationships, 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 promotions and dismissals, and contribute to the worker's social insurance benefits. In turn, the worker has various rights, among them to unionize and to sue the boss if dismissed for an unjustified cause. Non-salaried workers, on the other hand, may be associated with a firm, but are not subordinated to it. Examples include workers who sell door-to-door, workers on a temporary contract performing a non-recurrent task and, very importantly for 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 non-salaried workers, or to observe regulations regarding dismissal, promotions, or minimum wages. Further, non-salaried workers cannot take a firm to court for dismissing them because there is no relation of subordination, nor can they unionize.

The distinction between salaried and non-salaried employment is the basis for the distinction between formal and informal workers. We define a formal worker as a salaried

employee covered by labor regulations regarding minimum wages, unionization, and dismissal, among other things, and who benefits from social insurance paid by the firm that hires him or her. All other workers, including the self-employed, are informal. If the law were fully enforced, all salaried workers would be formal. However, this is not the case in Mexico (Busso, Fazio, and Levy, 2012). This implies that informal workers consist of the self-employed, non-salaried 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 because they mix salaried and non-salaried contracts and at times violate the law. Table 1 identifies five possible combinations (all observed in the data). One implication of firms with mixed contracts is that they make it difficult to identify the formal and informal sectors with precision. Clearly, the former consists of at least those workers and firms in column two, while the latter consists of at least those workers and firms in columns three, six, and seven. But there are some firms that are neither purely formal nor informal, mixing salaried and non-salaried employees but complying with the law (column four), or partly violating it (column five).

Table 1. Formality Status of Firms and Workers

	Contracts between Firms and Employees					
	(2) Only salaried, compliant with law	(3) Only non-salaried, not obligated by law	(4) Mixed, but compliant with law	(5) Mixed, but not fully compliant with law	(7) Only salaried, not compliant with law	(8) **Self-employed workers**
Firm	Formal and legal	Informal and legal	Semi-formal and legal	Semi-formal and semi-legal	Informal and illegal	Not applicable
Worker	Formal	Informal	Salaried formal and Non-salaried informal	Salaried compliant formal; the rest informal	Informal	Informal

Source: Prepared by the authors.

3.2 Data

Our analysis relies on Mexico's employment surveys and economic census, and focuses on 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 National Employment Survey (*Encuesta Nacional de Empleo*—ENE); after 2005 it was known as the National Occupation and Employment Survey

(*Encuesta Nacional de Ocupación y Empleo*—ENOE). We refer to it here as the ENE-ENOE, a nationally-representative quarterly survey on type 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, and years of schooling; and other dimensions of a job like a written contract and yearly bonus payments. The ENE-ENOE also records hours worked and earnings net of taxes and contributions, whether they take the form of wages, salaries, commissions, or bonuses. We use data from the second quarter of each year and apply the average of the corresponding monthly price indices. All earnings are measured per hour in prices of May 2008.

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

Importantly for the econometric analysis in Section 6, the employment surveys have a panel structure that allows for following the same worker through five consecutive quarters. Since in each quarter we can identify firm size and worker's labor status, we can measure individual worker transitions in the course of a year across firm size and formal-informal status.

An issue with the ENE-ENOE is that some workers fail to report earnings (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 earnings in each of these categories, and randomly impute their earnings to workers with the same observable characteristics but without reported earnings.[5] The Appendix provides more details.

While the ENE-ENOE is a household-based employment survey, the Economic Census is a firm-based data set published every five years. We have four available censuses for 1998, 2003, 2008, and 2013, a subset of the years for which we have employment data. The census collects information on economic activity in urban areas in fixed establishments of all sizes

[5] To check the robustness of our procedure we compute earnings with data from Mexico's household survey which, as documented in the Appendix, does not have the same under-reporting problem as the ENE-ENOE. 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.

(henceforth, firms). Economic activity in rural areas, and on the streets in urban areas (street vendors, street markets, and so on) is not captured. Public sector employment is also excluded. As a result, the census only captures about 45 percent of the occupied population captured in the ENE-ENOE. However, as detailed below, we restrict the ENE-ENOE sample to private sector employees in firms in urban areas, ensuring that they are a subset of the employees captured in the census.

The census has data on the total number of workers and the aggregate of earnings and social security payments at the firm level, which allows us to classify firms according to the typology described in Table 1 (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 have data on individual workers within each firm, and thus provides no information on hours worked, gender, or years of education. On the other hand, the ENE-ENOE identifies the size of the firm where the worker is employed, so, critically, we can classify workers by years of schooling, firm size, and formal-informal status.[6] However, our data do not allow for mapping formal and informal *workers* as identified in the ENE-ENOE into the formality status of *firms* as identified in the census. Given that in our analysis we focus on the earnings of individual workers by years of schooling, firm size, and formality status, we mostly rely on ENE-ENOE data. However, we use the census to measure misallocation and the size and type distribution of firms between 1998 and 2013.

4. Misallocation and Firm Formality and Informality[7]

This section documents four sets of findings. First, most firms are small and informal, and they absorb a large share of employment. Second, in Mexico there is large and persistent misallocation of resources. Third, misallocation systematically operates in the direction of channeling too many resources to informal firms. Lastly, between 1998 and 2013 the number of formal firms declined, while that of informal ones increased. Over the same period employment grew substantially more in informal firms.

[6] We do this following the methodology developed by INEGI (2014).
[7] This section borrows heavily from Busso, Fazio, and Levy (2012) and Busso and Levy (2016).

4.1 Size and Type Distribution of Firms

Table 2 serves to make four observations on the distribution of firms and employment in 2008. First, 89.7 percent of all firms are very small (up to five workers); and 90.1 percent are purely informal (legal and illegal). Second, purely informal firms have 2.9 workers, on average, but account for more than half of all employment captured in the census (54.5 percent). On the other hand, purely formal firms have 32.2 workers on average (and 22.3 if we extend the definition of formal firms to include mixed firms). Altogether these firms account for 45.5 percent of employment captured in the census.

Table 2. Size and Type Distribution of Firms and Employment, 2008
(in percent)

	Legal & Formal	**Legal & Informal**	**Legal & Semi-formal**	**Semi-legal & Semi-formal**	**Ilegal & Informal**	**Total**
Firms*						
[0-5 workers]	1.21	65.60	1.18	2.49	19.25	**89.73**
[6-10]	0.69	1.11	0.33	1.23	2.46	**5.82**
[11-50]	0.87	0.56	0.15	1.10	0.90	**3.58**
[50+]	0.29	0.19	0.03	0.30	0.05	**0.86**
Total	**3.06**	**67.46**	**1.69**	**5.12**	**22.66**	**100.00**
Workers**						
[0-5]	0.77	23.60	0.79	1.69	10.94	**37.79**
[6-10]	1.09	1.63	0.50	1.92	3.66	**8.80**
[11-50]	3.84	2.39	0.57	4.72	3.36	**14.88**
[50+]	14.67	7.74	1.19	13.75	1.18	**38.53**
Total	**20.37**	**35.36**	**3.05**	**22.08**	**19.14**	**100.00**

Source: Busso, Fazio, and Levy (2012). *3.643 million; **17.655 million.

Second, note that most very small firms are informal but legal. When a firm has few workers, it is easy to monitor effort, coordinate activities, and reach agreements to distribute profits. This is the case of cooperatives or, more relevant in Mexico, firms where workers are related (family firms). In these cases it may be efficient for the firm to establish non-salaried contracts. Indeed, as shown in Table 2, this is the most common contractual structure of firms in Mexico.

Our third observation is that as the number of workers in the firm increases, it is more difficult to coordinate tasks unless there are relationships of subordination, and to observe individual effort. Problems of shirking and free-riding thus soon appear. Moreover, if the production technology calls for a fixed place of work and close coordination between tasks

performed by different workers at the same time, salaried contracts will be more efficient.[8] As a result, on average, larger firms have proportionately more salaried contracts than smaller ones: firms with only salaried workers, formal and informal, constitute only 25.7 percent of all firms but employ 39.5 percent of all workers.

Our fourth observation is that considering only firms with salaried contracts, formal firms are larger than informal ones: on average, 32.2 versus 4.1 workers. This results from the imperfect enforcement of laws regarding salaried contracts. Since the probabilities of being detected by the authorities are proportional to the size of the firm, firms with illegal salaried contracts tend to be small, as the expected marginal costs of labor increase sharply with size (Anton, Hernandez, and Levy, 2012).

This brief discussion about the type and size distribution of firms is central to our analysis of labor earnings and the returns to education in Mexico. This is because firms' demands for workers of various schooling levels depend on those firms' size and formal/informal status. Transportation services can be provided by a hundred self-employed workers driving their own trucks, or by a single firm with a hundred salaried employees. In both cases there will be a hundred trucks and a hundred drivers, but in the latter case 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 needing engineers; the same holds for apparel and food processing, among many manufacturing activities. And the same holds for retail commerce: it can be carried out in small stores employing workers with only basic literacy and numeracy or through large supermarket chains requiring, say, industrial designers. In general, the complexity of tasks and the division of labor increase with firm size, and generate a need for more educated workers. But the issue is not only size: a small informal firm producing jeans for sale 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.

As discussed, the census data unfortunately contain no information on the schooling composition of workers. However, in Table 9 later in this paper we use the ENE-ENOE data to document two additional empirical regularities that are also central to our analysis: that larger

[8] These differences matter for earnings: for salaried workers they take the form of wages (a fixed amount of money per unit of time), while for non-salaried workers they take the form of payments per product regardless of the time required to produce it, profit-sharing, or commissions based on sales.

firms are more intensive in workers with more years of schooling than smaller ones; and that controlling for size, informal firms are less intensive in educated workers than formal ones.

4.2 Large and Persistent Misallocation

We next turn to evidence of misallocation. Following Hsieh and Klenow (2009), we define total revenue productivity, $TFPR_{is}$, as the value of the output 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 that firm). In turn, $TFPQ_{is}$ is the physical productivity of resources (a weighted average of the marginal products of labor and capital). In the absence of any distortions that would misallocate resources across firms, revenue productivity would be the same for all firms in a given sector and across all sectors. This implies that the greater the dispersion of TFPR, the greater the degree of misallocation. Table 3 presents three measures of the dispersion of TFPR and TFPQ.[9]

Table 3. Measures of Dispersion of Firm Productivity, 1998-2013

	1998		2003		2008		2013	
	TFPR	TFPQ	TFPR	TFPQ	TFPR	TFPQ	TFPR	TFPQ
Standard deviation	1.15	1.75	1.14	1.77	1.23	1.90	1.24	1.85
p75 - p25	1.55	2.44	1.50	2.38	1.60	2.60	1.55	2.48
p90 - p10	2.95	4.58	2.91	4.57	3.14	4.91	3.14	4.77

Source: Busso and Levy (2016).

