CEDLAS
Facultad de Ciencias Económicas
UNIVERSIDAD NACIONAL DE LA PLATA

DOCUMENTOS DE TRABAJO

The Labor Market Effects of an Educational Expansion. A Theoretical Model with Applications to Brazil

David Jaume

Documento de Trabajo Nro. 220
Enero, 2017
ISSN 1853-0168

www.cedlas.econo.unlp.edu.ar

The Labor Market Effects of an Educational Expansion. A Theoretical Model with Applications to Brazil*

(Job Market Paper)
Click here for latest version

David Jaume†

December 3rd, 2017

Abstract

Most countries are rapidly increasing the educational attainment of their workforce. This paper develops a novel framework to study, theoretically and empirically, the effects of an educational expansion on the occupational structure of employment and distinct aspects of the wage distribution—wage levels, wage gaps, and poverty and inequality indicators—with an application to Brazil. I proceed in three steps. First, I provide some stylized facts of the Brazilian economy between 1995 and 2014: A large educational expansion took place; the occupational structure of employment remained surprisingly fixed; workers of all educational groups—primary or less, secondary, and university—were increasingly employed in occupations of lower ranking as measured by average wages over the period; and wages of primary educated workers increased while wages of more educated workers declined, bringing forth reductions in poverty and inequality. Second, I build a model that traces these heterogeneous patterns of occupations and wages to the educational expansion. The model assigns workers with three levels of education to a continuum of occupations that vary in complexity and are combined to produce a final good. I investigate three different policy experiments: An increase in university level, an increase in secondary level, and a simultaneous increase in both. The predicted effects depend on the policy analyzed. Considering the educational expansion that took place in Brazil (simultaneous increases in university and secondary levels), the model predicts qualitatively all the observed labor market changes in the occupational structure of employment and the wage distribution. Finally, I calibrate the model with the data from 1995 and show that, through its lens, the observed educational expansion in Brazil explains 66 percent of the occupational downgrading and around 80 percent of the changes in wage levels, inequality, and poverty during the period of 1995-2014.

JEL-codes: I25; J24; O15.
Keywords: Educational Expansion, Brazil, Labor Market, Wages, Occupations, Poverty, Inequality.

---

*I am very grateful for the support, guidance, and invaluable discussions with Gary Fields and Julieta Caunedo. I thank Arnab Basu, Kaushik Basu, Francine Blau, Nancy Chau, Guillermo Cruces, Maria Fitzpatrick, Jorgen Harris, Lawrence Khan, Ravi Kanbur, Elissa Keller, Michael Paris, Victoria Prowse, Evan Riehl, Nicholas Sanders, Mallika Thomas, Erik Thorbecke, Mariana Viollaz, Alexander Willén, as well as conference participants at the 2017 North East Universities Development Conference (NEUDC), the 2017 Society for the Study of Economic Inequality (ECINEQ) annual meeting, the 2017 Research Institute for Development, Growth and Economics (RIDGE) meeting, the 2017 North American Summer Meeting (NASMES) and Cornell internal seminars for valuable comments and suggestions on earlier drafts of this paper.

†Ph.D. Candidate in Economics, Department of Economics, Cornell University. Email: djj56@cornell.edu.

1 Introduction

Increases in human capital through formal education are widely cited as a key driver of economic development (Hanushek & Woessmann 2008, 2012). In the last decades, most countries have experienced a rapid increase in the educational attainment of their workforce. Between 1990 and 2010 the worldwide average enrollment rates in secondary and university education grew by around 25 percentage points each (Barro & Lee 2013). An educational expansion, defined as an increase in the share of workers with secondary and/or university education, raises the educational level of some workers while leaving that of others unchanged. Those workers that receive the additional education increase their human capital, earn higher wages, and have access to better jobs. But an educational expansion likely affects the rest of the workforce as well. An influx of more educated workers can result in labor market effects that reshape the occupational composition of employment (what jobs are available and who performs them) and the entire wage distribution (the wages for those jobs). These effects are commonly overlooked when designing an educational policy, and they are by no means obvious. For example, the workers who remain with low levels of education may be hurt if they are displaced to worse jobs, where wages remain stagnant or decline, or they may become better off if their wages increase as result of a higher demand for their jobs (generated by general equilibrium effects) coupled with a lower supply.

This paper provides a novel framework to study the effects of an educational expansion on the occupational structure of employment and distinct aspects of the wage distribution, both theoretically and empirically, with an application to Brazil. I proceed in three steps. First, I provide some stylized facts for Brazil on the inter-linkages between changes in education, occupations, and wages over the period of 1995-2014. Second, I build a theoretical framework that relates changes in occupations and wages to an educational expansion that is qualitatively consistent with the patterns observed in Brazil. Finally, I calibrate the model and assess if the predicted effects of the Brazilian educational expansion are quantitatively relevant to explain its labor market changes during the analyzed period. This theoretical and empirical framework is of particular relevance to study other developing countries where educational attainment has increased or is increasing very rapidly.

In the first part of my paper, I rely on Brazilian data from the Pesquisa Nacional por Amostra de Domicílios (PNAD) during the period of 1995-2014. An exceptionally large educational expansion took place during these years, which is reported to be one of the fastest expansions in history (Bruns et al. 2011). The share of workers with secondary education doubled from 20.5 to 40.0 percent, the share with university level grew from 11.3 to 23.6 percent, and the share of workers with only primary education or less halved from 68.1 to 36.4 percent. During this period, the occupational structure of employment of the entire economy remained surprisingly fixed, while there was lower occupational attainment within each of the educational group—primary or less, secondary, and university—defined as an increase in employment in occupations of lower ranking, this was especially pronounced for workers with secondary schooling.[1] Despite this

[1] I rank ISCO-88 at 3-digit occupations between 0 and 1 according to the median wages over the entire period,

common decline in occupational attainment across educational groups, changes in wages were heterogenous: Wages of primary educated workers increased (37.3 percent), wages of secondary educated slightly declined (10.5 percent), and wages of university educated largely fell (21.6 percent), bringing forth reductions in poverty and inequality.[2] These facts are puzzling: a much larger increase in employment of higher rank occupations is expected if labor markets are fragmented (workers with different educational levels participate in separated labor markets containing particular occupations); average wages are expected to fall for each group and fall more for medium educated workers, if there is lower occupational attainment for all groups and wages are tied to occupations.

For the theoretical model that I construct in the second part of my paper to explain these observed patterns, I build upon the *Ricardian* model proposed by Acemoglu & Autor (2011). There is a unique good to be produced using an infinite number of occupations, which can be ordered by their level of complexity. There are three types of workers: low, medium, and high educated, which relate to the primary, secondary, and university educated workers in the data. Workers' types differ in their productivity of performing each occupation, so that there is an optimal assignment of workers' types to occupations. The key assumption of the model is that the comparative advantage of more educated workers relative to less educated workers increases with the complexity of the occupation. This assumption ensures positive assortative matching, where more educated workers are optimally assigned to more complex occupations.[3] The competitive equilibrium determines the following: 1) the occupational structure of employment (the share of workers performing each occupation), 2) the assignment of workers' types to occupations, and 3) the wages of each type of worker. I analyze the effects on these equilibrium outcomes under three alternative policy experiments: (*i*) an increase in the share of high educated workers through a decline in the share of low educated workers, (*ii*) an increase in the share of medium educated workers through a reduction in the share of low educated workers, and (*iii*) an increase in the shares of both high and medium educated workers. I evaluate for each case the changes in the occupational structure of employment, overall and for each educational group, and various aspects of the wage distribution: average wages for each type of worker, growth incidence curves, poverty and inequality.[4] I find that the predicted effects depend on whether the increase took place in high or medium education. When there is an increase in both, one of these effects will dominate depending on the extent of the changes in the relative supply.

The model qualitatively predicts all the observed labor market changes in the occupational

with 0 representing the occupation with the lowest wage and 1 representing the one with the highest wage. The average ranking diminished by 15.2 percent, 28.2 percent and 8.5 percent for primary, secondary and university educated workers, respectively, between 1995 and 2014.

[2]Reductions in inequality and poverty depend on the index being considered. For example, the Gini index declined by 0.085 points and the wage poverty rate (using a \$2.4 wage poverty line) diminished by 16.4 percentage points.

[3]Positive assortative matching is another stylized fact in Brazil. In a given year workers with higher levels of education are more likely to be employed in occupations of higher ranking.

[4]Growth incidence curves trace the anonymous winners and the losers for each percentile of the wage distribution. Poverty is measured using a wage poverty line between initial wages of low and medium educated. Inequality is approximated by the wage gaps between the different groups and other indexes of inequality.

structure of employment and the wage distribution for the policy experiment that resembles the educational expansion that took place in Brazil. The case of an increase in higher education provides a better intuition to understand the effects of the Brazilian educational expansion (a simultaneous increase in medium and high education). When the share of workers with high education increases, they become more abundant and their wages fall in the occupations that they were originally employed in, while the supply of low educated workers declined and their wages increase in the occupations where they were originally employed. Therefore, it becomes profitable for firms to start hiring high educated workers in occupations that were previously performed by medium educated workers and use medium educated workers in occupations previously performed by low educated workers, generating lower occupational attainment for each educational group. The changes in the occupational composition of employment of the whole economy are smoothed by more educated workers starting to perform occupations of lower complexity. Wages for high educated workers decline due to the increase in supply. Wages of low educated workers rise because of the decline in supply coupled with a higher demand for the occupations that they perform, given that the production of the rest of the occupations increased.[5] Changes in wages of medium educated workers are always between those of low and high educated workers and these changes could be positive or negative (the negative effect of the lower occupational attainment is offset by an increase in the value of the new occupations that they perform). Inequality declines, as measured by wages gaps of more educated workers with respect to lower educated workers. Poverty diminishes as a larger share of the population has access to wages of more educated groups and wages of the remaining low educated workers increase. These patterns exactly match the stylized facts from Brazil.

In the third part of my paper, I examine how much of the observed changes in the data in Brazil can be explained by its educational expansion through the lens of the model. To that end, I calibrate the model using the data from 1995 as the baseline year. The calibration procedure is simple. The set of parameters that need to be estimated are the skill supplies and the productivity across occupations for each type of worker. The parameters corresponding to the educational level of the workforce come directly from the data. I use the average wages for low, medium and high educated workers in 1995 to estimate the technology parameters according to the equilibrium conditions in the model by making the identifying assumption that the relative comparative advantages across occupations are linear functions. Then, I estimate the effects on the equilibrium outcomes of the Brazilian educational expansion between 1995 and 2014. I simulate the effects of an increase of 19.5 and 12.3 percentage points in the share of workers with secondary and university level, respectively, holding fixed the rest of the parameters in the model. Finally, I compare the equilibrium outcomes in the model with that observed in the data before and after the educational expansion. I find that the educational expansion generated most of the quantitative changes observed in the data during the period of 1995-2014. It predicts 38 percent of the (small) changes in the occupational structure of employment, 72 percent of

[5]Although the productivity to carry out each occupation does not change with an educational expansion for any group, there is an implicit complementarity in the production of the final good by combining occupations in a Cobb-Douglas production function with an elasticity of substitution equal to one.

the lower occupational attainment for each educational group, 91 percent of the changes in wage levels of different educational groups and their corresponding wage gaps, 85 percent of the decline in wage inequality, and 68 percent of the reduction in wage poverty. I conclude that the increase in education was of utmost importance to the changes in the Brazilian labor market in the last two decades.

By using the calibrated model to run counterfactuals, I show that the effect of increases in education on average wages and total output declines rapidly with successive educational expansions when technology is fixed. I find that a further educational expansion over a much more educated workforce in 2014 will have less than a third of the effects it had in 1995 when the workforce was less educated. The reason for this is that more educated workers are increasingly employed in tasks of lower complexity, where their relative productivity diminishes, reducing the additional impact of the increase in education on the total output of the economy and, therefore, on average wages. From this exercise, I conclude that further educational expansions in Brazil will have a lower impact, and this is especially true for increases in medium education.

This paper contributes to the existing literature in several important ways. First, it relates to the literature that evaluates the labor market effects of an increase in education. To the best of my knowledge, this is the first paper to study the theoretical predictions of the effects of different educational expansions on the occupational structure of employment and the distinct aspects of the wage distribution in a general equilibrium framework accompanied by a quantitative assessment of these effects. Other papers that use a similar theoretical framework analyze a different set of outcomes, do not bring the model to the data, and are mainly concentrated on the effect of a technological change as opposed to an educational expansion (Teulings 2005, Costinot & Vogel 2010, Acemoglu & Autor 2011). Other models study the effect of an educational expansion in a subset of the outcomes analyzed here: wage premiums among workers with different educational level (Katz & Murphy 1992, Goldin & Katz 2009), the occupational composition of employment (Fields 1974, 1995), or the wage levels (Becker 1994, Moretti 2004*a*, Khanna 2015). Second, this paper provides new evidence for the increase in education to be the main factor behind the changes in the wage distribution in Brazil between 1995 and 2014. By using a novel framework, this paper supports previous research finding that the increase in educational attainment declined wage inequality (Barros et al. 2010, Gasparini et al. 2011, López-Calva et al. 2016, Alvarez et al. 2017), extending the analysis to other aspects of the wage distribution and the occupational composition of employment. Finally, this paper also contributes to the literature by reporting on distinctive patterns of changes in the occupational structure of employment that may arise in developing countries when education increases, contrasting with the job polarization patterns that have been found in developed countries due to skill-biased technological change or to the falling cost of automating routine job tasks (Katz et al. 2006, Acemoglu & Autor 2011, Autor & Dorn 2013, Goos et al. 2014, Beaudry et al. 2016, Burstein et al. 2016, Deming 2017).

In summary, this paper provides a theoretical framework that relates the observed changes in the occupational structure of employment and in the wage distribution in Brazil to its educational expansion, developing a relatively simple method to quantitatively assess these effects. The framework is of particular relevance for developing countries where educational attainment of the workforce has increased or is increasing rapidly.

The rest of this paper is organized as follows: Section 2 discusses the related literature; Section 3 describes the data; Section 4 contains the stylized facts for Brazil; Section 5 introduces the model and the predictions of the model for different policy experiments; Section 6 calibrates the model with Brazilian data; Section 7 presents the results from the calibrated model; and Section 8 concludes.