Two facts follow from Table 3. First, there is a large dispersion in revenue productivity across firms, implying substantial misallocation of resources. For example, in 2013 the firm in the 75th percentile of the revenue productivity distribution was 55 percent more productive than the one in the 25th percentile.[10] Second, this dispersion persisted (and in fact increases slightly) over the 15 years considered, indicating the persistence of misallocation of capital and labor.

[9] Computations are done at the six digit sector level and include firms of all sizes in manufacturing, services, and commerce. There are 559 sectors in 1998, 699 in 2003, 707 in 2008, and 735 in 2013. Comparisons of TFPR and TFPQ are only made for firms within the same sector. The numbers in Table 3 are averages across all sectors. As expected, dispersion of TFPR is smaller than that of TFPQ, because when firms produce more physical output they sell at lower prices (so that revenue-based measures of productivity tend to underestimate variation in producers' physical efficiencies). See Syverson (2011).

[10] The difference between firms in the 90th and 10th percentile is 214 percent. This compares with a difference of 92 percent for the same range in the manufacturing sector of the United States, as reported by Syverson (2004). Importantly, Syverson's computations are carried out at the four-digit level, and one would expect smaller differences at the six-digit level. IDB (2010) compares the dispersion of revenue productivity between manufacturing firms in the United States and Mexico at the four digit level, and finds that dispersion is substantially higher in Mexico. Busso, Madrigal, and Pages (2013) find that dispersion is also higher in Mexico compared to other Latin American countries.

4.3 Productivity Differences between Formal and Informal Firms

Table 4 compares productivity across firm types.[11] In all years, formal firms are more productive than all other firms. If we focus only on pure informal firms, legal and illegal, ignoring mixed firms, it turns out that depending on the year considered, their physical productivity is between 158 and 28 percent lower than that of pure formal firms. If we focus on revenue productivity, the differences are between 60 and 10 percent. Note that legal informal firms, which as shown in Table 2 constitute the majority of firms in Mexico, are always the least productive of all, and that productivity differences between these firms and illegal informal ones are significant. That said, there are two critical results for our purposes. First, 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. Second, the fact that this result is observed in all periods considered shows that distortions operate systematically in the same direction.

Table 4. Productivity Differences by Firm Type, 1998-2013
(In percent, relative to formal legal firms)

	1998		2003		2008		2013	
	TFPQ	**TFPR**	**TFPQ**	**TFPR**	**TFPQ**	**TFPR**	**TFPQ**	**TFPR**
Legal & Semi-formal	-0.635 (0.0064)	-0.360 (0.0042)	-0.497 (0.0058)	-0.340 (0.0038)	-0.672 (0.0072)	-0.418 (0.0046)	-0.370 (0.0066)	-0.394 (0.0041)
Legal & Informal	-1.585 (0.0043)	-0.582 (0.0028)	-1.475 (0.0040)	-0.644 (0.0026)	-1.130 (0.0043)	-0.467 (0.0027)	-1.404 (0.0035)	-0.690 (0.0022)
Semi-legal & Semi-formal	-0.096 (0.0039)	-0.010 (0.0025)	-0.012 (0.0035)	-0.049 (0.0022)	0.035 (0.0041)	-0.053 (0.0026)	-0.058) (0.0036)	-0.185 (0.0023)
Illegal & Informal	-0.541 (0.0045)	-0.201 (0.0029)	-0.285 (0.0040)	-0.102 (0.0026)	-0.488 (0.0044)	-0.208 (0.0028)	-0.414 (0.0041)	-0.157 (0.0025)
Observations	2,138,976		2,398,341		2,192,322		2,694,7120	
R-squared	0.379	0.066	0.407	0.071	0.381	0.053	0.371	0.069

Source: Busso and Levy (2016).

[11] These ordinary least square regressions capture the log of firm *i* productivity in sector *s* over the log of average productivity of all firms in that sector (i.e., we only compare firms in the same sector). Regressions include controls for firm size and age and, without implying causality; show that some firm types are systematically more productive than others. Standard errors are in brackets. All coefficients are significant at the 99 percent level; see Busso and Levy (2016), who also perform these regressions separately for each size group and report similar results, but with larger productivity differences among smaller firms.

4.4 Evolution of Firm Size and Type between 1998 and 2013

Table 5 presents data on the size and type distribution of firms over the 1998-2013 period. Notice first that average firm size is constant. But despite this constancy, there are large compositional changes: the number of small and very small firms increases more than that of medium-size and large firms. In parallel, the number of formal firms *falls* (by over 12 percent) while that of informal firms *rises* quite substantially, by 66 percent.

Table 5. Size and Type Distribution of Firms and Employment, 1998-2013
(Thousands)

	1998	2013	Percent Change
Totals			
Firms	2,726.3	4,169.5	52.9
Employment	12,777.5	19,511.8	52.7
Average firm size	4.68	4.67	0.0
Firms by size			
[0-5]	2,479.7	3,802.6	53.3
[6-10]	123.4	195.3	58.2
[11-50]	97.3	136.7	40.4
[51+]	25.8	34.8	34.8
Firms by type			
Formal*	459.6	402.5	-12.4
Informal**	2,266.7	3,767.0	66.1
Employment by firm size			
[0-5]	4,352.9	7,010.0	61.0
[6-10]	920.7	1,441.7	56.5
[11-50]	2,020.8	2,855.9	41.3
[51+]	5,483.0	8,204.1	49.6
Employment by firm type	8,092.8	8,949.1	10.6
Formal*	4,684.7	10,562.7	125.4
Informal**			

Source: Prepared by authors.
*Sum of Formal and Legal, Legal and Semi-formal, and Semi-legal and Semi-formal. **Legal and Illegal; see Table 1.

Employment in firms captured in the census reflects these changes, and increases more in smaller than in larger firms (i.e., in firms that are less intensive in educated workers). Moreover, these differences are substantially magnified when we focus on firm types: employment in informal firms increases by 125.4 percent versus 10.6 percent in formal ones. These trends are noteworthy given that, as documented in the next section, 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 the reason behind them.

5. Workers' Demographics, Schooling, and Earnings

This section provides descriptive statistics on workers. We document the changes in the schooling composition of the labor force, the path of earnings, and the formal-informal composition of employment. We also present estimates of the returns to education between 1996 and 2015.

5.1 Descriptive Statistics

We define the working-age population (WAP) as all persons 18 years of age or older, and the economically active population (EAP) as the subset of the WAP that participates in the labor market. We interpret the WAP as the potential supply of labor, determined by demographics and schooling investments, and the EAP as the observed supply (equals demand) of labor, given the participation rate of each schooling group.

The sample of workers on which we focus 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. This group represents approximately 20 percent of the EAP, and is the most urbanized and educated segment of Mexico's labor force. Table 6 presents descriptive statistics for the first and last year of the period under study.

Table 6. Workers' Descriptive Statistics, 1996–2015
(Percent)

	Annual Average Growth Rate			Composition					
				1996			2015		
	WAP	EAP	Sample	WAP	EAP	Sample	WAP	EAP	Sample
Incomplete primary	-1.22	-1.68	-3.39	23.80	20.99	8.02	11.59	9.26	2.71
Complete primary	0.85	0.81	-1.12	24.60	23.05	19.67	17.77	16.34	10.22
Incomplete junior high	0.44	0.45	-1.99	4.09	4.42	5.23	2.73	2.93	2.31
Complete junior high	5.03	5.03	3.45	15.51	16.62	20.06	24.80	26.24	24.90
Incomplete high school	1.19	0.75	0.00	13.05	13.11	19.33	10.07	9.18	12.44
Complete high school	6.18	6.16	5.77	6.46	6.95	9.46	14.40	15.24	18.23
University	4.67	4.39	4.79	12.47	14.85	18.23	18.62	20.79	29.20
All	2.20	2.31	2.32	100.0	100.0	100.0	100.0	100.0	100.0
Years of schooling	.	.	.	4.7	7.8	9.8	9.2	9.8	11.5

Source: Prepared by authors.

We highlight five findings. First, the rates of growth of the EAP and the sample slightly exceed those of the WAP, as a result of a gradual increase in the aggregate participation rate (from 61.6 to 62.8 percent). Second, a rapid increase in working-age persons with completed junior high or more education leads to 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 had changed to 29.3 and 33 percent, respectively.

Third, the rates of growth of the EAP by schooling level by-and-large mimic those of the WAP, implying relatively constant participation rates.[12] This contrasts with the sample, where the employment of workers with incomplete junior high or less schooling falls in absolute terms. As a result, fourth, the schooling composition of our 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, fifth, our sample has more years of schooling than the EAP: in 2015, workers who had at most completed primary schooling represented 25.6 of the EAP, but only 12.9 of our sample; at the other extreme, workers with at least completed high school were 36 percent of the EAP, but 47.3 percent of our sample.

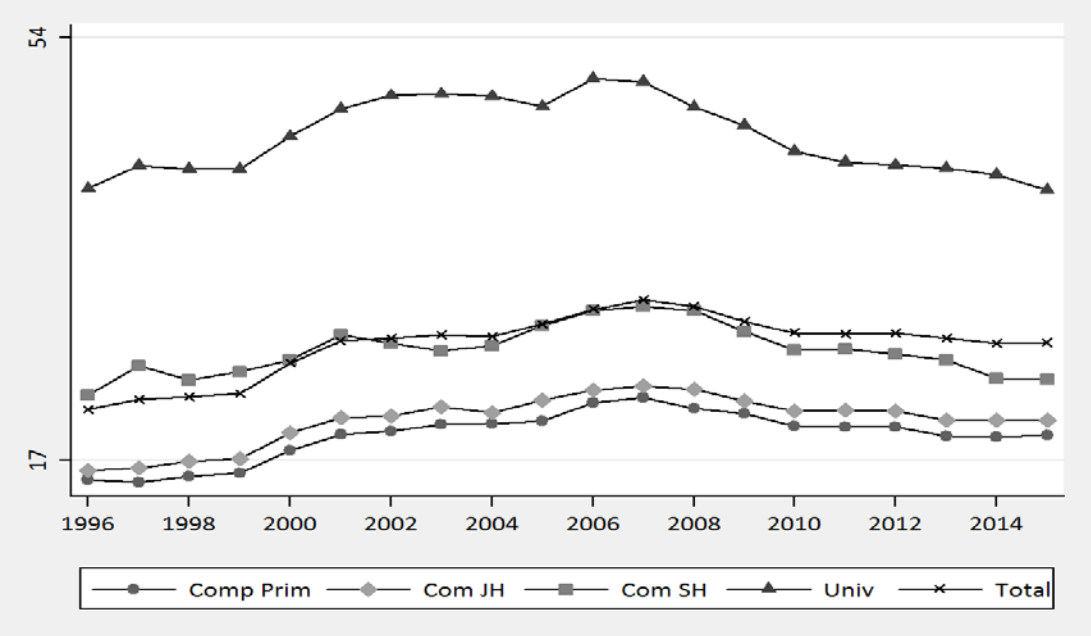
Figure 1 presents the evolution of earnings for our sample of workers. Earnings increase for all groups up to 2003, reflecting the recovery from the sharp fall during the 1995 financial crisis.[13] After that, earnings stagnate for employees with completed primary and junior high, and *fall* for those with completed high school and university education. In fact, remarkably, by 2015 earnings for the latter group are 7 percent *below* their 1996 level, while for the former group they are the same as in 1996. The contrast with employees with completed junior high is very revealing: as seen in Table 6, the supply of persons in this educational group also grows rapidly, in fact faster than those with a university education. Yet, earnings of employees with completed junior high do not fall, and in 2015 are 26 percent *higher* compared to 1996. These asymmetries clearly indicate that there are other forces aside from changes in supply determining the behavior

[12] Participation rates are constant for each schooling group, but higher for those with more years of schooling. Thus, the increase in the aggregate participation rate reflects mostly a change in the schooling composition of the labor force.

[13] Unfortunately, there are no employment surveys for 1992-94, and the available ones for earlier years cannot be compared with the ENE-ENOE. However, to corroborate our statements, the Appendix presents data from Mexico's household surveys that show, i) a sharp fall in earnings in 1995 as a result of that year's financial crisis, ii) that despite the increase observed after 1996, by 2012 earnings had yet to reach the levels observed in 1994, and iii) a similar trend in earnings after 1996 in the two surveys (contrast Figure 1 with Figure A.3 in the Appendix).

of earnings. The up-shot of the different path of earnings for each educational group (together with changes in their relative shares) results in a decrease in average employee earnings (labeled "total" in Figure 1), which by 2015 are the same as in 2000.

Figure 1. Sample Employee Earnings, 1996–2015
(May 2008 pesos per hour)



Source: Prepared by the authors.

Table 7 presents descriptive statistics for our sample of workers. 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 narrow overtime. There are no relevant changes in hours worked per week over the period, though formal employees work slightly more hours than informal ones (but again the differences are small). The share of women in the sample increases, reflecting higher female participation rates; overtime, the share of women in informal employment increases. Finally, as expected, formal employees have more years of schooling than informal ones.

Table 7. Characteristics of the Sample of Workers

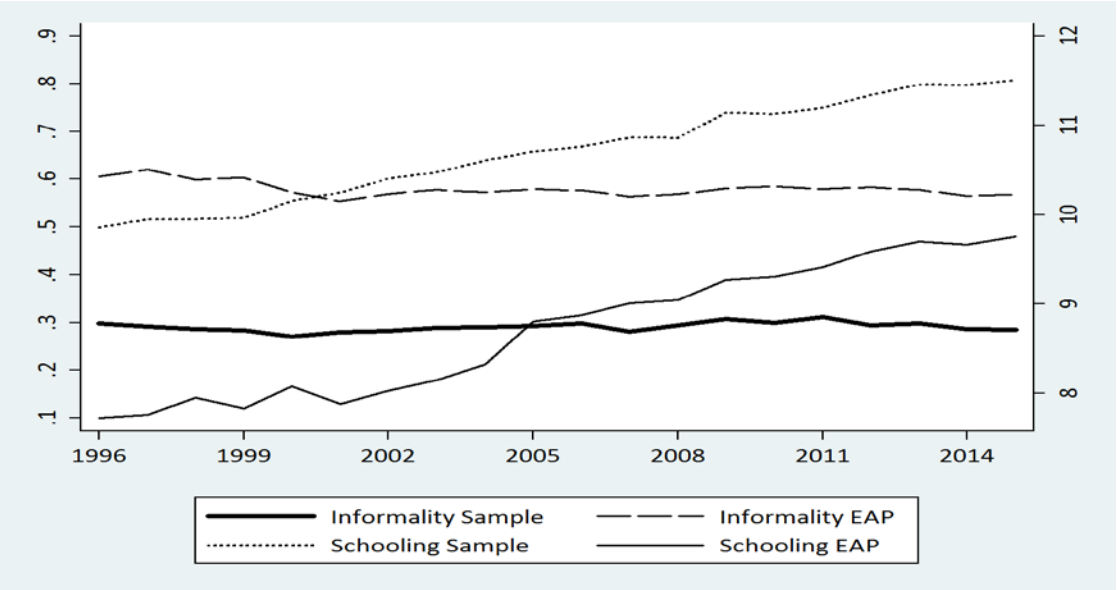
	1996			2006			2015		
	All	**Formal**	**Informal**	**All**	**Formal**	**Informal**	**All**	**Formal**	**Informal**
Mean age (years)	32.17	32.69	30.96	34.58	35.13	33.29	36.46	36.62	36.04
Median age (years)	30	31	28	33	34	31	35	36	34
Hours worked	42.24	42.41	41.85	42.48	42.83	41.66	42.79	43.16	41.84
Years 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		
Education shares (%):									
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 high school	19.33	21.44	14.38	15.57	17.04	12.13	12.44	13.06	10.88
Complete high school	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: Prepared by the authors.

Figure 2 depicts the paths of the share 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 share is lower for our sample, as we exclude the self-employed, rural workers, and urban employees working less than 30 hours a week or who live in relatively less urbanized localities. That said, it is remarkable that both shares are practically constant through time, despite the fact that 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.[14]

[14] Levy and Szekely (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 2 are not a result of changes in the age composition of the labor force.

Figure 2. Years of Schooling and Share of Informal Employment, Economically Active Population and Sample, 1996-2014



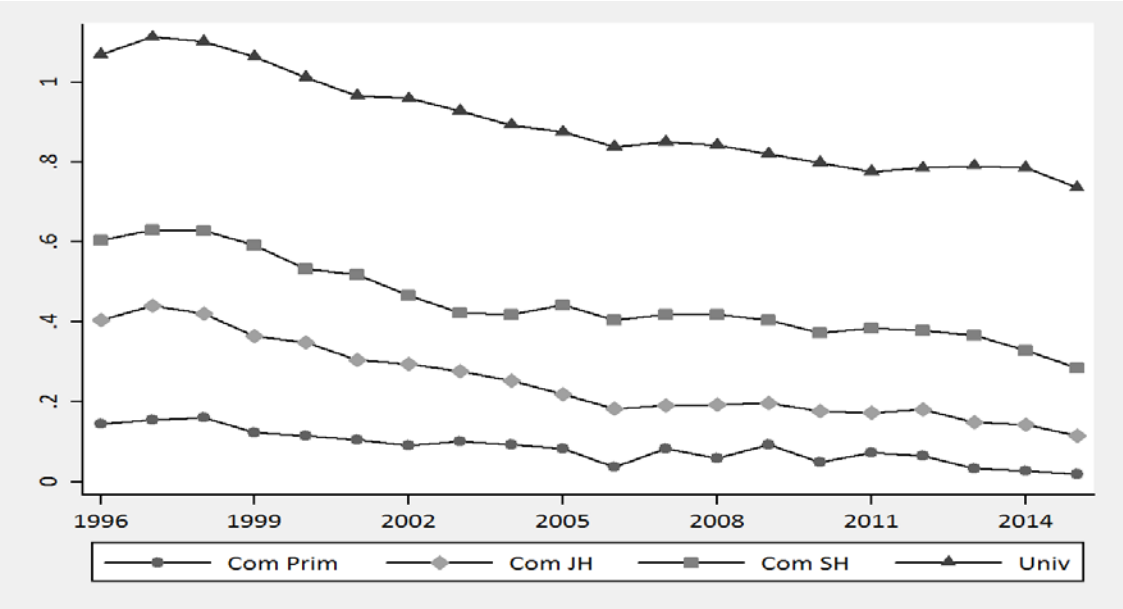
Source: Prepared by the authors.

5.2 Returns to Education: Preliminary Overview

Figure 3 depicts the returns to education for the 1996-2015 period.[15] The points in this figure are the coefficients for each year from an ordinary least squares regression where the excluded category is incomplete primary, and where we incorporate controls for age, experience, and the municipality where workers are located. The structure of returns is consistent with the findings in the literature reviewed in Section II. As can be seen, over the 20-year period returns fall for all groups, although the declines are more pronounced for employees with more years of education.

[15] Similar results are found using the household surveys; see Figure A.4 in the Appendix.

Figure 3. Returns to Education, 1996-2015
(In percentage, relative to incomplete primary education)



Source: Prepared by the authors.

Finally, to motivate the analysis in the next section, Table 8 presents the estimates for 2006—the mid-point of the period considered—of the returns to education separating formal from informal workers. The contrast between them is striking. Returns are higher for formal workers, and significantly so. These results are puzzling, insofar as we are comparing employees who have very similar observable characteristics, as was seen in Table 7 (and who, as will be shown in Table 11, move between sectors). Both groups of employees live in the most urbanized areas of Mexico, work almost the same number of hours, and have very similar gender composition. Formal employees are slightly older, but the difference is small and in any case it is controlled for. Are these large differences only the result of unobservable differences in talent and abilities between formal and informal employees? Or are these differences partly explained by observable characteristics of the firms in which these workers are employed?

Table 8. Returns to Education, Sample, 2006
(In percent, relative to incomplete primary)

	(1) All	(2) Formal	(3) Informal
Age	0.0428***	0.0430***	0.0299***
	(455.20)	(367.78)	(198.19)
Age # Age	-0.000405***	-0.000376***	-0.000320***
	(-324.78)	(-247.20)	(-159.35)
Complete primary	0.0474***	0.0398***	0.0283***
	(56.18)	(32.01)	(25.09)
Incomplete junior high	0.163***	0.212***	0.0753***
	(144.74)	(130.16)	(49.66)
Complete junior high	0.188***	0.206***	0.0763***
	(233.65)	(174.28)	(68.96)
Incomplete high school	0.330***	0.368***	0.106***
	(391.79)	(304.82)	(83.90)
Complete high school	0.409***	0.441***	0.210***
	(477.19)	(361.82)	(159.91)
University	0.849***	0.879***	0.564***
	(1047.62)	(755.34)	(455.95)
Observations	7,922,128	6,108,218	2,369,543
Adjusted R^2	0.359	0.363	0.229
Controls	Municipality	Municipality	Municipality

Source: Prepared by the authors.
Note: t statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6. Misallocation, Earnings, and the Returns to Education, 2006

We now turn to analyzing the effects of misallocation, as reflected in a large informal sector, on employees' earnings and returns to education. This section uses the ENE-ENOE data and focuses on the mid-point of our period, 2006; the next section covers the 1996-2015 period. We document differences in the schooling composition of the workforce between formal and informal firms, construct a panel of workers following them for up to five quarters, and present regressions by schooling level to measure the impact of firm formality on earnings, exploiting the panel structure of our data to control for time-invariant unobserved workers characteristics that could be correlated with employment status. Finally, we construct counter-factual simulations of earnings and the returns to education assuming that all firms behave as formal firms.