2 Related literature

The model that I use in this paper is closely related to Acemoglu & Autor (2011). I differ from them by shifting the focus of the analysis from the effects of technological changes towards the labor market effects of different policy experiments of educational expansions. To that end, I incorporate three dimensions. First, I modify the model to be more suitable to study educational expansions by adding the restriction of a measure of 1 of workers in the economy. Thus, if the number of workers in one group increases there must be a corresponding decline in the other groups.[6] Second, I look at changes in a broader set of outcomes related to the entire wage distribution by studying real wages in addition to relative wages, which I use to evaluate the effects on different percentiles of the wage distribution and on wage poverty. Third, I provide an empirical strategy to calibrate the model and quantitatively assess the importance of changes in education on the observed patterns in the data. Teulings (2005) and Costinot & Vogel (2010) also use a similar assignment model to look at changes in the skill distribution of the labor force on the wage distribution and the occupational composition of employment, but do not refer to increases in education.[7]

Most of the existing literature on the labor market effects of an increase in education is concentrated either in its effect on the wage distribution or in the effects on occupations, but not both. A descriptive analysis of this literature can be found in Fields (1995), who presents the labor market consequences of an educational expansion under different functioning of the labor market, distinguishing between two models. First, a model with fragmented labor markets by workers' educational level, where wages fluctuate when supply or demand changes but the occupations that workers with different educational level perform is fixed (the standard textbook model). Second, a segmented labor market where wages in the high earning sector are fixed (and therefore a given number of positions available for each job) and education is a way to be hired preferentially for the best jobs available.[8] In my model, wages and occupations (overall

[6] Acemoglu & Autor (2011) construct the model to study job polarization in the United States arising from technological changes. The restriction of a measure of workers is irrelevant in that context.

[7] These models include a continuum of workers skills which is not explicitly related to education. This makes the effects of changes in the distribution of skills difficult to interpret in terms of an educational expansion, and are harder to bring to the data.

[8] To consider fix wages in one sector of the economy is a common assumption in the literature about segmented labor markets or dual economies in development economics starting with Lewis (1954) and Harris & Todaro (1970). Fields (1974) develops a related model assuming that educated workers are favored by employers and

and for each educational group) are interlinked and simultaneously determined in equilibrium.

Another strand of the literature studies the effect of increases in education on wage premiums and inequality. It is a well known result that an increase in the relative supply of skills causes a decline in wage premiums if the technology is constant (Katz & Murphy 1992, Goldin & Katz 2009, Gasparini et al. 2011, Acemoglu & Autor 2011). This paper adds to the analysis of wage gaps by including the effects on real wages for each educational group, for different percentiles of the wage distribution, and the occupational structure of employment.

Other papers have empirically studied the effect of an educational expansion on wages of all educational groups by exploiting regional variation within the United States (Rauch 1993, Acemoglu & Angrist 2000, Moretti 2004*a*). They find that an increase in the supply of college graduates raises wages for all educational groups, even for college graduates who experienced an increase in supply, attributing these effects to productivity spillovers generated by college-educated workers. I find that an increase in the share of high educated workers also raises the wages of low educated workers in Brazil, but that wages of high educated declines. The main reason for this discrepancy could be that the empirical strategy of these papers only captures small increases in the share of high educated workers, as opposed to the major increase that took place in Brazil where the supply effect is more likely to dominate. Another reason could be that productivity spillovers require directed technology changes for the educational group that now is more abundant (Acemoglu 2002), which may take place over a longer period of time in developing countries than in developed ones.[9]

Another contribution of this paper is to provide evidence that the increase in education was the main factor behind the changes in the wage distribution in Brazil between 1995 and 2014. There is a heated debate in the literature on the contributing factors to the decline in inequality and wage gaps of workers with different educational levels in Brazil. These factors include: educational upgrading that declined returns to education (Barros et al. 2010, Gasparini et al. 2011, López-Calva et al. 2016, Alvarez et al. 2017); increases in the minimum wage that spread throughout the wage distribution (Engbom & Moser 2017); and the fall in returns to experience which compressed the wage distribution (Ferreira et al. 2016).[10] In my results, the educational expansion observed in Brazil not only explains most of the changes in key aspects of the wage distribution, but it is also consistent with the changes in the occupational structure of employment in the analyzed years. My results can be related to the other channels that have been identified in the literature. For example, it could be possible that the increase in the minimum wage was effective in raising wages and did not decrease formal employment because the labor market for low educated workers became much tighter as a result of the supply and

hired in preference to the less educated for unskilled jobs.

[9]Khanna (2015) uses similar techniques of exploiting regional and cohort variation to evaluate the general equilibrium effect of an educational reform in India. He finds that the reform largely increased the share of high educated workers, increasing the wages of low educated workers and declining the wages of high educated workers. This is consistent with the supply effects dominating in developing countries at least in short or medium-run.

[10]Other studies focus on different factors. Ulysseas (2014) and Meghir et al. (2015) study the of formal-informal gaps. Alvarez (2017) analyzes agricultural wage gaps. Dix-Carneiro & Kovak (2015) look at the effect of the trade liberalization during the nineties. Adão (2015) relates the decline in inequality to shocks in commodity prices

demand mechanisms generated by an educational expansion. Moreover, if returns to experience are larger in more complex occupations, returns to experience may have declined for each educational group because of the lower occupational attainment arising from the educational expansion. The relationship between these various sources of changes in the wage distribution constitutes an exciting research agenda.

This paper also contributes to the literature by reporting on distinctive patterns of changes in the occupational structure of employment that may arise in developing countries, contrasting with the job polarization patterns that have been found in developed countries. In the case of Brazil, the occupational structure has been particularly rigid and the educational expansion was mostly associated with lower occupational attainment, meaning that more educated workers are increasingly employed in occupations previously performed by lower educated workers. On the contrary, in developed countries the changes in occupational structure are characterized by a hollowing-out of middle-wage occupations, with a corresponding increase in low and high wage occupations, leading to job polarization (Katz et al. 2006, Goos et al. 2014). The literature for developed countries focuses on occupational changes resulting from the rapid change in technology in a context where education is relatively fixed (Acemoglu & Autor 2011, Autor & Dorn 2013, Deming 2017, Beaudry et al. 2016, Burstein et al. 2016).[11] For developing countries, especially for Brazil and most of Latin America, education may have been the most important factor in shaping the occupational composition of employment given its rapid increase over the last decades (Cruces et al. 2017).

Finally, I find that some educational expansions may generate lower occupational attainment across all educational groups. This result relates to the literature on overeducation (Leuven & Oosterbeek 2011), where there is also an allocation problem of deciding which workers will perform which jobs (Sattinger 1993, 2012), but there are mismatches arising from incomplete markets such as imperfect information or job-search frictions. In this literature, there are Pareto improvements to be made by switching workers across occupations, and policy interventions that ameliorate these frictions are desirable. In this paper, more educated workers also end up in occupations where their skills are less relevant (have a lower comparative advantage), but they are optimally assigned to those occupations and there is no wage or productivity gain to be made by switching workers across occupations. My finding indicates that more educated workers may be increasingly employed in occupations where their acquired human capital through formal education is less relevant. This lower occupational attainment results from labor market forces (demand and supply) acting in an environment where education increases and technology is fixed, as opposed to being originated by labor market frictions.

[11]It is reasonable that the literature on occupational changes for developed countries has focused on technological changes. In these countries, the expansion in education took place in the past century, and technological progress is the main force behind recent changes in the assignment of workers' types to occupations and their corresponding wages. In particular, there is a special interest in the effects of labor-saving technologies that have polarized the labor market in developed countries (Autor 2014).

3 Data

This paper use data from the Pesquisa Nacional por Amostra de Domicílios (PNAD), a nationwide household survey for the years 1995 to 2014. Workers are classified into low, medium and high educated. *Low educated* workers are those with completed primary education or less (less than 9 years of schooling). *Medium educated* workers are those with some or complete secondary education (between 9 and 11 years of schooling). *High educated* workers are those with some or complete university or another tertiary education (more than 11 years of schooling).

Hourly wages are expressed in 2005 purchasing power parity (PPP) dollars, and are estimated by dividing monthly labor earnings by hours worked in the corresponding month. The analysis is performed on employed workers between 18 and 55 years old, and living in regions that have been surveyed throughout the entire period.[12] Observations are re-weighted to hold constant the demographic composition within each educational group at the levels of 1995.[13]

Codes of different occupations are harmonized using the International Standard Classification of Occupations of 1988 (ISCO-88). In the household surveys occupations are classified according to the *Classificação Brasileira de Ocupações* (CBO), which changed between PNAD 1996-1999 and PNAD 2001-2014. I follow Salardi (2014) for recoding the occupations from the CBO for each period into ISCO-88 at 3 digits.

I follow the standard procedure in the literature of using the wage percentile rank of the occupation (Katz et al. 2006, Autor & Dorn 2013, Beaudry et al. 2016, Deming 2017). The wage percentile rank is usually estimated using the mean wage of the occupation in the baseline year. I opt for a different rank procedure that is better suited for my analysis: the median wage percentile rank of occupations taking into account the entire period 1995-2014. I prefer to use the median instead of the mean to avoid the influence of outliers. Moreover, I use all years between 1995 to 2014 for three reasons. First, to have more observations in each occupation. Second, the rank is less susceptible to changes in the characteristics of the occupations that may impact wages of particular occupations.[14] Third, the model below in Section 5 predicts that occupations have the same wage if performed by the same educational group in a given year so that differential changes in wages throughout the entire period will better reflect its ranking.

Occupations are usually considered in the literature as the best description a researcher has about the type of job a worker performs.[15] Some occupations are assumed to be more complex than others. For example, the tasks performed by an electrical engineer are arguably of higher

[12] I exclude regions from rural north. These regions were incorporated in the sample in 2004.

[13] I follow DiNardo et al. (1996) by constructing 20 cells within each educational group: two genders, five age categories and two sub-levels of education. I fix the participation of each cell to the initial share in 1995. This procedure avoids confounding changes in average wages of a particular group with a change in its gender composition, its age structure, or the education level within each educational group. Results are robust to not implementing this re-weighting procedure and are available upon request.

[14] For example, clerks had a decline in their relative position when compared to other occupations because of the change in the task content of their job. Library, mail and related clerks were in percentile 51 in 1995 and declined to percentile 17 in 2014, while using my classification they are in percentile 27. Except for some extreme cases like this one, the percentile an occupation belongs is similar by using the wage in 1995, the wage in 2014 or the wage during the period 1995-2014. The correlation is above .9 among all these alternatives, and the results are robust to using different classifications and available upon request.

[15] See Autor (2013) for a detail discussion on the information content of occupations.

complexity than that of a housekeeper. Although the tasks' content of different occupations may differ in more than one dimension (for example, routine manual, routine cognitive, and non-routine cognitive tasks), the relative wage of an occupation is assumed to be an indicative summary of how complex that occupation is when compared to other occupations in the economy. Therefore, to use wages to rank occupations provides useful information regarding their relative position in terms of how complex its tasks content is.[16]

4 Changes in the Brazilian labor market between 1995-2014

4.1 Education

During the 1995-2014 period, there was a major increase in the educational attainment of the Brazilian workforce. Table 1 shows the share of workers with low (primary or less), medium (some secondary) and high education (some university) in 1995 and 2014. The share of low educated workers decreased from 68.1 percent in 1995 to 36.4 percent in 2014, the share of medium educated increased from 20.5 to 40.0 percent, and the share of high educated workers rose from 11.3 to 23.6 percent. In only 19 years, the share of workers with medium and high educated doubled, while the share of low educated workers halved.[17]

To contextualize the extent and pace of the educational expansion that took place in Brazil, it is useful to compare it with changes in enrollment rates in the rest of the world over a similar period of time. Panel (A) of Figure 1 shows the changes in enrollment rates in secondary and university education across countries, according to the data from Barro & Lee (2013). Brazil is the country with the largest increase in secondary school enrollment, going from 16.0 percent in 1990 to 86.2 percent in 2010, closely followed by other Latin American countries and Portugal.[18] In university education, the enrollment rate in Brazil increased 20 percentage points, around the average of the countries considered in the data. The educational expansion in Brazil between 1990 and 2010 was also large from a historical point of view for the country. Panel (B) in Figure 1 plots the evolution of enrollment rates since 1975. Except for a moderate increase in primary education, enrollment rates were practically constant in Brazil between 1975 and 1990. From 1990 to 2010 enrollment in secondary schooling soared, while university enrollment increased steadily after 1995. The substantial increase in enrollment had an impact on the educational level of the labor force after the younger and more educated cohorts replaced older less educated cohorts in the labor market. According to Bruns et al. (2011), the rise in the educational attainment of Brazil's labor force since 1995 has been one of the fastest on record in history.

[16] This approach has been extensively used in the literature (Acemoglu & Autor 2011, Autor & Dorn 2013, Beaudry et al. 2016). However, there is an increasing literature that further discriminate among different tasks within an occupation by matching occupations with other datasets with detail tasks' information (for example, the Dictionary of Occupational Titles for the United States). This information is not available for the case of Brazil.

[17] The extent of the educational expansion is robust to using the share of total hours worked instead of the share of employed workers.

[18] The enrollment rate is defined as the ratio of students at a given level of schooling in the designated age group to the total population of that age group.

Brazil implemented several important educational reforms to increase the educational level of the population in primary, secondary, and university education.[19] Bruns et al. (2011) estimate that, as a consequence of these reforms, a six-year-old Brazilian child starting school in 2010 from the bottom quintile of the income distribution will, on average, complete more than twice as many years of schooling that her parents have. Data from the PNAD shows the large increase in secondary schooling between 1995 and 2014: the share of the population between 20 and 24 years old that finished secondary schooling increased from 24.0 to 63.3 percent during this period. Demand for higher education also skyrocketed. According to the data from Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), the number of students in tertiary education increased from 1.7 million in 1995 to 6.6 million in 2015. But the selection process became tougher: only 19 percent of the candidates where accepted 1995, as opposed to only 14 percent 2015. The number of candidates increased from 2.6 to 14.3 million. The supply and diversity of higher level institutions also increased. The number of institutions offering university/tertiary education went from 894 in 1995 to 2,364 in 2014.

Given the large educational expansion that took place in secondary and university level, it is a possibility that the quality of education went down in Brazil. On the contrary, the available evidence points out that the quality did not decline. In secondary education, the students' scores in the Program for International Student Assessment (PISA) increased slightly between 2000 and 2015 in each of the evaluated subjects (science, mathematics, and reading), and the scores in 2009 first-order dominates those of the year 2000, with the highest increase taking place on the bottom quintile (Bruns et al. 2011).[20] The scores are low by OECD standards, but they have not deteriorated in a time of a large educational expansion, which has been recognized as a remarkable achievement (Bruns et al. 2011). Regarding tertiary education, selection to top universities remains highly competitive, as was discussed previously. For example, the ratio of applicants to UNICAMP and USP (two of the largest universities in Brazil) is of 16 to 1. According to Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), approximatively 90 percent of the students in university/tertiary education attend an institution that applied some kind of selection in 2014.[21]

[19] Bruns et al. (2011) provide a detailed analysis of these reforms.