6.1 Formal-Informal Firm Differences

Table 9 shows the distribution of employment by schooling level and firm size in formal and informal firms.[16] 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 percent and reflect the schooling composition of firms' workforces by 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 9. Distribution of Employees by Education and Firm Size, 2006
(Percent)

	Informal Firms					Formal Firms				
	1-5	6-10	11-50	51+	**Total**	1-5	6-10	11-50	51+	**Total**
Incomplete primary										
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**
Complete primary										
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**
Incomplete junior high										
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**
Complete junior high										
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**
Incomplete high school										
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**
Complete high school										
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**
University										
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**
Total										
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: Prepared by the authors.

[16] As noted, the ENOE data refer to workers, not firms. If there were no mixed firms, all informal workers would be employed by informal firms and all formal workers by formal firms. The presence of mixed firms complicates this picture. Unfortunately, we cannot use ENOE data to correct for this problem, and Table 9 is the best approximation of differences in the schooling composition of the workforce between formal and informal firms. Ideally one would also want to separate firm types as in Table 1, but this can only be done with the census data, and thus we only have two firm types.

Three findings are worth highlighting. First, focusing down the columns and considering the totals for each sector, we find that formal firms are more intensive in educated workers: 42.6 percent of their workforce has at least completed high school while 14.3 percent has at most completed primary school, in contrast to 24.1 and 29.4 percent, respectively, for informal firms. Second, these patterns hold controlling for firm size. Thus, for instance, 16.8 percent of the workforce of informal firms with up to five workers has at least a high school education, while that figure is more than double for formal firms (38.5 percent). At the opposite end, 34.6 percent of the workforce in informal firms with up to five workers has at most completed primary school, versus almost half that, 17.8 percent, for formal firms.

Third, for any educational level, the distribution of workers across firm size is also very different, reflecting the fact that formal firms are on average larger. Thus, 72.1 percent of all informal workers with incomplete primary are employed in firms with up to five workers versus 11.8 percent of all formal workers with the same schooling level. For workers with a university education the corresponding numbers are 33.9 and 4.6 percent. Altogether, the last row of Table 9 shows that 62.9 percent of workers of all educational levels are employed in informal firms of up to five employees (versus 5.8 percent in formal firms), while in firms of 50 workers or more that figure is only 8.4 percent (versus 57.4 percent in formal firms).

Table 10 presents the distribution of yearly bonus payments and written contracts by firm size and workers' educational level and sector.[17] Briefly, although there are small variations by firm size and schooling levels, the vast majority (95.2 percent) of workers in the formal sector receives a yearly bonus and has a written contract (90.9 percent). This contrasts with workers in the informal sector, where only 22 percent receive a yearly bonus and 13.1 percent have a written contract. Further, differences across informal firms of various sizes are larger than across formal ones, with smaller firms showing lower rates of bonus payments and even lower rates of written contracts. Finally, note that in all cases the proportions of workers receiving bonuses and having written contracts broadly increase with educational levels, particularly for contracts.

[17] The yearly bonus is an additional month of salary paid at the end of the year (called an *aguinaldo*). Since all our earnings are measured per hour, in our computations below we add the pro-rata per hour amount.

Table 10. Bonus and Contracts by Education and Firms Size, 2006
(Percent)

	Informal Firms					Formal Firms				
	1-5	6-10	11-50	51+	**Total**	1-5	6-10	11-50	51+	**Total**
Yearly Bonus										
Incomplete primary	17.24	15.90	31.27	15.62	**18.69**	82.99	86.88	92.82	95.71	**92.39**
Complete primary	20.42	28.21	19.82	33.98	**21.66**	90.24	88.30	92.18	96.82	**94.41**
Incomplete junior high	14.06	26.44	15.50	3.46	**15.44**	92.44	94.11	85.79	96.31	**92.62**
Complete junior high	19.67	29.21	26.08	16.32	**21.47**	82.92	92.72	93.59	96.37	**94.65**
Incomplete high school	19.05	36.17	25.03	32.92	**23.65**	91.47	91.26	95.62	96.84	**95.71**
Complete sigh school	18.65	24.31	29.72	26.03	**22.28**	92.80	92.76	93.13	96.87	**95.05**
University	17.58	25.38	32.20	36.57	**27.04**	84.37	96.62	96.41	97.73	**96.64**
Total	**18.92**	**27.26**	**26.73**	**28.33**	**22.02**	**87.60**	**92.73**	**94.19**	**96.93**	**95.27**
Written Contract										
Incomplete primary	0.59	4.48	16.75	33.03	**4.47**	39.69	60.24	86.86	91.22	**80.77**
Complete primary	1.43	6.58	14.79	48.72	**5.40**	36.50	61.87	82.37	94.96	**85.39**
Incomplete junior high	2.06	3.10	18.84	52.74	**7.41**	47.76	84.86	83.51	92.90	**86.96**
Complete junior high	2.07	14.17	20.26	33.19	**8.16**	51.66	80.66	88.41	95.97	**90.53**
Incomplete high school	3.10	13.93	44.00	48.96	**17.30**	59.80	79.07	90.16	97.08	**91.30**
Complete high school	4.55	23.26	36.54	45.18	**18.39**	71.36	80.40	94.19	96.55	**92.99**
University	9.38	23.78	46.88	70.94	**36.54**	69.06	83.24	93.43	96.82	**93.65**
Total	**2.58**	**14.00**	**30.44**	**51.99**	**13.16**	**57.64**	**78.16**	**90.21**	**96.16**	**90.90**

Source: Prepared by the authors.

Summing up, we find that larger firms are more intensive in educated workers than smaller ones; that controlling for size the same is true of formal firms compared to informal ones; that workers of the same educational level are distributed very differently across firms sizes in the formal and informal sector; and that controlling for educational levels, rates of yearly bonus payments and written contracts are significantly higher for formal firms.

6.2 Employee Mobility across Firm Size and Type

Table 11 documents a central feature of Mexico's labor market: the large mobility of workers of all educational levels across sizes and types of firms. We construct a panel of all workers in our sample surveyed in the second quarter of 2006, following them through five consecutive quarters. In one extreme, some workers were observed in the four previous quarters and are last surveyed in 2006.Q2; at the other, some first entered the survey in 2006.Q2 and were observed for the next four quarters. The rest corresponds to intermediate cases with a mix of up to three quarters before or after 2006.Q2. Altogether we span the period from 2005.Q3 to 2007.Q1, and record the number of times that each individual worker changed firm type or size. Changes in firm type are from formal to informal or vice versa, while changes in firm size are from any one

of the four firm sizes considered to any of the other three sizes (i.e., they can be changes from larger to smaller firms or vice versa).

Table 11. Mobility of Individual Workers across Firms Size and Type, 2005.Q3-2007.Q1
(Shares)

	Firm Type	Firm Size	Number
Education	**Change* (%)**	**Change* (%)**	
Incomplete primary	15.26	31.86	426,013
Complete primary	16.41	33.93	1,226,837
Incomplete junior high	20.09	36.84	361,622
Complete junior high	19.33	38.98	2,453,960
Incomplete high school	21.02	44.83	1,666,239
Complete high school	20.3	45.99	1,549,004
University	21.15	47.67	2,749,087

Source: Prepared by the authors.
*Change refers to at least one movement over one year.

As can be seen, mobility is very high even for such a short period of time, and in fact increases slightly with educational levels. During one year, anywhere between 15 to 21 percent of workers change status, from formal to informal or vice versa. Mobility across firm sizes is even higher, and again increases with educational level: for workers with incomplete primary or just primary education, around 32 percent changed firm size during one year, whereas almost 48 percent of workers with a university education did so over the same period. This level of mobility, consistent with analyses presented elsewhere (e.g., Levy, 2008), allows us to more robustly estimate the effects of simulated changes in the size and type distribution of firms on workers' earnings.

6.3 Earnings Regressions by Educational Level with Workers' Fixed Effects

To measure the importance of observable firm characteristics on workers' earnings, for each educational level we estimate a within-educational group earnings function controlling for firm characteristics and an individual's observable characteristics. Since 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 returns to firm characteristics

may be biased upward if there is an unobserved workers' ability that is correlated with workers' 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 high enough to allow for identification using individuals' fixed effects. The panel regression with worker fixed effects, with all earnings measured per hour, is as follows:[18,19]

$$(1) \quad Log\ y_{it;e} = D_{i;e} + \beta_e X_{it;e} + \gamma_e Z_{it;e} + \partial_e FS_{it;e} + \epsilon_{it;e}$$

where $y_{it;e}$ is earnings for worker *i* of educational level *e*, at quarter *t*; $D_{i;e}$, is an individual fixed effect for individual *i* with educational level *e*, and $X_{it;e}$ is a vector of characteristics of individual *i* with educational level *e* at quarter *t* (age, experience, gender, location), with associated coefficients $\boldsymbol{\beta}_e$; $Z_{it;e}$ is a vector of firm characteristics for individual *i* with educational level *e* at quarter *t,* and $\boldsymbol{\gamma}_e$ is a vector of associated coefficients. We also include a dummy for whether the worker is formal or informal, and a set of interactions between formality and observable firm characteristics ($FS_{it;e}$). Finally, there is an error term $\boldsymbol{\epsilon}_{it;e}$, which is assumed to be uncorrelated with $Z_{it;e}$ and $FS_{it;e}$, which implies that firm size and characteristics are not correlated with unobserved time-varying individual characteristics.[20] Results are shown in Table 12.

[18] To test for the potential existence of this bias we ran a simple cross-section ordinary least square regression. Indeed, the returns to firm size are about 15 to 20 percent higher in the specification without fixed effects, consistent with the hypothesis of unobserved time-invariant characteristics associated with workers sorting into firms of different sizes.

[19] Bonus payments are excluded from both sides of the regression, but are then added to workers' earnings in proportion to the rates observed for each educational level and firm size, as specified in Table 10.

[20] This assumption may still imply a bias if there were time-varying individual characteristics associated with selection of workers into firms (such as training or investments in skills). The best way to deal with that problem would be to match employer-employee data over time, but such data are not available for Mexico. If any, the magnitude of this bias—particularly in terms of its relative importance across educational levels—should not affect the counterfactual simulations in the next section.