[20] PISA is regarded as one of the best measures of student outcomes. It allows comparison of results across countries and within a country in different periods.

[21] Changes in education quality can also be inferred by comparing data on wages of workers with the same educational level but that were educated at different periods. In the online Appendix B.1 I follow Bowlus & Robinson (2012) decomposition to disentangle what part of the changes in wages for each educational group is explained by a decrease in education quality (the quantity of human capital inherent to each education level) and which part corresponds to a decline in market prices. Intuitively, the procedure estimates changes in prices by looking at yearly changes in wages for workers aged 40-50 for which the quantity of human capital presumably does not change, and the changes in quality are estimated as the residual of the difference between total changes in wages and changes in prices. I conclude that practically all the changes on wages come from a change in the market price of human capital for each educational level, consistent with the human capital content within an educational level remained unchanged.

4.2 Wage distribution

Along with the educational expansion, the wage distribution in Brazil changed in several important ways between 1995 and 2014. Table 1 presents changes in different dimensions of the wage distribution. With respect to real wages, wages of low educated workers increased by 37.3 percent, while wages for medium and high educated workers fell by 10.5 and 21.6 percent respectively. Average wages increased by 27.8 percent during this period, drove by two factors, the increase in wages of low educated workers, and the large increase in the share of workers receiving wages of medium and high educated (the wage bill increased for these two groups despite the fall in average wages).

Wage inequality declined sharply in Brazil, contrary to what has happened in the United States since the 1970s. The wage gaps between workers with different educational level diminished, and all the indexes of relative income inequality declined.[22] In 1995, the average wage of high educated workers was 2.5 times those of medium educated, and 4.8 times the wage of low educated workers. These gaps declined to 2.2 and 2.7 in 2014, respectively. The Gini index went down 0.085 points, from 0.445 in 1995 to 0.360 in 2014. To better understand the importance of changes in wages for workers with different educational levels in the reduction of inequality, I compute other indexes that can be decomposed in between and within-group inequality. The Theil index diminished 0.109, with 20 percent of that fall explained by changes in inequality between the three educational groups—low, medium, and high educated—consider in this paper. The importance of the between-group inequality is larger in the case of the Atkinson index with an inequality aversion parameter equal to 2, which put more weight to changes at the bottom of the wage distribution. The index declined 0.132, with 59.2 percent of that decline due to falling inequality between workers with different educational level.

For any anonymous welfare function that is increasing in wages, the welfare level in Brazil increased during this period. Figure 2 display the cumulative density function (CDF) of wages in 1995 and 2014. It is easy to see that the wage distribution in 2014 first-order dominates that of 1995, therefore welfare increase for any welfare function belonging to the class of anonymous and increasing in individual wages (Saposnik 1981).[23] Because of the properties of first-order dominance, any wage poverty index for any line decreased between 1995 and 2014. This is depicted by considering a poverty line of 2.4: wage poverty rate fell from 62.4 to 46.2 percent.[24]

4.3 Occupational structure of employment

I find that the large educational expansion in Brazil was not matched by a corresponding occupational upgrading between 1995 and 2014. Table 2 displays the share of the workforce in each educational level in 1995 and its changes between 1995 and 2014 according to the 1 digit

[22]See Barros et al. (2010) for a detail characterization of the decline in inequality in Brazil.

[23]Usually welfare analysis is related to household per capita income. This paper is only concerned about labor market outcomes (note not household wellbeing) and the welfare analysis is done over the wage distribution.

[24]The 2.4 wage poverty line is ad hoc, but there is a rationale to use this level. It is the wage needed for a household composed of 4 individuals to be above the $4-dollar-a-day poverty line if the employed individual works 45 hours a week.

ISCO-88 classification. The occupational categories are ordered from higher pay to lower pay according to their median wage during the period 1995-2014. The changes are characterized by a small increase in the share of employment in top ranking occupations and a short decline in low ranking occupations. The share of professionals (the higher paid category) grew, but other high paid categories declined (managers, legislators, and military occupations). In the middle of the distribution of occupations, the share of clerks increased, compensated by a decrease in technicians, plant and machine operators and craft and related trade occupations. Finally, the share of service workers ros eon the low end of the distribution, while the share of agricultural or other elementary occupations fell. Overall, it is striking that while the share of professionals and managers increased only 2.5 percentage points from 13 to 15.5 percent between 1995 and 2014, the share of highly educated workers went up 12.3 percentage points from 11.3 to 23.6 percent.

Small changes in the occupational structure of employment are also found when considering occupations at the 3-digit level. I construct a ranking based on the median wage of each occupation considering all households surveys from 1995 to 2014.[25] The ranking of occupations allows to estimate statistics that are informative about changes in the occupational structure of employment. Table 2 shows the mean, median, and standard deviation of this ranking for 1995 and 2014. It also displays the changes in the employment share of the bottom-third, the middle-third and the top-third of the ranking of occupations. I also follow Acemoglu & Autor (2011) and Autor & Dorn (2013), among others, in estimating smooth regressions of the changes in employment share over the ranking of occupations, overall and for each educational group.

By looking across the rows of Table 2 for the total workforce, it becomes evident that the employment structure of occupations did not change much overall. The mean, median, and standard deviation of the ranking do not change. Furthermore, when occupations are divided into bottom-third, middle-third and top-third the shares of employment remained practically unchanged, which is consistent all movements across occupations happening among those of similar ranking. The locally weighted regression for the total workforce is shown in the blue line in Figure 3. The figure shows a small decrease in the share of employment in low ranking occupations and a similarly small increase in high ranking occupations. Note that the slightly positive slope in Brazil differs from to the U-shaped pattern that has been estimated for the United States and other developed countries (Acemoglu & Autor 2011, Goos et al. 2014), where occupations in the middle declined and those at both ends of the occupational distribution increased in the last three decades.

The red line in Figure 3 displays the result of a thought experiment simulating the changes in occupations between 1995 and 2014 if workers with a given educational level in 2014 were distributed across occupations as in 1995. This exercise simulates how the occupational structure of employment would have looked like in 2014 if the increases in education did not change the assignment of workers' type to occupations. The employment share in low ranking occupations (the 40 percent of occupations with lower ranking) would have largely declined, and

[25]The ranking is robust to using other measures as well. For more details see Section 3.

the employment share would have increased for occupations at the middle and at the top of the ranking. The difference between the blue and the red line is due to the lower occupational attainment within each educational, defined as an increase of employment in occupations of lower ranking relative to a previous time to which I now turn.

The small changes in the occupational structure of employment for the total workforce coupled with a large educational expansion resulted in a deterioration of the occupational distribution within each educational group, especially for medium educated workers. When looking at 1-digit occupations, the employment share in the five higher paid occupations declined for each educational type. It declined 5.0, 23.3, and 11.4 percentage points for low, medium, and high educated respectively. The lower occupational attainment becomes even more clear when looking at 3-digit level classification of occupations. In the case of low educated workers, the mean and the median rank declined, and the share employed in the bottom-third occupations increase 7 percentage points. For medium educated workers, the share in bottom-third occupations increased by 24.9 percentage points. For high educated workers, there was an increase in bottom-third occupations of 9 percentage points. This lower occupational attainment is depicted in Figure 3. All educational groups lost employment shares in higher rank occupations and increased their employment share in lower rank occupations, and the pattern is much stronger among medium educated workers.[26]

In summary, there was a very large educational expansion in Brazil between 1995 and 2014. Wages of low educated workers increased and wages of medium and high educated workers fell. Wage poverty and wage inequality declined. At the same time, the occupational structure of the employment of the Brazilian economy remained surprisingly stable during this period, while there was a large decline on occupational attainment for workers with of a given educational level.

5 A theoretical model

5.1 Environment

There is a unique final good, the economy is closed and there is no trade in tasks. The unique final good is produced by combining a continuum of tasks represented in the unit interval $[0, 1]$, with a Cobb-Douglas technology mapping tasks to the final good. All markets are competitive

[26]In the online Appendix B.2 I explore in more detail the relationship between changes in average wages of each educational group and their occupational composition of employment. Using the Oaxaca-Blinder decomposition, I disentangle what part of the changes in averages wages are due to an occupational composition effect (changes in the occupations workers perform) and what part is due to a pay structure effect (changes in wages within occupations). I find that for all workers the composition effect is small, consistent with small changes in the occupational composition of the economy; while the increase in wages is driven by higher wages at each occupation, especially those that started with a lower wage. I also find that all educational levels experienced a negative occupational composition effect, consistent with lower occupational attainment for each educational level. For low educated workers wages within each occupation largely increased, compensating for the negative composition effect, and average wages increased. For medium educated workers, wages increased only in a small number of occupations but not enough to compensate for the negative composition effect. In the case of high educated workers, wages declined within occupations, reinforcing the negative composition effect, and average wages largely fell.

and the final good is chosen to be the numeraire. Let Y denote the production of the unique final good and let $y(i)$ be the production level of the task i, then:

$$Y = exp\left[\int_0^1 ln\, y(i)\, di\right]. \tag{1}$$

Each task i can be produced by using three types of workers—low, medium, and high educated—which are perfect substitutes. Workers' types differ in their productivity to perform each task. Consider A_J to be the factor-augmenting technology of input J and $\alpha_J(i)$ to be the task-specific productivity of input J in task i. The production of each task i is defined as:

$$y(i) = A_L\alpha_L(i)L(i) + A_M\alpha_M(i)M(i) + A_H\alpha_H(i)H(i),$$

where $L(i)$, $M(i)$ and $H(i)$ are the amount of workers employed in producing task i with low, medium, and high education, respectively.[27] I assume tasks can be ordered by a unidimensional level of complexity represented by the index $i \in [0, 1]$, where 0 represents the less complex task and 1 stands for the most complex task. I further assume that the comparative advantage of more educated workers relative to less educated workers increases with the complexity of the task and they are continuously differentiable on i. That is,

$$\frac{\partial \frac{\alpha_M(i)}{\alpha_L(i)}}{\partial i} > 0, \; \frac{\partial \frac{\alpha_H(i)}{\alpha_M(i)}}{\partial i} > 0. \tag{2}$$

This is a key assumption of the model establishing the structure of comparative advantage across tasks. This assumption ensures positive assortative matching since more educated workers are optimally assigned to more complex tasks.

Let l, m and h be the share of low, medium, and high educated workers in the economy. There is no unemployment in the model. Factor clearing conditions require that:

$$\int_0^1 L(i)di = l; \int_0^1 M(i)di = m; \int_0^1 H(i)di = h. \tag{3}$$

Finally, I assume that there is a measure of workers equal to 1 in the economy,

$$l + m + h = 1; \tag{4}$$

The labor supply for each of the three types can be defined by two parameters, e.g. h and m. When either h or m goes up there must be a corresponding decrease in the share of workers with another level of education. This is the most significant difference between the environment of my model and that of Acemoglu & Autor (2011). This restriction is a fundamental feature of an educational expansion where some percentage of the workforce will move from one educational level to the another, increasing the supply of a more educated group by reducing the supply of a less educated group.

[27]Because of the assumption of a measure 1 of workers, $L(i)$, $M(i)$ and $H(i)$ are interpreted as the share of the population with a specific level of education employed in task i.

5.2 Equilibrium

The competitive equilibrium in this economy consists of an assignment of workers' types to tasks, an overall distribution of employment, and real wages for each type of worker, such that producers maximize profits and labor markets clear, given the supply of skills and the productivity across tasks of each type of worker.

The equilibrium conditions in this model are similar to those in Acemoglu & Autor (2011) and its derivations are detailed in Appendix 1. The equilibrium is easy to characterize given the positive assortative matching that arises from the supermodularity in the production function and the assumption that the comparative advantage of more educated workers with respect to less educated workers is increasing in i. In particular, there exist two thresholds that determine which tasks are carried out by low, medium and high educated workers. Restating Lemma 1 from Acemoglu & Autor (2011):

Lemma 1. *In any equilibrium there exist $\{I_L, I_H\} \in (0,1)$ with $I_L < I_H$ such that for any $i < I_L$, $L(i) > 0$ and $M(i) = H(i) = 0$; for any $I_L < i > I_H$, $L(i) = H(i) = 0$; and $M(i) > 0$, and for any $i > I_H$, $L(i) = M(i) = 0$ and $H(i) > 0$.*

These thresholds naturally arise from the assumption that comparative advantage is increasing in i, generating positive assortative matching. Intuitively, it is optimal for employers to use the most productive workers (high educated) to the more complex tasks (closer to the index 1), where they have a larger comparative advantage. On the opposite, it is in their profit-maximizing interest to employ low educated workers in the tasks of low complexity (closer to 0). And finally, medium educated workers will be employed in the remaining task with a medium level of complexity. The location of these thresholds depends on the relative supply of workers' types and on their relative productivities in each task, both of which are assumed to be exogenous in the model.

In equilibrium, thresholds I_L and I_H are obtained by solving the following system of equations:

$$\frac{(I_H - I_L)}{I_L} \frac{\alpha_L(I_L)}{\alpha_M(I_L)} = \frac{A_M}{A_L} \frac{m}{(1 - m - h)}. \tag{5}$$

$$\frac{(1 - I_H)}{I_H - I_L} \frac{\alpha_M(I_H)}{\alpha_H(I_H)} = \frac{A_H}{A_M} \frac{h}{m}. \tag{6}$$

Equation (5) follows from equalizing the cost of producing task I_L with low and medium educated workers. It states that there is no profit to be made by employers in switching workers with medium and low education across tasks. Similarly, equation (6) comes from equalizing the cost of task I_H if produced by medium or high educated workers. These equations provide a unique mapping between the thresholds I_H and I_L, and the relative supply and relative productivities of low, medium, and high educated workers across different tasks. These thresholds are uniquely determined. To see this, consider the case of equation (5). If I_L is equal to zero, the left-hand-side (LHS) goes to infinite; if I_L is equal to I_H the LHS is equal to zero. The LHS is decreasing in I_L given the assumption that more educated workers are more productive in more complex tasks; the right-hand-side is positive and does not depends on I_L. Taken together, there is a unique value of $I_L \in (0, I_H)$ that solves (5), given a value of I_H. Using similar arguments, there is a unique value of $I_H \in (I_L, 1)$ that solves equation (6). Finally, there is a unique pair $I_L, I_H \in (0, 1)$ that solves both equations simultaneously.