Table 12. Earnings Regressions by Educational Level, Anchor 2006.Q2, Panel 2005.Q3-2007.Q1

(Excluded firm size is that of 0-5workers)

	(1) Incomplete primary	(2) Complete primary	(3) Incomplete junior high	(4) Complete junior high	(5) Incomplete high school	(6) Complete high school	(7) University
Age	0.0368***	0.0382***	0.103***	0.0143***	0.0688***	0.000694	0.0431***
	(52.19)	(89.27)	(62.00)	(47.15)	(127.57)	(1.21)	(101.47)
Age * Age	-0.000539***	-0.000496***	-0.00131***	-0.000137***	-0.000890***	0.000164***	-0.000480***
	(-66.80)	(-94.28)	(-54.43)	(-32.66)	(-127.02)	(21.20)	(-91.35)
6-10 workers	0.108***	0.0482***	-0.0593***	0.0197***	0.0884***	0.0870***	0.0695***
	(85.39)	(53.04)	(-42.03)	(26.54)	(71.11)	(62.12)	(48.52)
11-50 workers	0.130***	0.0569***	0.0883***	-0.00508***	0.108***	0.0217***	0.156***
	(99.33)	(58.84)	(49.80)	(-6.54)	(89.70)	(15.15)	(112.40)
51+ workers	0.0432***	-0.0169***	0.198***	0.0122***	0.0918***	0.00625***	0.212***
	(19.62)	(-10.64)	(66.22)	(10.36)	(58.02)	(3.34)	(134.81)
Written contract	-0.0884***	0.00702***	-0.0516***	0.0467***	0.0503***	0.0992***	0.0844***
	(-43.03)	(6.41)	(-28.32)	(57.39)	(46.10)	(74.71)	(79.84)
Formal	0.00577***	0.0128***	-0.00382*	0.0477***	0.0857***	0.0371***	0.126***
	(4.11)	(13.92)	(-2.05)	(61.16)	(72.43)	(26.41)	(84.73)
Formal * contract	0.117***	-0.0124***	0.0765***	-0.0422***	-0.0319***	-0.0707***	-0.0143***
	(51.68)	(-10.08)	(35.33)	(-44.97)	(-25.03)	(-45.62)	(-11.02)
Formal * [6-10]	-0.0310***	-0.0610***	0.0550***	-0.0216***	-0.0455***	-0.0256***	0.00335
	(-15.71)	(-46.88)	(24.57)	(-20.30)	(-29.67)	(-14.60)	(1.88)
Formal * [11-50]	-0.0756***	-0.0666***	-0.0674***	0.0530***	-0.0300***	0.0704***	-0.0215***
	(-40.64)	(-54.07)	(-28.29)	(52.39)	(-20.92)	(41.86)	(-12.98)
Formal * [51+]	0.0378***	0.0492***	-0.156***	0.0432***	-0.0158***	0.125***	-0.00584**
	(14.81)	(27.87)	(-45.48)	(32.27)	(-9.17)	(60.57)	(-3.27)
Observations	1,453,795	3,991,436	1,090,511	7,354,974	5,005,826	4,181,318	7,057,757
Adjusted R^2	0.746	0.718	0.755	0.708	0.733	0.746	0.704
Controls	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality

Source: Prepared by the authors.

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 to very small ones [0-5 workers] is small; the difference with being employed in a small [6-10 workers], medium [11-50 workers] or large [50+ workers] one is 7.7, 5.4, and 8.1 percent, respectively (and in the last case, not statistically significant). Contrast this with workers with a university education: relative to very small firms, the difference of being employed by a small, medium, or large firm is 7.3, 13.4, and 20.6 percent, respectively. Similar results hold contrasting workers with completed primary education with those who have completed high school: for the former, the difference between working for a large firm relative to a very small one is 3.2 percent; for the latter it is 13.1. More generally, Table 12 shows that there is a premium for working with larger firms and that this premium increases with educational level (net of interactions between firm size and formality).

Having a written contract also matters for earnings, and again has a different effect depending on educational level. In the case of workers with incomplete primary schooling, the net effect is a 2.8 percent increase in earnings; for those with a university education, it is 7 percent. A similar contrast holds for workers with complete primary versus complete high school: 0 versus 7.3 percent.

Finally, recall from Table 4 that, controlling for size, formal firms have higher productivity than informal ones. Since we cannot identify firm type in the 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 like size and a written contract, other dimensions of firm formality positively affects workers' earnings, increasing them between 0.5 and 12.6 percent.

6.4 Counterfactual Earnings Distributions and Returns to Education

We now combine the results of Tables 9, 10, and 12 to carry out our counter-factual simulation. The broader objective is to identify the impact of misallocation on workers' earnings. 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 where distortions induce firms to separate into formal and informal, each with very different size distributions and with larger average size for formal firms (as in Table 2). He then finds that in

the absence of distortions there would be no informal firms, while the average size of formal firms would increase (and the distribution would be closer to that observed in the United States).

We take advantage of these results and operationalize the no-misallocation scenario 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, we ask: 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? Importantly, 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 it, as our exercise 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, we redistribute workers in informal firms to replicate the proportions across sizes of formal firms in which workers of the same educational level are employed. Consider, for example, workers with incomplete primary education. As seen in Table 9, in the formal sector 44.2 percent are in large firms and 11.8 percent in very small ones, in contrast to 3.9 and 72.1 percent, respectively, in the informal sector. Further, 34.8 percent of all formally employed workers are in medium-size firms and 9.1 percent in small ones, versus 11.9 and 12 percent, respectively, of all those informally employed. To replicate the proportions observed in the formal sector, we randomly subtract 60.3 percent (=72.1 minus 11.8) of informally employed workers from very small firms, and 2.9 percent (=12 minus 9.1) from small firms, and then randomly assign them to medium-size and large firms to reach the desired proportions (34.8 and 44.2 percent, respectively). Note that we only redistribute at the margin, that is, we only remove excess workers from firm sizes that employ proportionately more than formal firms of the same sizes, and place them in firm sizes that employ proportionately less. Note as well that this exercise is performed for each educational level separately.

Second, for each individual worker of a given educational level who is moved, we use the regression coefficients from Table 12 to impute the implied change in earnings derived from

working for a different-size firm. To capture the effects of changing firm type, we impute the value of the dummy for formality to all informal workers, and randomly assign the effect of a written contract to those who did not have one, in order to reach the same proportions as formal workers of that educational level and firm size. All other determinants of earnings are left intact, including workers' unobservable characteristics, 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 randomly adding it to informal workers who did not receive such a bonus to achieve, again 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 13 reports the observed and simulated mean and standard deviation of the distribution of workers' earnings by educational level, and the observed share of formal and informal workers. By construction, columns two and five are the same, while columns three and six show the observed and simulated 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 synthetizes the cost to workers for being employed in the informal rather than in the formal sector or, the misallocation penalty. Critically, *this penalty does not result from workers lacking education or abilities*; rather, it results from distortions that allocate too many resources to low-productivity informal firms. Second, and in line with our hypothesis, note that the penalty is not evenly distributed across educational levels: for those with incomplete primary schooling it is 9 percent, but for those with a university education it is significantly higher, at 29 percent. Similarly, for those with complete primary schooling it is 3 percent, while for those with completed high school it is 17 percent.

Table 13. Observed and Simulated Earnings, 2006.Q2 Anchor
(pesos per hour)

	Observed			Simulated						
	All	**Formal**	**Informal**	**All**	**Formal**	**Informal**				
	(1)	(2)	(3)	(4)	(5)	(6)	(6)/(3)	(4)/(1)	(2)/(3)	(5)/(6)
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 high school										
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 high school										
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: Prepared by the authors. *Note:* SD = standard deviation.

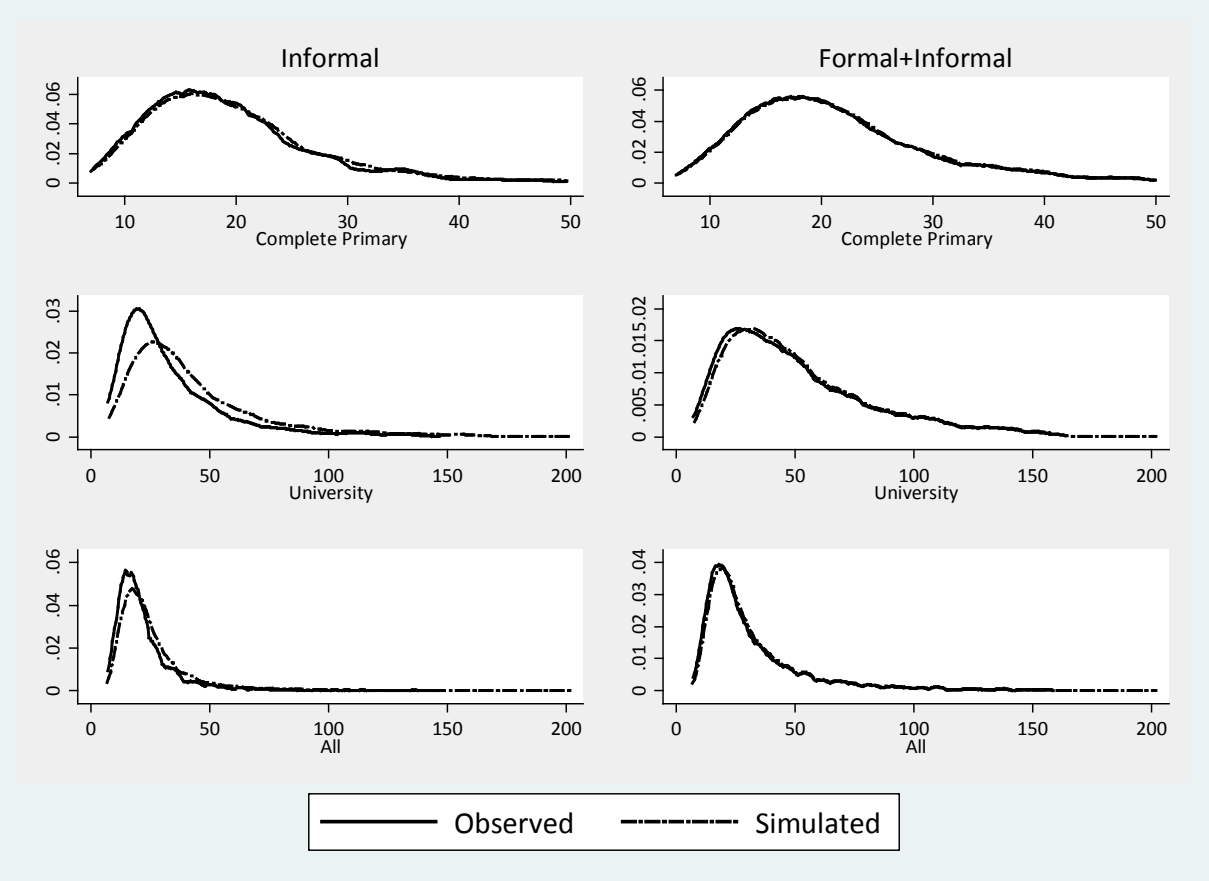
The simulated mean of earnings for all workers by educational level is shown in column four. Since this is simply the weighted average of earnings for newly formal workers and already formal workers, the increase is lower than that of newly formal workers by themselves, as earnings of already formal workers did not change. Overall, the mean increase is 4 percent, and is very similar across educational levels, a result that follows from the fact that the proportion of informally to formally employed workers falls as educational levels increase. Thus, the 9 percent

increase in earnings for informally employed workers with incomplete primary schooling benefits 56.7 percent of all workers with that level of education, while the 29 percent increase in earnings for informally employed workers with a university education applies to only 19.9 percent of those workers. That said, recall that the share of formal workers in our sample is 70 percent, while for the occupied labor force as a whole the share 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.

The last two columns in Table 13 show the formal-informal earnings differentials. The observed one is positive and increases by educational levels. The differential narrows in the simulated scenario, and narrows proportionately more for workers with more education: for those with incomplete primary schooling it falls from 1.21 to 1.11 versus 1.62 to 1.26 for those with a university education (and from 1.58 to 1.35 for the full sample). We interpret the post-simulation earnings differentials to result from a combination of innate differences in underlying workers' abilities, or other differences in firm characteristics that are in principle observable (like unionization or job tenure) but that were not considered in our earnings regressions because they are not available for every year in the ENE-ENOE data.

Figure 4 depicts the observed and simulated 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 can be seen that the mean and the dispersion of earnings increase in the simulated scenario, although the contrast between the two educational levels is notable. These contrasts are reduced when considering the sum of formal and informal employees, in the right panels, because the share of informal employees with complete primary schooling among all employees of that educational level is higher than that of employees with a university education. Still, the mean and the dispersion are larger in the simulated scenario for both groups. This holds as well for all educational groups combined, as can be seen in the bottom graphs in Figure 4 (and verified in Table 13).

Figure 4. Observed and Simulated Earnings Distributions, 2006



Source: Prepared by the authors.

Table 14 presents the estimates of the returns to education in the observed and simulated earnings distribution. (The numbers for the observed distribution differ slightly from those presented in Table 8 because here we also consider the yearly bonus.) The most significant result, of course, is the significant increase in the returns to education for informal workers, which again highlights the point that when workers are employed by informal firms, accumulating more years of education is substantially less valuable than when they are formally employed. That said, returns to education for informal workers are still below those for formal workers, which is just the other side of the coin of our finding that earnings differences with formal workers were narrowed by our counter-factual exercise, but not eliminated.

Table 14. Observed and Simulated Returns to Education, 2006.Q2 Anchor
(relative to incomplete primary)

	Observed			Simulated	
	All	**Formal**	**Informal**	**All**	**Informal, Simulated**
Age	0.0472***	0.0456***	0.0312***	0.0430***	0.0311***
	(32.71)	(25.88)	(13.86)	(30.57)	(13.69)
Age * Age	-0.000445***	-0.000400***	-0.000342***	-0.000401***	-0.000345***
	(-23.16)	(-17.05)	(-11.38)	(-21.39)	(-11.36)
Complete primary	0.0732***	0.0691***	0.0213	0.0350**	-0.0406*
	(5.63)	(3.71)	(1.27)	(2.76)	(-2.40)
Incomplete junior high	0.186***	0.203***	0.0787***	0.196***	0.130***
	(10.79)	(8.48)	(3.45)	(11.63)	(5.64)
Complete junior high	0.219***	0.224***	0.0758***	0.217***	0.146***
	(17.74)	(12.71)	(4.62)	(18.01)	(8.81)
Incomplete high school	0.362***	0.384***	0.0969***	0.349***	0.175***
	(28.16)	(21.38)	(5.26)	(27.80)	(9.40)
Complete high school	0.450***	0.459***	0.194***	0.431***	0.263***
	(34.48)	(25.32)	(10.12)	(33.79)	(13.62)
University	0.863***	0.880***	0.500***	0.859***	0.668***
	(69.82)	(50.49)	(27.65)	(71.16)	(36.56)
Observations	35,059	24,440	10,619	35,059	10,619
Adjusted R^2	0.340	0.362	0.188	0.346	0.249
Controls	Municipality	Municipality	Municipality	Municipality	Municipality

Source: Prepared by the authors.
Note: t statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6.5 General Equilibrium Effects

We have simulated the earnings distribution of informal employees assuming they are employed by formal firms. From their perspective, our simulations ensure 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 perspective of firms, the simulated 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, as documented above.

Table 9 helps to 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 of each educational level, and thus the total number of informal workers. Thus, despite the redistribution, it is still the case that of all newly formalized workers, 9.5 percent have incomplete primary schooling, 19.8 complete primary, and ….3.9 percent a university education. This composition of the workforce differs substantially from that observed in the formal sector, where only 3.3 percent of workers have incomplete primary schooling, 11 percent complete primary, and …27.1 percent a 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 a higher education, or too many workers with low educational levels. From whichever perspective, at the simulated earnings distribution there would be a mismatch between firms' demand and workers' supply 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 be the total number of formal and informal workers. From the perspective of firms the issue is that:

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

It is of interest to ask what adjustment would be required to the supply of informally employed workers of each schooling category to match the proportions observed in the formal sector. This can be done taking as given α_e and N_{ie}, and finding the values of δ_e that solve:

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

When expanded, equation (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.[21] To obtain a non-trivial solution, we arbitrarily set $\delta_7 = 1$, implying that we measure any excess supply or demand of workers of educational level *e* relative to workers with a 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 that observed in the formal sector (or increased if the inequality is reversed). Differently put, when $\delta_{e\neq 7} \leq 1$ there is excess supply of workers of that educational level, with $(1-\delta_e)N_{ie}$ measuring the absolute size of the excess supply, and $(1-\delta_e)N_{ie} / (N_{ie} + N_{fe})$ measuring the excess supply relative to the total number of workers of that educational level.

The second column of Table 15 presents the solution to equation (3). Two results are of interest. 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. Second, and more importantly, in relative terms excess supply is largest for workers with the least amount of education.[22]

Table 15. 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	227,171	186,260	413,389	0.451
Complete primary	0.288	469,525	334,510	1,080,167	0.310
Incomplete junior high	0.243	137,191	103,819	289,111	0.359
Complete junior high	0.424	673,273	387,796	1,968,788	0.197
Incomplete high school	0.723	289,234	80,233	1,235,346	0.064
Complete high school	0.776	243,601	54,632	1,102,977	0.049
University	1.000	328,109	0	1,834,499	0
Total	.	2,368,104	1,147,250	7,924,277	0.144

Source: Prepared by the authors.

[21] 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,…,6.

[22] 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, 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 informal firms.

These results can be interpreted as follows: the simulated earnings distribution presented in column four of Table 13 captures the effects that, from the perspective of workers, would be observed if they were all employed by formal firms. On the other hand, Table 15 indicates that formal firms would not be willing to employ all of them at the simulated distribution. More precisely, firms would be least willing to employ those workers with the fewest years of schooling: 45 percent of workers with incomplete primary would be in excess supply, 31 percent of those with completed primary would be in excess supply, 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 employment in the magnitudes shown in the last column of Table 15 (through an exogenous drop in participation rates, for example). This would imply that the simulated earnings distribution is the equilibrium distribution. On the other extreme, the adjustment would occur through prices, with earnings of workers of all educational levels falling as necessary, relative to those with a university education, to clear excess supply.

Our approach does not allow us to quantify the changes in earnings and employment levels required to establish 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, earnings of workers with fewer years of education would fall relative to those with more years of education. In turn, this implies that the simulated earnings and returns to education shown in Tables 13 and 14 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 so far can be seen as a lower bound of the effects of eliminating misallocation and associated firm informality.

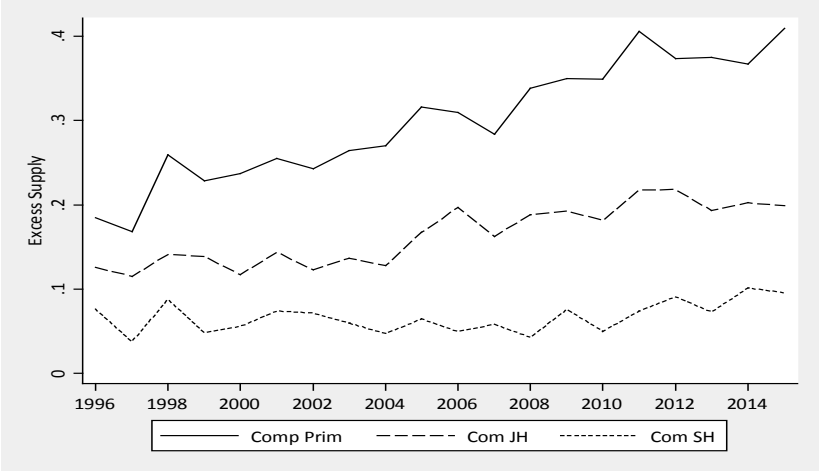
7. The Wage Premium, 1996-2015

We now use the results of the previous sections to shed some light on the evolution of the wage premium between 1996 and 2015. To set the stage, it is useful to contrast the evolution of the size and type distribution of firms, obtained from the census and presented in Table 5, with the evolution of the supply of workers by schooling levels, obtained from the employment survey as shown in Table 6. 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 significantly, as the number of educated persons and workers increased faster than the average of all persons/workers (and indeed the number of those with the least education fell in absolute terms). In parallel, however, the degree of misallocation of resources persisted (actually, increased slightly) and always in the direction of firms that are less intensive in educated workers. So the broad picture that emerges from Tables 5 and 6 is that of an economy where, on one hand, the supply of educated workers is increasing relatively fast; but where, on the other hand, given persistent misallocation, firm and employment growth is faster in the informal sector.

To illustrate the effects of these trends, we solve system (3) for each year between 1996 and 2015. Figure 5 plots the percentage excess supply of workers of educational categories with complete school cycles as shown in the last column of Table 15 (results for incomplete school cycles are between 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, excess supply of workers with fewer years of education would have increased overtime, particularly for those with complete primary education and, to a lesser extent, with complete junior high. In the first case, excess supply more than doubles in the 20-year span, from 18.5 to 41 percent; for the second case it increases from 12.6 to 19.8 percent.

Figure 5. Excess Supply of Workers Relative to Those with University Education, 1996-2015



Source: Prepared by the authors.