Let AVP be the average productivity in the production of tasks in the economy. $C_{HM}(i)$ be the comparative advantage of high with respect to medium educated workers at task i, and $C_{ML}(i)$ be the comparative advantage of medium relative to low educated workers at task i. Wage levels can be express in terms of AVP, $C_{HM}(I_H)$, $C_{ML}(I_L)$ and the thresholds levels I_H and I_L, that is:

$$ln\, W_H = AVP + I_H C_{HM}(I_H) + I_L C_{ML}(I_L) \tag{7}$$

$$ln\, W_M = AVP - (1 - I_H) C_{HM}(I_H) + I_L C_{ML}(I_L) \tag{8}$$

$$ln\, W_L = AVP - (1 - I_H) C_{HM}(I_H) - (1 - I_L) C_{ML}(I_L) \tag{9}$$

where,

$$AVP = \int_0^{I_L} ln\, A_L \alpha_L(i)\, di + \int_{I_L}^{I_H} ln\, A_M \alpha_M(i)\, di + \int_{I_H}^{1} ln\, A_H \alpha_H(i)\, di.$$

and

$$C_{HM}(j) = ln\, A_H \alpha_H(j) - ln\, A_M \alpha_M(j)$$

$$C_{ML}(j) = ln\, A_M \alpha_M(j) - ln\, A_L \alpha_M(j).$$

Equations (7)-(9) arise from the assumption of competitive labor markets, where workers of the same type earn the same wage, which is equal to the value of their marginal product.[28]

The wages for each type of worker have two components, one is common for all workers and the other contains two terms that are specific to each type of worker. The common factor is the average productivity of tasks' production in the economy. The type-specific component adds (subtracts) to the average labor productivity depending on how more (less) productive workers of a type are in the task thresholds, weighted by the number of tasks each type performs.

From taking the difference of equations (7)-(9), the wage gaps in the model are equal to the productivity differential at the task thresholds. That is:

$$ln\, W_H - ln\, W_M = C_{HM}(I_H) \tag{10}$$

$$ln\, W_M - ln\, W_L = C_{ML}(I_L) \tag{11}$$

$$ln\, W_H - ln\, W_L = C_{HM}(I_H) + C_{ML}(I_L). \tag{12}$$

[28]Each task i in this economy will have a price that will exactly compensate for the difference in labor productivity with any other task i' that is produced by the same type of worker. For example, $p(i)\alpha_H(i) = p(i')\alpha_H(i')\, \forall \{i, i'\} \in (I_H, 1)$, where $p(i)$ is the price of tasks i. See Appendix 1 for more details.

Finally, tet E_i be the employment in each task. It can be defined as:

$$E_i = \begin{cases} L = L(i) = \frac{(1-m-h)}{I_L}, & if \;\; 0 < i < I_L \\ M = M(i) = \frac{m}{(I_H - I_L)}, & if \;\; I_L < i < I_H \\ H = H(i) = \frac{h}{(1-I_H)}, & if \;\; I_H < i < 1. \end{cases} \tag{13}$$

Equation (13) implies that employment is the same among tasks performed by the same type of worker. This result follows from the Cobb-Douglas technology of the production function with a unitary elasticity of substitution, which implies that expenditure across all tasks must be equalized in order to minimize cost, and the fact that task's prices produced by a given type exactly offset the workers' productivity differential.[29] These expressions, together with (5) and (6), imply that $L > M > H$, which means that the employment share in tasks performed by low educated workers (those of lower complexity) is higher than the employment share in tasks performed by medium educated workers, which in turn are higher than those of high educated workers. The intuition is that because high educated workers are more productive in all tasks, employers find it profitable to use them in a broader set of tasks.

An intuitive depiction of the equilibrium is shown in Figure 5. The figure contains the level of task complexity (i) on the horizontal axis and the relative productivities in the vertical axis. The thresholds I_L and I_H define the assignment of workers' types to tasks, such that low educated workers perform tasks below I_L, medium educated workers those between I_L and I_H, and high educated workers carry out the most complex tasks above I_H. The function that defines the comparative advantage across tasks, $C_{HM}(i)$ and $C_{ML}(i)$, are increasing in the complexity of the task defined by i, and relative wages are defined by the comparative advantage (relative productivity) among different types at the thresholds I_L and I_H.

All the outcomes of interest in this paper are defined by equations (5)-(13). They determine: the distribution of workers' type to tasks; the wage levels for each type; the wage gaps between types; and the employment distribution of tasks in the economy. I now turn to the question of how an educational expansion affects these outcomes.

5.3 Labor Market Effects of an Educational Expansion

An educational expansion in the model is defined as:

Educational expansion: There exist an educational expansion between time t and t' if $m'_t \geq m_t$ and $h'_t \geq h_t$, with at least one of those relationship holding with strict inequality.[30]

In order to study the changes in the task composition of the economy, it is useful to define *bottom-level*, *middle-level* and *top-level* tasks according to whether they were originally performed for low, medium or high educated workers, respectively, before the educational expansion took place. These employment levels are represented by E_B, E_M, and E_T.

[29] The equal expenditure of producers across tasks follows from the dual problem of minimizing costs for a given level of production. For more details see Appendix 1.

[30] In the rest of the paper the subscript t is dropped to ease notation.

The summary of the labor markets effects of different educational expansions is shown in Table 3. I study 3 policy experiments of educational expansions: (i) an increase in h by declining l; (ii) an increase in m by declining l; and (iii) an increase in both m and h. Case (iii) is a mixture of the first two cases and can be further divided into two sub-experiments (iii.a) or (iii.b) depending on whether it is case (i) or case (ii) that dominates.

The effects of an educational expansion on the equilibrium conditions of the model are solved sequentially: equations (5) and (6) first determine the distribution of workers to tasks and, combined with equation (13), determine the overall employment composition of the economy. These equations only depend on the relative supply of skills, which changes with the educational expansion, and the comparative advantages across tasks, which is assumed to be fixed. Then, equations (7)-(12) determine the new wage levels and relative wages, depending on the new task thresholds and holding fix the technology of the economy. In other words, I first determine the effect of each policy experiment on the task thresholds, and the effects on wages follow from it.

I postulate one proposition for each policy experiment (i)-(iii).

Proposition 1 *Under the policy experiment* (i)*, the share of high educated workers increases from* h *to* h' *with a corresponding decline in the share of low educated such that* $\Delta h = h' - h = -\Delta l > 0$*, holding the share of medium educated constant, generating:*

- **(i.1)** *Changes in thresholds:* $I'_L < I_L$, $I'_H < I_H$;
- **(i.2)** *Changes in the occupational structure:* $\Delta E_B > \Delta l$, $\Delta E_M < 0$, *and* $\Delta E_T < \Delta h$;
- **(i.3)** *Changes in real wages:* $\Delta W_L > 0$, $\Delta W_H < 0$, *and* $\Delta W_H < \Delta W_M < \Delta W_L$;
- **(i.4)** *Changes in relative wages:* $\{\Delta \frac{W_M}{W_L}, \Delta \frac{W_H}{W_M}, \Delta \frac{W_H}{W_L}\} < 0$.

Proof: See Appendix A.3.

Condition (i.1) indicates that, under policy experiment (i), all workers' types end up more concentrated in occupations of lower complexity than before (lower occupational attainment). Condition (i.2) establishes that the employment share of low-level occupations declines less than the reduction in the share of low educated workers because some medium educated workers start to carry out some of these tasks. Similarly, the employment share in top-level occupations increases by less than the change in the share of workers with high education because some of these workers start to perform medium-level tasks. Conditions (i.3) and (i.4) result from the general equilibrium effects of changes in supply and demand for workers with different skill levels.

These results are intuitively appealing. When the share of high educated workers increases, they become more abundant and their wages fall in the tasks they were originally employed, while low educated workers become less abundant and their wages increase in the tasks where they were originally employed. Therefore, it becomes profitable for firms to start hiring high educated workers on tasks previously performed by medium educated workers, and to use medium educated workers in tasks previously performed by low educated workers. Then, all workers' types end up more concentrated in occupations of lower complexity than before (lower occupational attainment). The changes in the tasks composition of employment of the whole economy are smoothed by more educated workers starting to perform occupations of lower complexity. Wages for high educated workers decline because of the increase in supply. Wages of low educated workers rise because of the decline in supply coupled with a higher demand for the tasks they perform, given that there is an increase in production of other tasks in the economy.[31] Changes in wages of medium educated workers are always between those of low and high educated workers and it could be positive or negative (the negative effect of the lower occupational attainment is offset by an increase in the value of the new tasks they perform). Wages gaps of more educated workers with respect to lower educated workers decline. The workers that benefit the most from an increase in the share of high educated are those who get educated (their wages raise to those of high educated workers), but also the remaining low educated for which wages increase. The workers that are hurt the most are those that already were high educated.

Panel (A) in Figure 6 displays some of the effects of policy experiment (*i*). When the share of higher educated workers increase, their wages decline and it becomes profitable for employers to use them in tasks of lower complexity (decline in I_H), moving medium educated workers downward in the occupational ladder which in turn also displaced low educated workers towards tasks with lower complexity (decline in I_L). Wage gaps decline given the fall in the comparative advantages of higher to medium and that of medium to low educated workers at each threshold.

Proposition 2 *Under the policy experiment* (*ii*), *the share of medium educated workers increases from* m *to* m' *with a corresponding decline in the share of low educated such that* $\Delta m = m' - m = -\Delta l > 0$, *holding the share of high educated constant, generating:*

(ii.1) *Changes in thresholds:* $I'_L < I_L$, $I'_H > I_H$, $(I'_H - I'_L) > (I_H - I_L)$, *and* $(I'_L + I'_L)/2 \lesseqgtr (I_L + I_H)/2$;

(ii.2) *Changes in the occupational structure:* $\Delta E_T > 0$, $\Delta E_B > \Delta l$, *and* $\Delta E_M < \Delta m$;

(ii.3) *Changes in real wages:* $\Delta W_M < 0$, *and* $\{\Delta W_L, \Delta W_H\} \lesseqgtr 0 > \Delta W_M$;

(ii.4) *Changes in relative wages:* $\Delta W_H/W_M > 0$, $\Delta W_M/W_L < 0$, *and* $\Delta W_H/W_L \lesseqgtr 0$.

Proof: See Appendix A.3.

Condition (ii.1) establishes that low educated workers are displaced to occupations of lower complexity, high educated workers are employed in more complex tasks, medium educated workers take over a larger number of tasks in the middle of the tasks distribution such that the average task that they perform may increase or decrease.[32] Condition (ii.2) follows from

[31] Although the productivity to perform each task does not change with an educational expansion for any group, there is an implicit complementarity in the production of the final good by combining occupations in a Cobb-Douglas production function with an elasticity of substitution equal to one.

[32] The average task is informative of how complex are the jobs each worker is performing, and it can be compared to the average occupational ranking in the data for each educational group.

condition (ii.1). It shows that the employment share in top-level tasks increase because a share of medium educated workers is incorporated into these tasks. Similarly, employment share in bottom-level tasks declines less than the reduction in the share of low educated workers because some medium educated workers start carrying out some of these tasks. Because some share of medium educated workers start to perform tasks they did not use to do, changes in employment in medium-level tasks are lower than the increase in the share of medium educated workers. Condition (ii.3) and (ii.4) result from the general equilibrium effects of changes in supply and demand for workers with different skill levels. Wages of medium educated workers decline, while wages of the other groups may increase or decrease but the changes are always higher than the decline in wages of medium educated. The changes in real wages decline of wage gaps between medium and low educated, and an increase in the wage gaps of high to medium educated.

These results can be intuitively interpreted as follows. When the share of medium educated workers increases, they become more abundant and their wages fall. It becomes profitable for employers to start using them in tasks previously performed by low and by high educated workers. Therefore, there is an expansion of the number of tasks carried out by medium educated. As the tasks performed by medium educated expands, there is a displacement of low educated workers towards tasks of lower complexity and of high educated workers to tasks of higher complexity. The employment share in top-level tasks expands because there is an influx of medium educated workers on these tasks. In terms of wages, the supply effect dominates for medium educated and their wages decrease. For low and high educated workers, wages can increase or decrease. For high educated workers, they are now concentrated into more complex tasks for which they are more productive, but the price for these tasks diminishes because of an increase in the amount produced. Therefore, the value of their marginal product (because we assume markets are competitive, wages equal the value of their marginal product) may increase or decrease. On the contrary, low educated workers are now concentrated in tasks of lower complexity, where they are less productive (although having the comparative advantage on those tasks, productivity is assumed to increase with the complexity of the task). But the prices for these tasks may increase (they are relatively more scarce than before). The value of their marginal product (wages) may increase or decrease depending on which effect is stronger. Relative wages of medium with respect to other workers decline given the larger drop in its wages, while that of high educated with respect to low educated may increase or decrease. The workers that benefit the most from an increase in the share of medium educated are those who get educated (their wages raise to those of medium educated workers) and the workers that are hurt the most are those that already have medium education (their wages will have the largest fall).

Panel (B) in Figure 6 depicts the effects of policy experiment (*ii*). As medium education expands, wages for medium educated workers decline and they start to perform a broader set of tasks in the economy, pushing low educated workers towards tasks of lower complexity (decline in I_L) and driving high educated workers to tasks of higher complexity (increase in I_H). The wage gaps of medium to low declines and that of high to medium educated increases (there is movement along the curves C_{ML} and C_{HM} respectively).

Proposition 3 *Under the policy experiment (iii), the share of medium and high educated workers increases from m to m' and from h to h', with a corresponding decline in the share of low educated, such that $\Delta m = m' - m > 0$, $\Delta h = h' - h > 0$, and $\Delta m + \Delta h = -\Delta l$, generating:*

(iii.1) $\Delta I_H \lesseqgtr 0$. *If $\Delta I_H > 0$ it must be the case that $\frac{m'}{h'} > \frac{m}{h}$.*

(iii.2) *If $\Delta I_H < 0$, preposition 1 holds.*

(iii.3) *If $\Delta I_H > 0$, preposition 2 holds.*

Proof: See Appendix A.3.

If the share of medium and high educated workers increases at the same time, there is a mixture of the two cases discussed above. Condition (iii.1) shows that the threshold that separates the tasks of medium and high educated workers may increase or decrease, depending on the extent of the changes in relative supply and on the technology reigning the production function in the economy. It also establishes that for I_H to increase it is necessary (but not sufficient) that the relative supply of medium with respect to high educated workers increases. Conditions (iii.2) and (iii.3) determine that the changes in the equilibrium of the model are dominated by the increase in h or the increase in m depending on the directional change in I_H. In other words, which one dominates depends on whether in the new equilibrium high educated workers displace medium educated workers towards tasks of lower complexity ($\Delta I_H < 0$, effect of increase in high education dominates), or if middle educated workers displace high educated workers in some of the tasks they used to perform ($\Delta I_H > 0$, effect of increase in medium education dominates).

By looking at propositions 1,2 and 3 and across columns of Table 3, there are some common patterns of adjustment across all the policy experiments. First, there is a decline of the occupational attainment for low educated workers under any educational expansion (I_L always decreases). Second, the changes in the occupational composition of employment are smooth when compared to changes in the educational composition of employment, due to changes the assignments of workers' types to tasks when there is an educational expansion in the economy. Finally, wages always decline for at least one educational group.

Corollary 1: When any educational expansion takes place, there is always a decline of the occupational attainment for low educated workers, smooth changes in the occupational composition of employment, and a decline in wages for at least one educational group.

5.4 Welfare analysis: changes in the CDF of the wage distribution, poverty and inequality

This section studies the effects of different policy experiments on the wage distribution by looking into cumulative density functions (CDF), growth incidence curves (that reflect wage changes at each percentile of the anonymous wage distribution) poverty and inequality. In the model, the CDF is estimated using the share of workers with different educational level and their respective wages, before and after each educational expansion. Some of the directional changes in wages are ambiguous and depend on the parameters of the model, according to prepositions 1-3. When this is the case, there are two opposite forces pushing wages up and down. Wages in those cases may not change much due to the compensating forces. In this section, I make a simplifying assumption that wages remain at a similar level than before the educational expansion when the model provides ambiguous predictions. Figure 7 shows the CDF before and after the educational expansion for each policy experiment. The changes in CDF are better characterized by the growth incidence curves of wages (GIC), which captures graphically the wage changes for every anonymous percentile of the wage distribution. Figure 8 depicts the GIC for each policy. It shows the absolute change in wages in the vertical axis and the percentile of the wage distribution (divided by 100) in the horizontal axis. The label "J'"(prime) symbol indicates the value of variable J after the corresponding policy experiment.

Panel (A) in Figures 7 and 8 displays the effect policy experiment (i). An educational expansion in high education increases wages across the entire wage distribution, except for the wages at the top, those that originally where high educated. The new high educated workers move up in the wage distribution on top of those with medium education, increasing the educational level of most percentiles in the middle of the wage distribution. There is an increase in wages for the lower percentiles that remain with low education ($\bar{L}$), as predicted by the model. Although the share of medium educated workers does not change, they move to the left of the wage distribution, increasing the wages of percentiles that move from being low to medium educated ($L \rightarrow M$). Some percentiles remain medium educated and wages could increase or decrease ($\bar{M}$), and are depicted by a dashed line. Wages for percentiles the move from medium to high educated increase ($M \rightarrow H$) because more educated workers have higher earnings (albeit the fall due to the educational expansion). Finally, the wage falls for the percentiles that already were high educated located at the top of the distribution ($\bar{H}$).

Panel (B) shows the effects of policy experiment (ii). With an expansion in medium education, the wages of percentiles that were originally low educated and are now medium educated ($L \rightarrow M$) increase to the new level of medium educated wages ($W_M' - W_L$), while that of percentiles that were already medium educated ($\bar{M}$) decline as predicted by the model. The changes in wages for those percentiles that remain low and high educated ($\bar{L}$ and $\bar{H}$, respectively) are ambiguous. It is clear from the figure which anonymous percentiles benefit and which are hurt with an educational expansion in medium education. Not surprisingly, the percentiles where the educational level increases benefit by a rise in their wages, while percentiles that already had wages of medium educated are hurt with the fall in wages as supply increases.

Panel (C) presents the effect of policy experiment (iii) when the increase in supply of high educated dominates (case $iii.A$ in Table 3). The effects in the wage distribution are similar to those in Panel (A), with the only difference of a larger share of percentiles that were originally low educated are now medium educated ($L \rightarrow M$).

Finally, Panel (D) shows the effect of policy experiment (iii) when the increase in medium education prevails (case *iii.b* in Table 3). In addition to the effects in Panel (B), there is an increase in the wage of percentiles that become high educated ($M \rightarrow H$). Note that the larger the increase in the share of high educated the smaller the number of percentiles that remain medium educated ($\bar{M}$). If the increase in high educated is large enough, these percentiles for which wage diminishes disappear ($\Delta h > m$). This is the only policy experiment in the context of the model where a wage distribution after an educational expansion can first order dominate the original wage distribution.[33]

Looking across panels of Figures 7 and 8, I find that an educational expansion concentrated in medium education mostly changes the middle of the wage distribution, while an educational expansion concentrated in high education reshapes the entire wage distribution.

I now turn to the welfare analysis. Because an educational expansion always declines the wages of at least one educational group (see corollary 1), in most educational expansions there is no welfare improvement characterized by a first order dominance. But a first order dominance is a very demanding welfare criteria to evaluate the policy experiments analyzed here. A more practicable welfare analysis can be performed by constructing an abbreviated welfare function, with welfare depending negatively on poverty and inequality, and positively on the total production of the economy. Next, I analyze the predictions of the model on these three relevant outcomes.

Changes in poverty follow from the changes in the CDF of the wage distribution depicted in Figure 7. The poverty rate is defined as the percentage of workers below a wage poverty line. If the wage poverty line is anywhere between W_L and W_M, the poverty rate declines under policy experiments (i) and ($iii.a$), but it may increase in policy experiments (ii) and ($iii.b$) if the new wage for medium educated workers is below the wage poverty line. Note also that other poverty indexes that aggregate poor individuals according to how far they are from the poverty line will also decline under policy experiment (i) and ($iii.a$), given the increase in wages of those that remain in poverty (workers with low education, considering that the wage poverty line is lower than their new wage level). This is not ensured in experiments (ii) and ($iii.b$) because wages of low educated workers may fall and the gap between wages and the poverty line may increase.

In terms of inequality, the most natural measures of inequality in this context are the wage gaps. Inequality declines under policy experiments (i) and ($iii.a$), given that all wage gaps diminish. For policy experiments (i) and ($iii.b$), changes in inequality are ambiguous due to a decline in one wage gap (W_M/W_L) and an increase in another (W_H/W_M).

Finally, total production increases with any educational expansion.[34] This result arises from the assumption that more educated workers also have absolute advantage in the production of tasks, so that a more productive labor force that is fully employed under the equilibrium conditions of the model always produces more output.[35]

Corollary 2: Without considering the cost of education, an educational expansion that

[33]It is also necessary that the parameters of the model are such that the changes in wages of low and high educated workers, which are ambiguous in the model, are not negative.

[34]Given that there is a measure of workers equal to 1, total production is equal to production per worker.

[35]The formal proof will be part of the appendix in a future version of the paper.

increases in the share of high educated workers (or when it dominates) always improves welfare measured by an abbreviated social welfare function which depends on poverty (negatively), inequality (negatively), and total output (positively). This may not be the case of an expansion in the share of medium educated workers (or when it dominates).

5.5 The importance of the comparative advantage across tasks

This section discusses the importance of the shape of the comparative advantage across tasks for the effects of the different policy experiments on occupations and wages. To that end, I evaluate the effects of an increase in high education. The only assumption in the model related to the comparative advantages across tasks is that it must be a function that is continuous, differentiable, and increasing in i, according to equation (2). The slope of this relationship is crucial to determine the effects of an educational expansion. In particular, if the comparative advantages are steep, the changes in the thresholds are small and the changes in wages are large. On the contrary, if the comparative advantages are flat, an educational expansion results in large changes in the thresholds and small changes in wages.

To see this, consider four different combinations of comparative advantage schedules. The comparative advantage of high educated with respect to medium educated workers, $C_{HM}(i)$, could largely or barely increase with i. Let each of these cases be *steep* $C_{HM}(i)$ and *flat* $C_{HM}(i)$, respectively. Similarly, define two cases for the comparative advantage of medium with respect to low educated, *steep* $C_{ML}(i)$ and *flat* $C_{ML}(i)$. There are four possible combinations of steep and flat comparative advantages across tasks. A numerical example of the labor market effects of an increase in h under each of the four cases is presented in Table 4 and depicted in Figure 9.[36]

By looking across columns, it is clear that in the case of flat comparative advantages, an educational expansion produces large changes in occupations and small changes in wages. On the contrary, in the case of steep comparative advantages, an educational expansion mainly affects workers' wages, with a small impact in their assignment to different occupations. If the curves are flat, employers find it profitable to use the more abundant high educated workers in many new tasks that were previously performed by medium educated workers, with only a small cost in terms of relative productivity on those tasks. In other words, the demand for high educated workers is more elastic due to the easiness of switching to occupations of lower complexity for which they are almost as productive as in the occupations they were originally performing. When the comparative advantages are steep, wages have to diminish sharply before employers find it profitable to start employing high educated workers in a small number of tasks of lower complexity. In this case, the labor demand is more inelastic and wages react more to

[36]The table and figure show the effect of an increase in h of 10 percent points, from 0.2 to 0.3. The rest of the parameters are fixed and they are defined as follows: $m = 0.4$, $A_H = 4$, $A_L = 2$, $A_M = 2.5$. In the flat scenario,$\alpha_L = 1 + 0.0001 * i$; $\alpha_M = 1 + 0.01 * i$, $\alpha_H = 1 + 0.05 * i$. In the steep scenario, $\alpha_L = i$; $\alpha_M = i^2$, $\alpha_H = i^3$, while A_L, A_M, and A_H are estimated so that the thresholds and the initial wage is the same than in the flat scenario.

changes in supply.

6 Calibration

The procedure to calibrate the parameters of the model with data from Brazil is fairly simple. The set of parameters that need to be estimated are the skill supplies and the productivity across tasks for each worker type. I used the baseline year, the PNAD survey of 1995, to calibrate all these parameters.

The parameters corresponding to the educational level of the workforce come directly from the data. The values of l, m, and h are the share of the employed workers with low (less than secondary), medium (some secondary), and high education (some university) in the survey of 1995.

What remains to be estimated are the factor-specific augmenting technologies common to all tasks A_L, A_M and A_H, and the specific productivity across tasks $\alpha_L(i)$, $\alpha_M(i)$ and $\alpha_H(i)$. To that end, I use the average wages for low, medium and high educated workers in 1995. My identifying assumption is to assume a functional form for the functions $\alpha_L(i)$, $\alpha_M(i)$ and $\alpha_H(i)$ that largely simplify the equilibriums conditions in the model.

With wage gaps and relative supplies in 1995 it is possible to estimate the thresholds levels I_L and I_H from the equilibrium conditions of the model. From equations (6), (5), (10), and (11), it is possible to express:

$$\frac{(1 - I_H^*)}{I_H^* - I_L^*} = \frac{W_H}{W_M}\frac{h}{m}, \tag{14}$$

$$\frac{(I_H^* - I_L^*)}{I_L^*} = \frac{W_M}{W_L}\frac{m}{(1 - m - h)}. \tag{15}$$

where I_L^* and I_H^* are the task thresholds in 1995. In terms of Figure 5, it solves for the thresholds levels and the wage gaps corresponding to each threshold. The function that goes through the point where the thresholds meet with the relative wages, which determines the relative productivities across all tasks defined as $C_{HM}(i)$ and $C_{ML}(i)$, is estimated next. I assume a functional form for these functions. In particular:

$$\alpha_L(i) = i;\ \alpha_M(i) = i^2;\ \alpha_H(i) = i^3.$$

Then, the functions that determine the relative productivity across tasks are

$$C_{HM}(i) = ln\,\frac{A_H}{A_M}i,$$

$$C_{ML}(i) = ln\,\frac{A_M}{A_L}i.$$

The above expressions combined to (10) and (11) solve for $\frac{A_H}{A_M}$ and $\frac{A_M}{A_L}$

$$\frac{W_H}{W_M} = \frac{A_H}{A_M} I_H^*, \Rightarrow A_H = \frac{W_H}{W_M} \frac{A_M}{I_H^*}$$

$$\frac{W_M}{W_L} = \frac{A_M}{A_L} I_L^*, \Rightarrow A_M = \frac{W_M}{W_L} \frac{A_L}{I_L^*}.$$

Finally, I use a simulated method of moments to target A_L so that (9) solves for the observed average wage for low educated workers in 1995. As result of this calibration, $A_L = 2.81$; $A_M = 12.39$ and $A_H = 47.36$, which implies that $C_{HM}(i) = ln\, 3.82i$ and $C_{ML}(i) = ln\, 4.42i$.

The calibrated model using these parameters matches the initial moments in the data relatively well. Table 5 contains the results of the calibration for the targeted and non-targeted moments. The supply of skills and average wages in 1995 are used to calibrate the model, so that its values perfectly match that of the model. The average task performed by each educational level and the overall distribution of occupations are not targeted in the model. According to the model the average task of low educated workers is 0.21 in 1995, while in the data it is 0.24, while the average tasks in 1995 for medium and high educated are slightly overestimated in the model.[37] The model matches that occupations of lower ranking (bottom-third) have the higher share of employment in the economy, while employment in higher rank occupations where high educated workers are employed (top-third) have the lower share of employment. This results in the model from the optimal decisions of employers that is it profit maximizing to use the most productive workers in a larger range of tasks, diminishing the employment share on those tasks.

The inequality of the wage distribution in the model is, as expected, lower than that of the data, given that in the model there is no inequality within each educational group. The Gini index in the model is 30 percent lower than that of the data, but the level of inequality in the model is higher than the between component of the Theil and the Atkinson(2) indexes. The model does a better job on fitting different indexes of wage poverty with a poverty line of $2.4. Although the model cannot capture the level of inequality at a given point in time because it only produces three different wages for the total workforce, it is useful to understand changes in inequality that arise from variations in these wages and in the share of workers that earns them.

7 Results

7.1 Wage distribution and occupational structure

This section compares changes in labor market outcomes in the Brazilian data between 1995 and 2014 with the predictions of the model when the Brazilian educational expansion takes place, holding constant the rest of the parameters of the model. I first discuss the qualitative results,

[37] I consider that the ranking of an occupation is informative of its complexity, such that a higher ranking occupations is more complex than a lower ranking occupation. In a strict sense, there is a correspondence between the ranking of the occupation and the index of task complexity in the model, but one index can be a monotonic transformation of the other, preserving the order but not necessarily the distance among each other.

followed by a quantitative assessment of the model's predictions arising from the calibrated model.