As mentioned, a richer model is required to measure the changes in earnings across all educational categories needed to clear the excess supplies shown in Figure 5. In the absence of that, we carry out a simple exercise focusing only on the ratio of the mean of earnings of different schooling levels (i.e., the wage premium), and considering only two educational levels. The assumption is that there are no changes in participation rates, so 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:

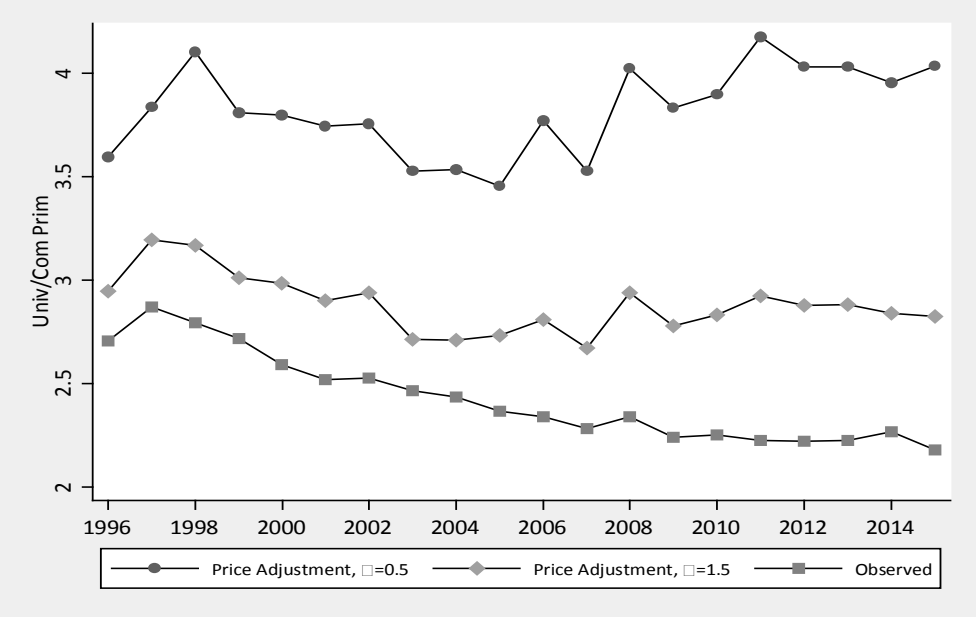
$$(4)\ \ \partial \ln(w^H / w^L) = -(1/\sigma)\partial \ln(H / L),$$

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

Figure 6 shows three paths of the wage premium between workers with a university education and those with complete primary schooling. The observed path is declining, in line with the earnings trends described in Figure 1. The paths labeled "price adjustment" use equation (4) and the estimates of excess supply of workers with primary education for each year depicted in Figure 5, to correct the simulated wage premium obtained from Table 13 for two assumed values of the elasticity of substitution, 1.5 and 0.5.[23]

[23] 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, however, correspond to comparisons between workers with a university versus high school education. Benita (2014) provides estimates for Mexico on the order of 3, but again between workers with a university versus primary education. Importantly, note that Benita's estimates implicitly reflect the formal-informal firm composition observed in Mexico. In our case, first, we are comparing workers with a university versus complete primary education; and second, we are considering a scenario where there are only formal firms. For these two reasons, the elasticity should be considerably lower. We thus arbitrarily chose two values, 1.5 and 0.5, which are respectively 50 percent higher and lower than the elasticity of substitution associated with a Cobb-Douglas production function.

Figure 6. Earning Premiums: University versus Completed Primary, 1996-2015



Source: Prepared by the authors.

We underline two results. First, in the simulated scenarios 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 simulated in Section VI, which did 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 overtime, as opposed to the decreasing trend of the observed premium; in the low-value case the trend is in fact increasing overtime.

Our analysis is limited for three reasons. First, we only considered two educational levels, while if firm informality disappeared, the earnings of workers of all educational levels would be affected. Second, we have, so-to-speak, stitched together the static results of each year to simulate the path of the wage premium, but in the counter-factual scenarios there are no inter-temporal effects from the equilibrium of one year on the next. Third, we have 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 as suggestive. Our objective is not to pin-point with precision the informality-free path of the wage premium, but rather to illustrate that the continuous growth of employment in informal firms as a result of persistent misallocation while the schooling of the labor force increases goes a long way in accounting for the observed decline of the wage premium.

8. Conclusions

This paper has presented a preliminary assessment of the effects of misallocation on labor earnings and the returns to education in Mexico. Misallocation was 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 (i.e., informal firms). We showed that because the production process in these firms is less intensive in educated workers than that of formal firms, misallocation tilts the demand for labor in favor of workers with fewer years of schooling. In doing so, it reduces the mean and variance of the distribution of labor earnings, and lowers the returns to education. We showed this by simulating the earnings distribution that would be observed in a counterfactual scenario where the schooling composition of the demand for labor from all firms was the same as that of existing formal firms. Our basic finding is that misallocation lowers average earnings of informally employed workers by 17 percent, but with substantial differences across educational levels: for workers with incomplete primary schooling, earnings fall by 9 percent, while for those with a university education earnings fall by 29 percent. Differently put, misallocation is more costly to workers with more years of schooling.

Our analysis was limited in scope, as we considered only the direct effects of firm informality on employees' earnings, although we showed that if general equilibrium effects were considered our results would be strengthened. That said, our analysis fell 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 that are employees, those that are self-employed, and those that are entrepreneurs. Further, firm-worker relations would change, with longer job tenures and greater

investments in worker training.[24] Firm dynamics would also differ.[25] None of these effects are captured in this paper, though they would quite likely have first-order effects on earnings.

Despite these limitations, we claim to have established that misallocation is a central part of the explanation as to 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 failing to find jobs that fully value their additional education 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. And this story has a corollary: the distortions that misallocate resources are more costly today than 20 years ago because there are more educated workers now than before. And, if the distortions persist while the schooling of Mexican workers continues to improve, they will become even costlier in the years ahead.

We conclude with three observations. First, our approach may help understand the decline in the returns to education observed in other Latin American countries Mexico (Gasparini et al., 2011). 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. Our 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—and concentrate on misallocation. Are there large distortions? Are they de facto taxing firms with some characteristics and subsidizing other firms? Are there systematic differences in the schooling composition of taxed and subsidized firms? Although not all these countries have sufficient firm-level data to carry out computations like the ones presented in Section 3 of this paper, the ones presented in Sections 5 and 6—which rely on employment surveys—may still shed some light.

The second 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

[24] Alaimo et al. (2015) argue that excessive worker rotation is associated with lower investments by firms in worker training and lower returns to experience. Further, 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 receive such training have higher education and work mainly in formal firms.

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

the main causes behind the fall in the Gini coefficient from 0.55 to 0.50 between 2002 and 2010. Reduced inequality is in principle 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. By shifting resources in favor of informal firms that are more intensive in less-educated workers, misallocation in Mexico reduces earnings dispersion. But because it also reduces productivity, it lowers workers' living standards—particularly for more educated workers. It is far from clear that this source of reduced inequality can be seen in a positive light.

Finally, our 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, our paper indicates that lack of educated workers is *not* one of them. The notions that "investments in human capital will gradually eliminate informality", or that "investments in human capital will gradually eliminate misallocation" are, in our view, flawed.[26] Rather, our paper suggests the opposite: the persistence of misallocation is impeding Mexico from taking full advantage of its investments in education.

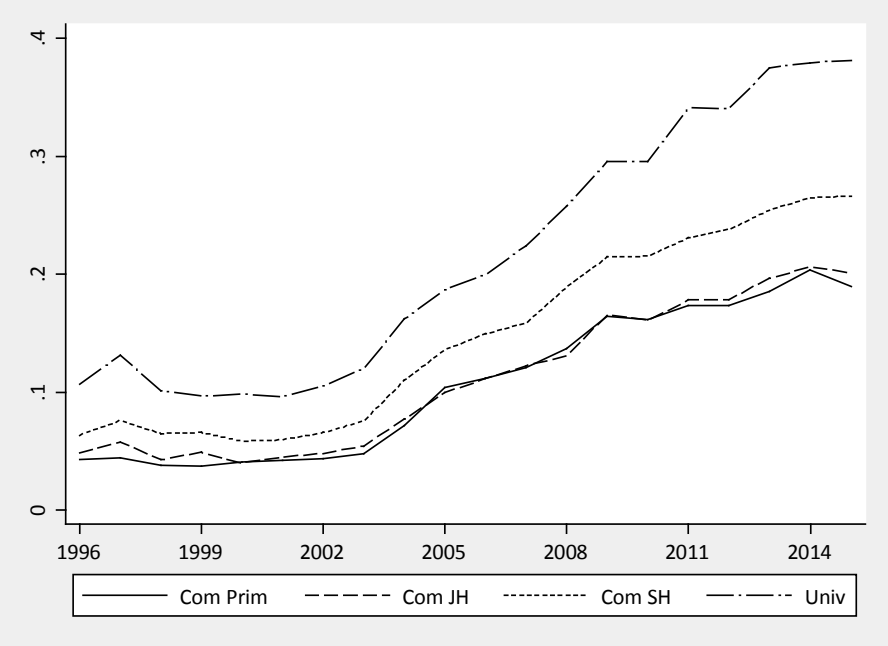
[26] More broadly, our results suggest that there is no automatic connection between increased schooling and higher productivity and earnings; the relation is mediated by the degree of misallocation in the economy.

Appendix: The Problem of Missing Observations

1. Statistical Approach

A common 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 A1 highlights two facts in the case of the ENE-ENOE: first, under-reporting rates increase with the level of education; and second, under-reporting has increased for all educational levels.

Figure A1. Rates of Under-reporting of Earnings, ENE-ENOE, 1996-2014
(Percent)



Source: Prepared by the authors.

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 who do and do not report earnings. These variables are gender, the seven educational levels, the individuals' age quartile (computed over the economically active population), 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 who do and do not report earnings.[27]

The results of the imputation are displayed in Table A1 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 A1. Employee Earnings, 2006

	Pre-Hot Deck			
Educational Level	**Number**	**Mean**	**Standard Deviation**	**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 high school	1,081,256	28.75	19.15	23.64
Complete high school	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 high school	1,244,486	28.61	19.22	22.95
Complete high school	1,113,035	30.57	20.19	25.37
University	1,916,397	55.20	44.16	43.60
Total	**8,065,660**	**32.33**	**28.77**	**23.55**

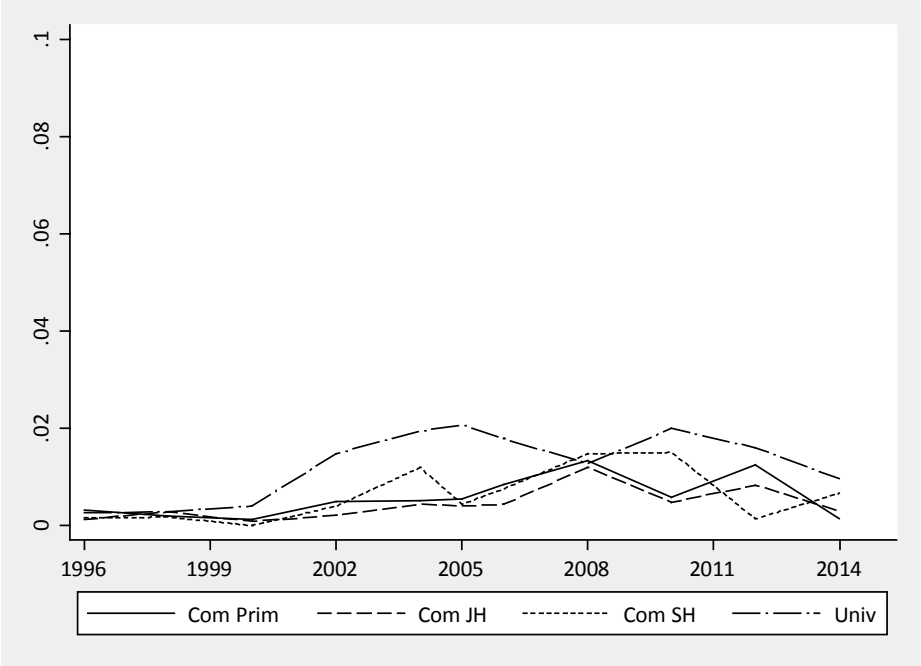
Source: Prepared by the authors.