Table 6 displays the directional changes in the data for Brazil between 1995 and 2014, and the qualitative results of the model when there is an increase in the share of medium and high educated workers.[38] The qualitative predictions of the model exactly match the changes observed in the data: the average complexity of the task performed by each type of worker declines (lower educational attainment for each educational group); the occupational composition of the workforce changes smoothly, so that the share of employment in tasks originally performed by high (low) educated workers increase (decrease) by less than the increase in the share of high educated workers (decline in the share of low educated workers); wages of low educated increases, wages of high educated workers decline, and changes in wages of medium educated workers are between those of low and high educated; and wage gaps of more educated workers with respect to less educated workers fall.

In the model, changes in the labor market can only arise from supply shocks or from technological shocks. I showed that a supply shock as the one that took place in Brazil matches all the patterns observed in the data. It remains to be shown that changes in other parameters of the model cannot generate these patterns. In the model, if the productivity of all workers increases by the same ratio (factor-neutral technical change), there is no effect on the assignment of workers to tasks, and wages of all types increase in the same proportion, leaving wage gaps are unaffected. In the case of a skill-biased technological change, one the main factors influencing the labor markets during the 1990s in the United States and other developed countries (Acemoglu 1998, Bekman et al. 1998) it is portrayed in the model by an increase in A_H. Although it generates a pattern of lower educational attainment for each group consistent to what is observed in the data, it also generates an increase in wages of high educated workers, which is not consistent with the data. The only technological change that can reduce the wage gaps is an increase in the productivity of low educated workers A_L, but it would also imply occupational upgrading instead of lower educational attainment: less educated workers will start to perform tasks of higher complexity. In short, only an educational expansion can generate the patterns on occupations and wages observed in the data in the context of the model.

I turn now to the quantitative assessment of the effects of an educational expansion through the lens of the model. To this end, I estimate the changes in several outcomes of interest in the calibrated model generated by the educational expansion that took place in Brazil, that is: l decrease from 0.68 to 0.36, m increases from 0.21 to 0.40, and h goes from 0.11 to 0.24. Table 7 presents the changes in the occupational structure of employment and the wage distribution. The table compares the changes observed in the data between 1995 and 2014 and that of the model before an after the educational expansion. The objective of this exercise is to estimate how much of the changes in the data can be predicted by the model when *only* the educational level of the workforce increases as it did in Brazil while all other parameters (technology) are

[38] Section 5 describe two possible scenarios when both m and h increases, depending on whether I_H increases or decreases. Given that in the quantitative exercise I_H declines and that the sufficient condition for I_H to decline ($\Delta \frac{h}{m} < 0$) is not meet in the data, only that case is relevant here.

held constant. The last column of the table shows the percentage in which the model fits the changes data for each case.

I found that the model's predictions of the effect of an educational expansion are remarkably accurate. The model is not only able to predict the qualitative changes but it also makes a close quantitative prediction of the observed changes in most of the outcomes of interest.

Changes in the average ranking of occupations in the model follow closely that of the data: there was lower educational attainment for all educational groups, and it was larger for medium educated workers. For medium educated workers, the average occupational ranking declines by 0.126 in the model compared to 0.128 in the data. For high educated workers, the average ranking declines by 0.042 in the model and by 0.057 in the data. The model predicts a somehow larger decrease in the average task of low educated workers (0.083) than the one observed in the data (0.037). Part of the reason is that the model cannot explain the decline in the share of employment in agriculture that is due to some structural change taking place during these years, which decreased the share of low educated workers in agricultural-related occupations that are at the bottom of the ranking.[39] Overall, the model explains 66 percent of the decline in educational attainment within each educational group that took place during this period.

With respect to the overall composition of employment, the general prediction of the model is that changes in the occupational composition of the workforce are smooth in comparison with the large educational expansion that took place. The model predicts larger changes than the ones in the data, but the same qualitative results (a decline in bottom-third and an increase in the top third occupations). When the educational expansion takes place, high and medium educated workers moved down to tasks of lower complexity, implying that, not all the new high and medium educated workers find a job in the same tasks they used to carry out before, smoothing the changes in the overall occupational composition of employment. The model predicts, on average, around 40 percent of the changes in the overall composition of employment.

In the case of real wages, the model predicts, on average, 91.1 percent of the changes overall and for each educational group. Average wages rise 29.3 percent in the model and 27.8 percent the data; wages of low educated workers increase 46.9 percent in the model compare to 37.3 percent in the data; wages of medium educated workers decline 10.9 percent in the model and 10.5 percent in the data; and wages of high educated workers decrease 22.2 percent in the model and by 21.6 percent in the data. The model predicts that wages of medium educated workers decrease less than wages of high educated workers, which is precisely what we observe in the data. Obviously, the changes in the wage gaps follow directly from the changes in wages, which the model predicts in the same precise way (92.5 percent on average). Wage gaps of medium to low falls 39.4 percent in the model and 34.8 percent in the data. Similarly, wage gaps of high to

[39]Structural change may also be driven, in part, by an educational expansion. Consider a model with two sectors, agriculture and non-agriculture. In the agriculture sector labor productivity declines with the number of workers in that sector. The non-agricultural sector produces a final good combining an infinite number of task, as in the model presented here. Low educated workers can be employed in the agriculture sector or in tasks of low complexity in the non-agricultural sector. An educational expansion as the one that happened in Brazil may increase the wages of low educated workers in tasks of lower complexity in the non-agricultural sector, pulling workers out of agriculture. To build this model is part of my research agenda.

medium declines 12.7 and 12.4 in the model and the data respectively. Finally, the wage gaps of high to low declines 47.0 percent in the model and 42.9 percent in the data.

Welfare changes in the model are somehow distinct from those in the data. Figure 10 shows the growth incidence curve of the wage distribution for Brazil between 1995 and 2014. Although the GIC is above zero for all the percentiles in the data, in the model wages decline for some percentiles at the top-end of the wage distribution. This is the case of the percentiles that originally where high educated and for which wages decreased. It is also the case of some percentiles that remained medium educated, for which wages decline as well. Despite these differences, the model predicts similar magnitudes of changes in wages for the first 70 percentiles. One reason for this discrepancy is that there is a larger dispersion in wages among high educated workers, so when more people are added to that category the percentage of the populations that earns higher wages does not necessary diminishes even when the mean of the wage distribution of high educated workers declines.

Changes in inequality and poverty in the model are also close to that of the data. In the case of inequality, Figure 11 displays the Lorenz curves for Brazil in the data (Panel A) and in the model (Panel B). The Lorenz curve in 2014 is always above the one in 1995, which implies that any index of relative inequality decreased during this period. The same pattern is observed in the model. In terms of specific indexes of inequality, the Gini coefficient declined 0.085 points in the data and 0.082 in the model. The reduction in the Theil index is also similar, but the between education groups (that may be more comparable to the changes in the model) is overestimated. This is not the case in the Atkinson index with the inequality aversion parameter equal to 2, which puts more weight to changes at the bottom of the wages distribution. In this case, the between inequality declined 0.078 in the data while in the model it decreased 0.088.[40] The fact that the fit of the model is better for the changes in the Atkinson (2) than in the Theil index is consistent with the model not being able to capture some of the movement at the top of the income distribution related to the dispersion of wages among high educated workers. Taking the average fit of the model over the changes in the indexes of inequality, the model predicts 85 percent of the observed changes.[41]

The generalized increase in wages at the bottom of the wage distribution reduced poverty in the data and the model. Wages for the percentiles that remained low educated increased and wages rose even more for those percentiles that switched from low to medium and high education. This represents the bottom 70 percent of the wage distribution in the model (the percentage that originally was low educated). However, the changes in the poverty rate in the model are overestimated given its sensitivity to different poverty lines when it moves above or below the wage level of any group. The model better predicts changes in the poverty gap—FGT(1)—and the depth of poverty—FGT(2). On average, the model fits 68.4 percent of the changes in wage poverty in Brazil measure by FGT(0)-FGT(2).

[40]In the model there is no within groups inequality, therefore total inequality is equal to between-group inequality.

[41]If only the between components and the Gini index is taking into account, the fit of the model diminishes slightly to 68.4.

Although the model accurately predicts most of the patterns in the occupational structure of employment and the wage distribution observed in Brazil between 1995 and 2014, it has some limitations to predict the year by year changes in a satisfactory way. Figure 12 shows the evolution of key variables and the predictions from the model. Brazil had a financial crisis in 1999 that lasted until 2003. The average wages during this period dropped overall and for the three educational groups. The model does not contain any structure to analyze business cycles so it is expected to perform poorly during this period. However, the crises mainly affected the levels of real wages, but it can be argued that the wage ratios where less affected if the crisis equally hurt workers with any educational level. This seems to be the case because even during this period the model closely predicts the evolution of wage gaps.

In summary, the relatively simple assignment model developed and calibrated in this paper allows isolating the effects of an increase in the education level of the workforce on occupations and the wage distribution when all the other factors, such as technology and educational quality, are held constant. This exercise resulted in estimations that follow very closely the observed changes in the data, providing evidence that the educational expansion in Brazil was the main factors behind the changes in the labor market between 1995 and 2014.

7.2 Output changes and the declining effects of an educational expansion

An expansion in education has distributional effects by changing the wages and shares of workers with different educational levels. It also affects the amount of output produced because more educated workers are more productive in the tasks that they start to perform after the educational expansion. In this section, I evaluate whether the model is informative with respect to changes in output per worker that took place in Brazil. Then, I evaluate different counterfactuals to study the differences between an increase in education in 1995 and in 2014.

The trend in output per worker is correctly captured by the model. Figure 13 shows the evolution of output per worker in Brazil and in the calibrated model.[42]. Output per worker in Brazil increased 18 percent between 1995 and 2014 and 15 percent in the model, accounting for 82 percent of its changes.[43] This result suggests that the large investment in education in Brazil resulted in a considerably higher output.[44]

Despite the large effect of education on output during the period 1995-2014, I find that further increases in education are predicted to have a much lower impact. I estimate the effects of an increase in 1 percentage point in the share of medium or high educated workers in 1995 and 2014. The results are in table 8. I found that an increase of 1 percentage point in the share of medium educated in 1995 raised output 0.1 percent and average wages in 0.5 percent. However, the same increase in 2014 slightly declines output and increases average wages by only

[42]Output per workers is measured as the coefficient of gross domestic product and the number of employed workers. The statistics are taken from World Development Indicators (WDI)

[43]Changes in average productivity in the model can be estimated from (1).

[44]My estimation may underestimate the real impact in economy-wide productivity if a more abundant supply of high educated workers might lead to skill-based technological change (Acemoglu 2002), or increases in human capital might generate positive externalities in terms of productivity that are not taken into account here, such as reduction of crime and better democratic institutions in the long term (Moretti 2004*a*).

0.1 percent, five times lower than in 1995.[45] In the case of an increase in 1 percent point of high educated workers, output and average wages increase by 1.8 and 2.3 percent, respectively, in 1995. The same increase in high education in 2014 leads to a rise in output of 0.7 percent and in the average wage of 1.1 percent. These are still sizable effects, but they account for between half and one-third of the impacts in 1995.[46]

The previous result implies that further increases in secondary schooling might not continue to foster increases in output and average wages. This result must interpret with caution for several reasons. First, in order to increase h without reducing m, as implied in the last two columns in Table 8, the share of workers with at least secondary education has to increase as fast as h (otherwise there would be a decline in m). Second, it does not necessary means that the government should stop investing in secondary schooling. Instead, It shows that *only* investing in secondary schooling will have a small impact on output, given the restrictions on the demand side by holding the technology constant. The fact that there is rapid decline to how much an extra percentage of medium or high educated workers can contribute to output if the demand side does not change implies that policy makers have to focus on both. Finally, the effects on wages or output estimated here do not have to be interpreted as social returns to education. It is a well know fact that education generates several positive externalities that might be difficult to quantify, including (but not limited to): decrease in crime, positive effects in health, and better family planning.[47]

7.3 Robustness Check

The comparative advantages across tasks are crucial for estimating the labor market effect of an educational expansion, as discussed in Section 5.5. Therefore, the results presented above are sensitive to the calibration procedure. In this section, I evaluate different alternatives to calibrate the parameters of interest. I show that the results are robust to using these alternatives. First, I show the results change little when other years are used to calibrate the model instead of 1995 data. Secondly, I discuss whether the linear assumption for $\alpha_L(i)$, $\alpha_M(i)$ and $\alpha_H(i)$ is reasonable according to the patterns observed in the data.

In principle, any year could be used to calibrate the parameters A_L, A_M and A_H as in Section 6. Figure 14 shows the value of the productivity parameters with respect to 1995, when parameters are estimated for every year separately, using data on wages and skill supplies from each year. All productivity parameters decreased around 25 percent between 1995 and 2004,

[45]The decrease in output takes place because in the calibrated model medium educated workers are starting to perform tasks for which low educated workers have an absolute advantage. In reality, it could be the case that after a certain task $\hat{i}$, low and medium educated workers are equally productive, and output will no longer change with an educational expansion when medium educated workers start to perform tasks $i < \hat{i}$.

[46]The changes in real wages for each group and in relative wages are similar in 1995 than in 2014. But the share of people affected by these changes is very different. For example, an increase in m in 2014 will decrease wages of 2/3 of the workforce while it only decreased wages on 1/3 in 1995. The educational composition of the workforce in the baseline year matters when estimating the effect of an educational expansion on output and average wages, even when the effects in the wage gaps are very similar in different years (there is a movement along the curves of comparative advantage in Figure 6).

[47]See Moretti (2004*b*) for a detail discussion on human capital externalities.

and increase thereafter reaching similar levels in 2014 with respect to 1995. These patterns are consistent with the precise predictions of the model for wage levels in 2014—assuming that productivity was constant between 1995 and 2014—and the lack of precision for the years in between.The model seems to be better suitable to explain changes over an extensive period of time because it lacks the flexibility to accommodate short-term effects related to the business cycle, e.g. financial crises. This is an important limitation of the model given that business cycles in Latin America (including Brazil) tend to be more pronounced than in other regions of the world (Mejia-Reyes 1999). However, it is worth noticing that because all the productivity parameters tend to move in the same direction and extent, the ratios A_M/A_L and A_H/A_M do not change much during the analyzed period. These ratios determine the assignment of workers to tasks in the model, as well as the relative wages. This is the reason why the model does a better job at predicting the trends in relative wages than in real wages throughout the analyzed period.

To evaluate the sensibility of the results to different values of the parameters A_L, A_M and A_H, I compare the results for four different alternatives: 1) the base calibration from 1995; 2) the mean value of the series 1995-2014; 3) the minimum value of the series 1995-2014; and 4) the maximum value of the series 1995-2014. The results are shown in Table 9. From looking across the columns of the table, I find that the initial and final thresholds estimated in the model do not change much under the different calibrations. Note that the wages of 1995 are no longer targeted under estimations 2-4. More importantly, the predictions of proportional changes in real wages, relative wages, occupations within each educational group and overall distribution of occupations are robust to the different calibration exercises. What matters for estimating these effects are the ratios A_M/A_L and A_H/A_M, which are similar across all specifications.

The second assumption that I discuss here is whether the linearity for $\alpha_L(i)$, $\alpha_M(i)$ and $\alpha_H(i)$ is consistent with the data when the assumption of no change in productivity holds. Yearly data on wage gaps and skill supplies allows estimating year-by-year changes on thresholds levels according to Section 6. Let c_J be equal to $exp(C_J)$. Then, It is possible to estimate $c_{HM}(I_H)$ and $c_{ML}(I_L)$ for the different thresholds and to check whether these points lie on a linear curve. We can then define:

$$c_{HM}(I_{Ht}) = \frac{A_H}{A_M}\frac{\alpha_H(I_{Ht})}{\alpha_M(I_{Ht})}, \quad c_{ML}(I_{Lt}) = \frac{A_M}{A_L}\frac{\alpha_M(I_{Lt})}{\alpha_L(I_{Lt})}, \; for\, t = \{1995, ..., 2014\}.$$

The identifying assumption in Section 6 was $\frac{\alpha_H(I_{Ht})}{\alpha_M(I_{Ht})} = \frac{\alpha_M(I_{Lt})}{\alpha_L(I_{Lt})} = i$. If this assumption is consistent with the data, the points estimated by the model should lie close the curve with no constant and with the slopes estimated in Section 6. Figure 15 shows the results of this exercise. In the case of the curve $c_{ML}(i)$, the points lie very close to the estimated curve. However, in the case of $c_{HM}(i)$ it seems that the slope is larger than the one predicted by the model, and it goes mostly below the points estimated year by year. There are two reasons why the curve may not fit very well the points in the data. First, the assumption of the function $\alpha_H(I_H)/\alpha_M(I_H) = i$ is accurate, but A_H may also be changing. Second, the function $\alpha_H(I_H)/\alpha_M(I_H)$ is misspecified and should be approximated by a linear function with a constant ($c+bi$). I checked how sensible the results are to the linear function being misspecified. To that end, I considered each of the yearly estimated points as an observation in an OLS regression. The result is a linear function that minimizes the sum of the quadratic distance of the points to the linear function. This exercise provides an alternative estimation for $c_{HM}(i)$ and $c_{ML}(i)$ based on the data of the entire period. Figure 16 shows the predictions of the model in some key variables when this calibration is used. The results do not significantly differ from those in Figure 12.

The main take away from this exercise is that the results are robust to using different functional forms for the productivity schedules across tasks that are also consistent with the data. A reason for this is that in Brazil the threshold I_H did not diminish much since both m and h increased during this period.[48] Therefore, moderate changes in the functional form of the curve $C_{HM}(i)$ have a small impact on the results.

Although the specific value of some models' prediction changes when using alternatives years to calibrate the model or different curves that may fit the data better, the same general conclusion holds: the model matches very closely the changes in the occupational structure of employment and in the wage distribution.

8 Conclusions

The educational level of the workforce is increasing rapidly in most countries. This paper developed a novel theoretical and empirical framework to study the effects of an educational expansion on the occupational structure of employment and the wage distribution. I applied the framework to Brazil, a country that underwent a major educational expansion between 1995 and 2014. I proceed in three steps. First, I showed the interlinks between education-occupations-and wages in Brazil. Second, I developed a theoretical model that traces changes in occupations and wages to different educational expansions. Third, I calibrated the model to quantitatively assess how much of the observed changes in the data can be explained by the educational expansion.

In the first part of my paper, I found that along with an educational expansion, the following patterns are present in Brazil: the occupational structure of employment was relatively fixed, there was lower occupational attainment for each educational group, wages for workers with primary education increased, wages of more educated workers declined, wage gaps and other indexes of inequality fell, and poverty diminished. In the second part, I use an assignment model were workers with three levels of education are sorted across a continuum of occupations that vary in complexity and are combined to produce a final good. I investigated the effects of different policy experiment in the model, showing that it predicts qualitatively all the observed labor market changes in the occupational structure of employment and the wage distribution in Brazil when an educational expansion as the one observed in the data takes place. Next, I calibrated the model using the data from 1995 to find that the educational expansion quantitatively

[48] It went from 0.66 to 0.58, compared to a decline twice as big in I_L from 0.42 to 0.26.

predicts most of the observed changes in Brazil. I conclude that the increase in educational attainment was of utmost importance to the changes in the Brazilian labor market in the last two decades. Finally, I also found that further educational expansions in Brazil are predicted to have a much lower impact than in the past if they are not accompanied by other reforms aimed to change the technology of the economy.

The framework presented here can be used to study the labor markets effects of an educational expansion in other countries as well. It can also be used to evaluate ex-ante the general equilibrium effects of a particular policy aimed to increase the educational level of one particular group of the population. Finally, it can be generalized to more groups, and it can be used to evaluate the general equilibrium effects of other supply shocks, such as an increase in the share of immigrants with a specific level of human capital (low or high).

My results reassert educational expansions as a key driver of economic development. In the case of Brazil, it largely improved the income distribution by reducing poverty and inequality, and it raised the total output of the economy. But educational expansions should be accompanied by policies directed to increase the job opportunities in occupations where more educated workers can exploit the productivity differentials that they have obtained through education. Otherwise, more educated workers will end up in occupations where schooling added adds little value to workers' their productivity, undermining their additional effect on economic development.

References

Acemoglu, D. (1998), 'Why do new technologies complement skills? directed technical change and wage inequality', *The Quarterly Journal of Economics* **113**(4), 1055–1089.

Acemoglu, D. (2002), 'Directed technical change', *The Review of Economic Studies* **69**(4), 781–809.

Acemoglu, D. & Angrist, J. (2000), 'How large are human-capital externalities? evidence from compulsory schooling laws', *NBER macroeconomics annual* **15**, 9–59.

Acemoglu, D. & Autor, D. (2011), *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Vol. 4, Elsevier B.V.

Acemoglu, D. & Zilbotti, F. (1999), Productivity differences, Working Paper 6879, National Bureau of Economic Research.

Adão, R. (2015), 'Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil', *Mimeo* .

Alvarez, J. (2017), 'The agricultural wage gap: Evidence from brazilian micro-data'.

Alvarez, J., Benguria, F., Engbom, N. & Moser, C. (2017), Firms and the decline in earnings inequality in brazil, Technical report, Columbia Business School Research Paper No. 17-47.

Autor, D. (2013), 'The" task approach" to labor markets: an overview', *Journal for Labour Market Research* **46**(3), 185–199.

Autor, D. (2014), 'Skills, education, and the rise of earnings inequality among the"other 99 percent"', *Science* **344**, 843–851.

Autor, D. & Dorn, D. (2013), 'The growth of low-skill service jobs and the polarization of the us labor market', *The American Economic Review* **103**(5), 1553–1597.

Barro, R. J. & Lee, J. W. (2013), 'A new data set of educational attainment in the world, 1950–2010', *Journal of development economics* **104**, 184–198.

Barros, R., Mirela de Carvalho, S. F. & Mendonca, R. (2010), 'Markets, the state and the dynamics of inequality: The case of brazil'.

Beaudry, P., Green, D. A. & Sand, B. M. (2016), 'The great reversal in the demand for skill and cognitive tasks', *Journal of Labor Economics* **34**(S1), S199–S247.

Becker, G. S. (1994), Human capital revisited, *in* 'Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)', The University of Chicago Press, pp. 15–28.

Bekman, E., Bound, J. & Machin, S. (1998), 'Implications of skill-biased technological change: international evidence', *The quarterly journal of economics* **113**(4), 1245–1279.

Blinder, A. S. (1973), 'Wage discrimination: reduced form and structural estimates', *Journal of Human resources* pp. 436–455.

Bowlus, A. J. & Robinson, C. (2012), 'Human capital prices, productivity, and growth', *The American Economic Review* **102**(7), 3483–3515.

Bruns, B., Evans, D. & Luque, J. (2011), *Achieving World-Class Education in Brazil: The Next Agenda*, World Bank Publications.

Burstein, A., Morales, E. & Vogel, J. (2016), Changes in between-group inequality: computers, occupations, and international trade, Technical report.

Costinot, A. & Vogel, J. (2010), 'Matching and inequality in the world economy', *Journal of Political Economy* **118**(4), 747–786.

Cruces, G., Fields, G. S., Jaume, D. & Viollaz, M. (2017), *Growth, Employment, and Poverty in Latin America*, Oxford University Press.

Deming, D. J. (2017), 'The growing importance of social skills in the labor market', *The Quarterly Journal of Economics* pp. 1–48.

DiNardo, J., Fortin, N. M. & Lemieux, T. (1996), 'Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach', *Econometrica* **64**(5), 1001–1044.

Dix-Carneiro, R. & Kovak, B. K. (2015), 'Trade liberalization and the skill premium: A local labor markets approach', *American Economic Review* **105**(5), 551–57.

Engbom, N. & Moser, C. (2017), 'Earnings inequality and the minimum wage: Evidence from brazil', *CESifo Working Paper Series No. 6393* .

Ferreira, F. H., Firpo, S. P. & Messina, J. (2016), *Understanding Recent Dynamics of Earnings Inequality in Brazil*, Oxford University Press.

Fields, G. S. (1974), 'The private demand for education in relation to labour market conditions in less-developed countries', *The Economic Journal* **84**(336), 906–925.

Fields, G. S. (1995), *Educational Expansion and Labor Markets*, Pergamon.

Gasparini, L., Galiani, S., Cruces, G. & Acosta, P. (2011), 'Educational upgrading and returns to skills in latin america', *Documentos de Trabajo del CEDLAS* .

Goldin, C. D. & Katz, L. F. (2009), *The race between education and technology*, Harvard University Press.

Goos, M., Manning, A. & Salomons, A. (2014), 'Explaining job polarization: Routine-biased technological change and offshoring', *The American Economic Review* **104**(8), 2509–2526.

Hanushek, E. A. & Woessmann, L. (2008), 'The role of cognitive skills in economic development', *Journal of economic literature* **46**(3), 607–668.

Hanushek, E. A. & Woessmann, L. (2012), 'Do better schools lead to more growth? cognitive skills, economic outcomes, and causation', *Journal of economic growth* **17**(4), 267–321.

Harris, J. R. & Todaro, M. P. (1970), 'Migration, unemployment and development: a two-sector analysis', *The American economic review* **60**(1), 126–142.

Heckman, J. J., Lochner, L. & Taber, C. (1998), 'Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents', *Review of economic dynamics* **1**(1), 1–58.

Katz, L. F., Kearney, M. S. et al. (2006), 'The polarization of the us labor market', *American Economic Review* **96**(2), 189–194.

Katz, L. F. & Murphy, K. M. (1992), 'Changes in relative wages, 1963–1987: supply and demand factors', *The quarterly journal of economics* **107**(1), 35–78.

Khanna, G. (2015), 'Large-scale education reform in general equilibrium: Regression discontinuity evidence from india', *unpublished paper, University of Michigan* .

Leuven, E. & Oosterbeek, H. (2011), 'Overeducation and mismatch in the labor market1', *Handbook of the Economics of Education* **4**, 283.

Lewis, W. A. (1954), 'Economic development with unlimited supplies of labour', *The manchester school* **22**(2), 139–191.

López-Calva, L. F., Lustig, N., Valderrama, D. et al. (2016), Understanding the dynamics of labor income inequality in latin america, Technical Report No. 1608, Tulane University, Department of Economics.

Meghir, C., Narita, R. & Robin, J.-M. (2015), 'Wages and informality in developing countries', *American Economic Review* **105**(4), 1509–46.

Mejia-Reyes, P. (1999), 'Classical business cycles in latin america: turning points, asymmetries and international synchronisation', *Estudios Económicos* pp. 265–297.

Moretti, E. (2004*a*), 'Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data', *Journal of econometrics* **121**(1), 175–212.

Moretti, E. (2004*b*), 'Human capital externalities in cities', *Handbook of regional and urban economics* **4**, 2243–2291.

Oaxaca, R. (1973), 'Male-female wage differentials in urban labor markets', *International economic review* pp. 693–709.

Rauch, J. E. (1993), 'Productivity gains from geographic concentration of human capital: evidence from the cities', *Journal of urban economics* **34**(3), 380–400.

Salardi, P. (2014), 'The evolution of gender and racial occupational segregation across formal and non-formal labor markets in brazil, 1987 to 2006', *Review of Income and Wealth* pp. 1–22.

Saposnik, R. (1981), 'Rank-dominance in income distributions', *Public Choice* **36**(1), 147–151.

Sattinger, M. (1993), 'Assignment models of the distribution of earnings', *Journal of economic literature* **31**(2), 831–880.

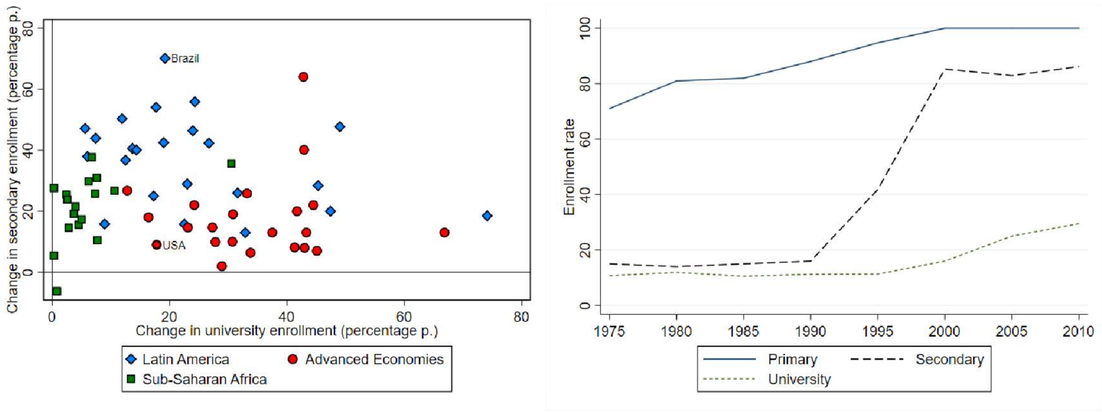
Sattinger, M. (2012), 'Qualitative mismatches', *Foundations and Trends in Microeconomics* **8**(1-2), 1–168.