[27] Further details can be found in Andridge and Little (2010).

2. Corroboration with Data from Mexico's Household Surveys for 1994-2012

As mentioned earlier, the hot deck technique implicitly assumes that, within each group, individuals who do not report earnings are a random sub-sample of all individuals in that group. Since, by definition, we have no information on the earnings of those who do not report, we cannot test for this assumption. Nevertheless, to corroborate our results we also measure the rates of under-reporting 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 main text, we exclude the self-employed and government employees and focus on private sector employees between 18 and 65 years of age who live in towns with over 100,000 inhabitants and work more than 35 hours a week. Figure A2 shows the rates of under-reporting in the ENIGH.

Figure A2. Rates of Under-reporting of Earnings in ENIGH, 1996-2014
(Percent)

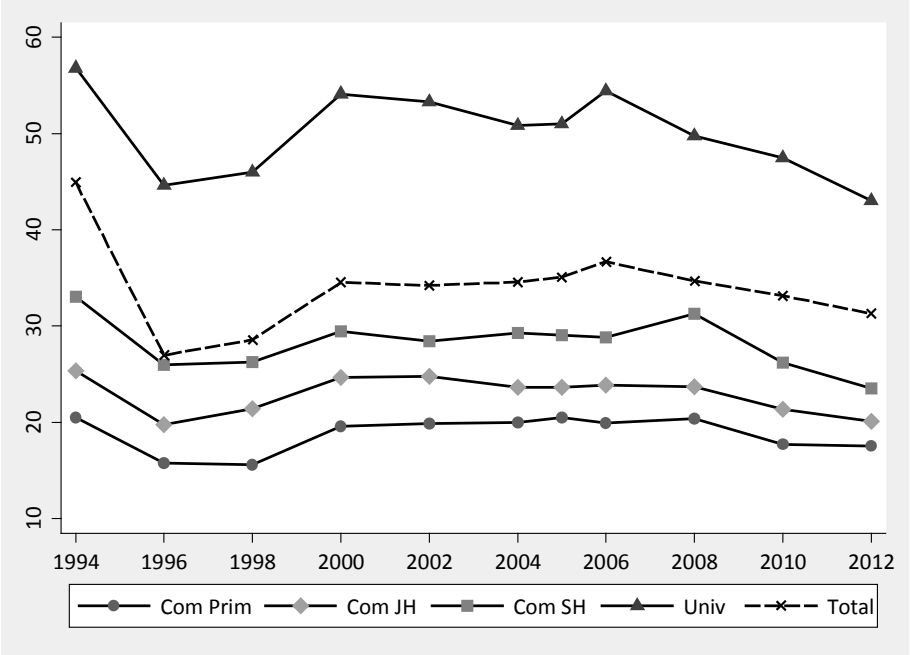


Source: Prepared by the authors.

Contrasting Figures A1 and A2, it is clear that: i) rates of under-reporting are substantially lower in ENIGH compared to ENE-ENOE, ii) differences in under-reporting across educational levels are very small, and iii) rates of under-reporting increased slightly over time, but then declined.

Figure A3 reproduces the same results presented in Figure 1 in the main text using 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 A3 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 up-ward trend in earnings from 1996 to the early 2000s is basically a recovery from that sharp fall.

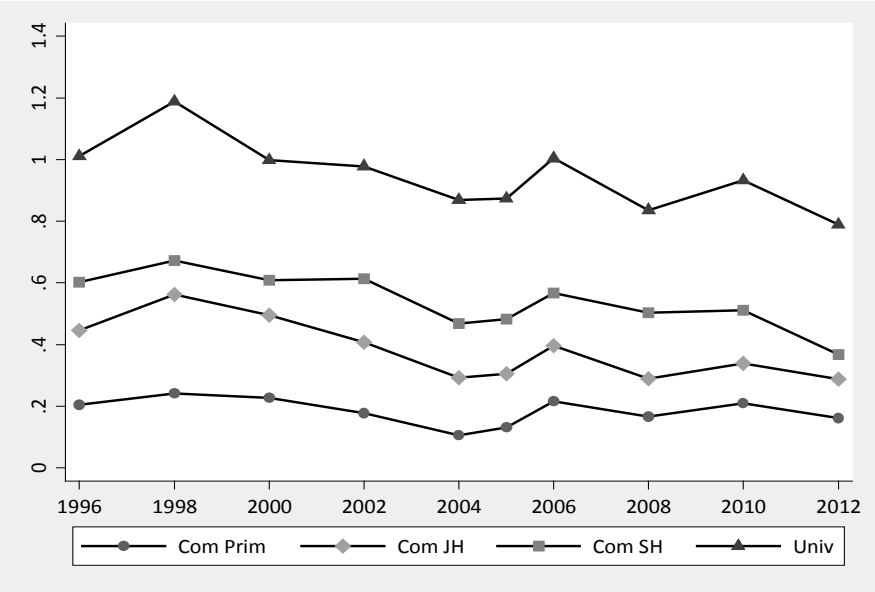
Figure A3. Earnings by Educational Level, ENIGH, 1994-2012
(Real hourly wages, November 2014 pesos)



Source: Prepared by the authors, based on ENIGH data.
Note: Data exclude self-employed workers.

Finally, Figure A4 presents estimates of the returns to education using ENIGH data, focusing on the same sample described in the main text. Again, results are similar to those presented in Figure 3 in the main text using ENE-ENOE data after applying the hot deck technique.

Figure A4. Evolution of Returns to Schooling by Educational Level, ENIGH, 1996–2012
(Percent, relative to incomplete primary)



Source: Prepared by the authors.

References

Acemoglu, D., and D. Autor. 2011. Skills, Tasks and Technologies: Implications for Employment and Earnings. In *Handbook of Labor Economics,* Volume 4, Part B, edited by O. Ashenfelter and D. Card. Elsevier.

Alaimo, V., M. Bosch, D. Kaplan, C. Pagés, 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.

Anton, A., F. Hernandez, and S. Levy. 2012. *The End of Informality in Mexico? Fiscal Reform for Universal Social Insurance.* Washington, DC: Inter-American Development Bank.

Autor, D., L. Katz, and L. Kearney. 2008. Trends in US Inequality: Revising the Revisionists. *The Review of Economics and Statistics* 90(2): 300–23.

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. Fabbi, and S. Levy. 2016. Labor Market Frictions, Search and Schooling Investments. Manuscript, Inter-American Development Bank, Washington, DC.

Bouillon, C. 2002. Inequality and Mexico's Labor Market after Trade Reform. IDB Working Paper 53718. Inter-American Development Bank, Washington, DC.

Busso, M., and S. Levy. 2016. Informality, Productivity and Misallocation, Mexico 1998-2013. Manuscript under preparation. Inter-American Development Bank, Washington, DC.

Busso, M., M.V. Fazio, and S. Levy. 2012. (In)Formal and (Un)Productive: The Productivity Costs of Excessive Informality in Mexico. IDB Discussion Paper No. 341. Inter-American Development Bank, Washington, DC.

Busso, M., L. Madrigal, and C. Pagés. 2013. Productivity and Resource Misallocation in Latin America. *B.E. Journal of Macroeconomics* 13(1): 903–32.

Busso, M., A. Neumayer, and M. Spector. 2015. Skills, Income Distribution, and the Size Distribution of Firms. Department of Economics, Universidad Torcuato di Tella, Argentina. Unpublished.

Campos, R., G. Esquivel, and N. Lustig. 2012. The Rise and Fall of Income Inequality in Mexico. WIDER Working Paper 2012-10. World Institute for Development Economics Research, Helsinki.

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

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

Esquivel, G., and J.A. Rodríguez-López. 2003. Technology, Trade and Wage Inequality. *Journal of Development Economics* 72(2): 543–65.

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. Centro de Estudios Distributivos, Laborales y Sociales Documento de Trabajo 127. Universidad Nacional de la Plata.

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 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): 1403–448.

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

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

Instituto Nacional de Estadística, Geografía e Informática (INEGI). 2014. La Informalidad Laboral, Encuesta Nacional de Ocupación y Empleo, Marco Conceptual y Metodológico. Distrito Federal, México.

Lächler, L. 1998. Education and Earnings Inequality in Mexico. Policy Research Working Paper No. 1949. World Bank, Washington, DC.

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.

Levy, S., and M. Szekely. 2015. Más Escolaridad, Menos Informalidad? Un Análisis de Cohortes para México y América Latina. Inter-American Development Bank, Washington, DC. Unpublished.

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

Washington, DC: Brookings Institution Press.

López-Calva, L.F., and E. Ortiz-Juarez. 2014. A Vulnerability Approach to the Definition of the Middle Class. *Journal of Economic Inequality* 12(1): 23–47.

López-Martin, B. 2015. Informal Sector Misallocation. Research Department, Banco de Mexico.

Lustig, N., L.F. López-Calva, and E. Ortiz-Juarez. 2014. Deconstructing the Reduction in Wage Inequality in Latin America. In *Shared Prosperity and Growth*, edited by K. Basu and J. Stiglitz. Oxford, UK: Oxford University Press.

Restuccia, D., and R. Rogerson. 2008. Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics* 11(4): 702–20.

Robertson, R. 2007. Relative Prices and Wage Inequality: Evidence from Mexico. *Journal of International Economics*, 64(2): 387–409.

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

Syverson, C. 2011. What Determines Productivity? *Journal of Economic Literature* 49:2: 327–62.

Torres-Coronado, J. 2015. Size-dependent Firm Regulations in a Many-to-one Assignment Model: Talent Misallocation and the Return to Skill. Instituto Tecnologico de Monterrey, Mexico. Unpublished.