Teulings, C. N. (2005), 'Comparative advantage, relative wages, and the accumulation of human capital', *Journal of Political Economy* **113**(2), 425–461.

Ulysseas, G. (2014), 'Firms, informality and development: Theory and evidence from brazil'.

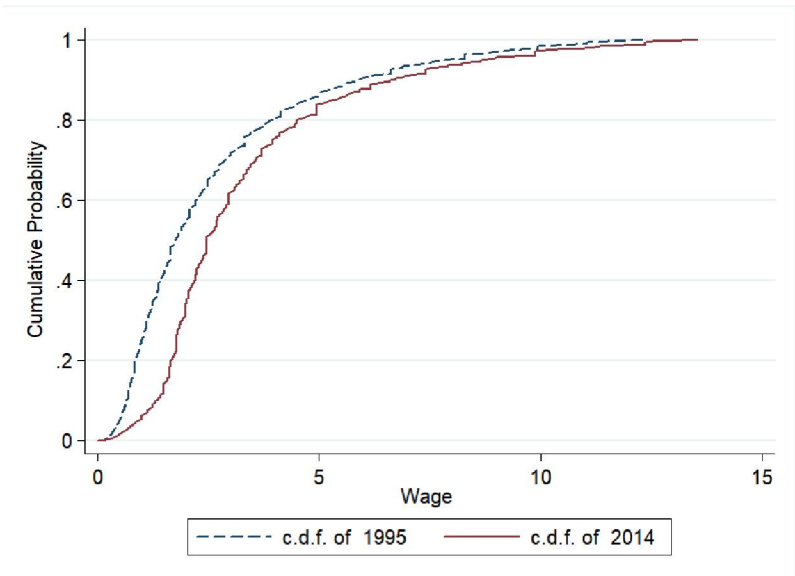
Figure 1: Changes in enrollment rates in Brazil

(a) Across countries comparison (1990-2010) (b) Historical comparison



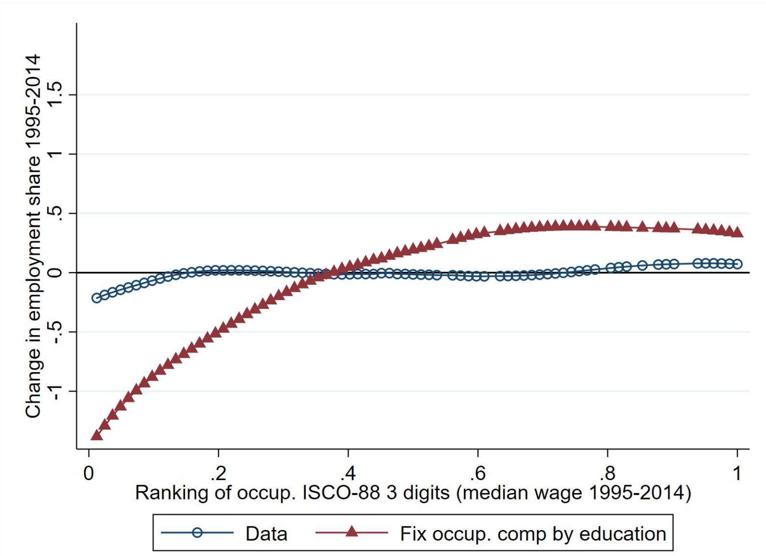
Notes: Panel (a) shows the changes in enrollment rates across countries between 1990 and 2010 in secondary and university/tertiary education. Panel (b) shows the historical evolution of enrollment rates for primary, secondary and university education in Brazil. Enrollment ratios are defined as the ratio of students at a given level of schooling in the designated age group to the total population of that age group.
Source: Barro & Lee (2013).

Figure 2: Cumulative Density function of the wage distribution in 1995 and 2014



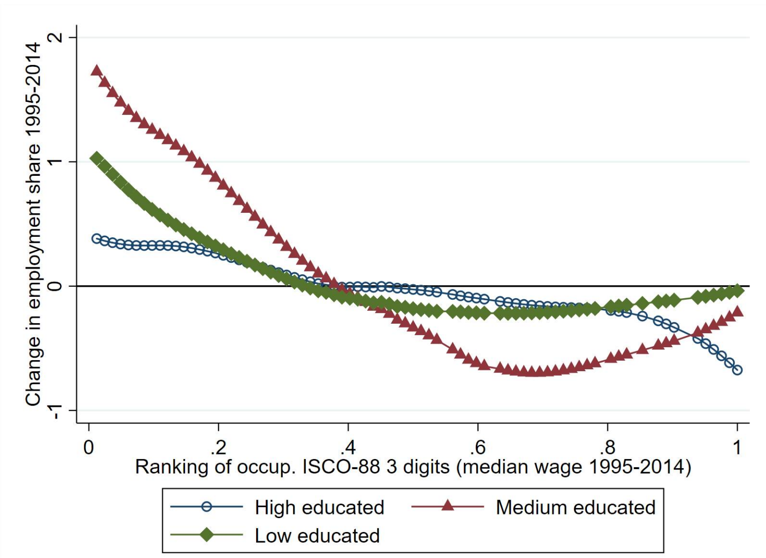
Notes: The figure shows the cumulative density function of real wages expressed in dollars at 2005 PPP in Brazil for the years 1995 and 2014.

Figure 3: Changes in the occupational composition of the workforce 1995-2004



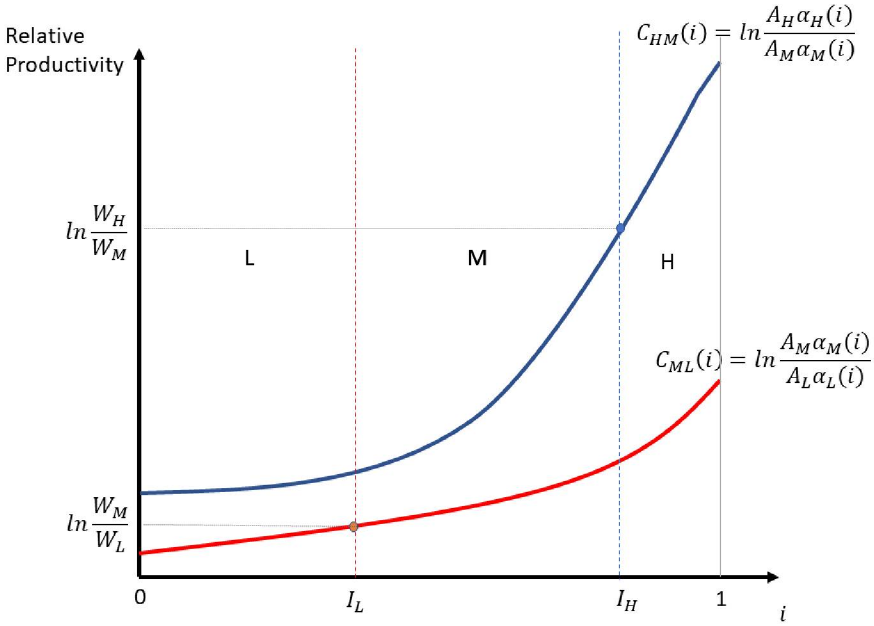
Notes: Occupations at 3 digit level from ISCO-88 are ranked according to their median wage from lower to higher in the horizontal axis. The blue line (circle) in the figure is a locally weighted smoothing regression of the changes in employment shares (percent points) across the occupational ranking based on the median wage of each occupation over the period 1995-2014. The red line (triangle) displays a counterfactual change in the occupational composition of employment between 1995 and 2014 if the occupational composition within each educational group is fixed to the levels of 1995.

Figure 4: Occupational downgrading for each educational group between 1995-2014



Notes: The figure plots a locally weighted smoothing regression of the changes in employment shares (percent points) across the occupational ranking based on the median wage of each occupation over the period 1995-2014. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university). Occupations are ranked as in Figure 3.

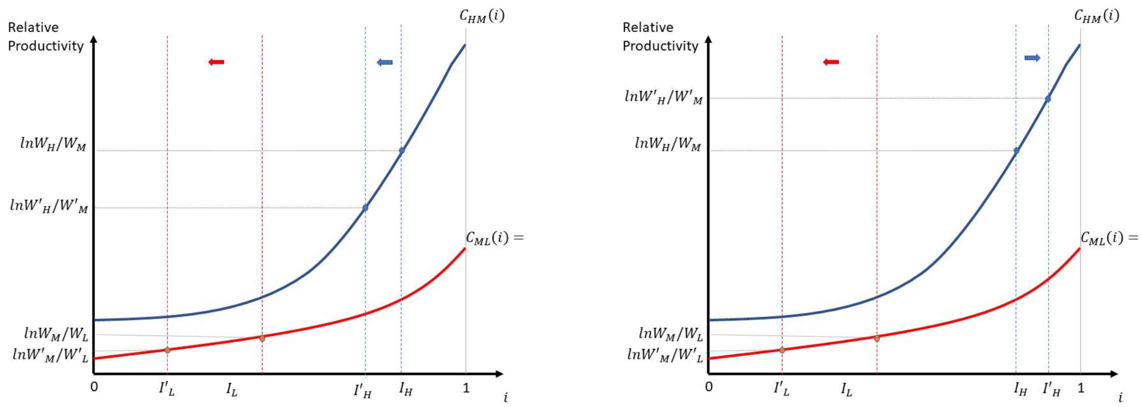
Figure 5: Equilibrium in the model



Notes: The figure shows an infinite number of tasks in the horizontal axis indexed by its level of complexity from 0 (the less complex tasks) to 1 (the most complex tasks) and the relative productivity in the vertical axis. The function $C_{HM}(i)$ represents the comparative advantage of high with respect to medium educated workers at task i and $C_{ML}(i)$ is the comparative advantage of medium relative to low educated workers at task i, both are increasing in i by assumption. The threshold levels I_L and I_H determine the task performed by low (L), medium (M) and high educated (H). In equilibrium, $C_{HM}(I_H)$ ($C_{ML}(I_L)$) equals relative wages if high and medium educated workers (medium and low educated workers) otherwise employers can profit from switching workers across occupations.

Figure 6: Effects of an educational expansion

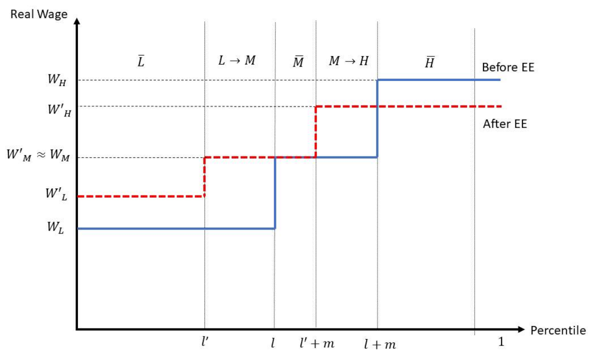
(a) Panel A: Policy experiment (i) (Δh) (b) Panel B: Policy experiment (ii) (Δm)



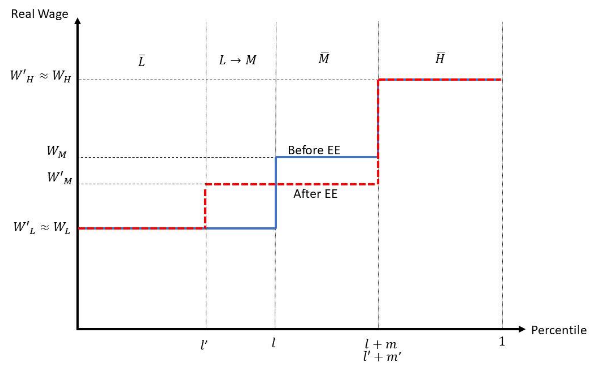
Notes: The figure shows the theoretical effects of an educational expansion on the assignment of workers' type to task and relative wages. Panel A shows to the effect of policy experiment (i) consisting of an increase in the share of high educated workers and a corresponding decrease in low educated workers. Panel B displays the effects of policy experiment (ii) increasing the share of medium educated workers by a corresponding decline in low educated workers.

Figure 7: Cumulative density function (CDF) before and after each policy experiment

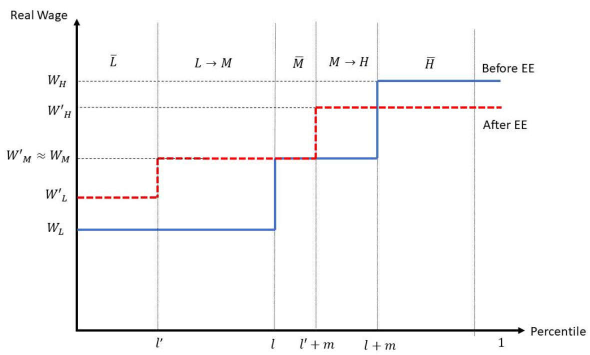
(a) Panel A: Policy experiment (i) $(\Delta h > 0)$



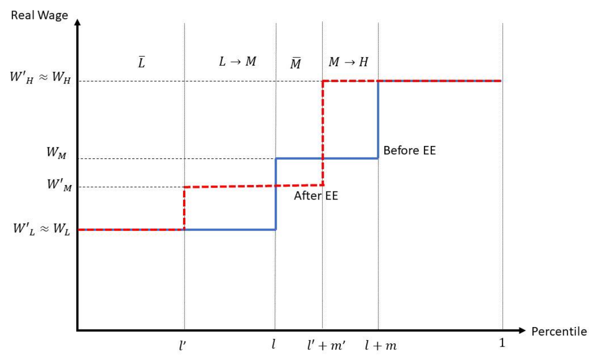
(b) Panel B: Policy experiment (ii) $(\Delta m > 0)$



(c) Panel C: Policy experiment $(iii.a)$ $(\Delta h$ dominates)



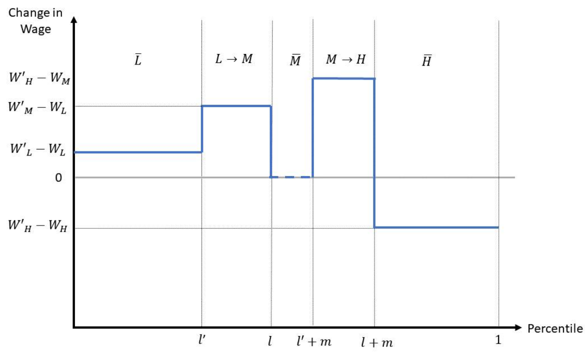
(d) Panel D: Policy experiment $(iii.b)$ $(\Delta m$ domintes)



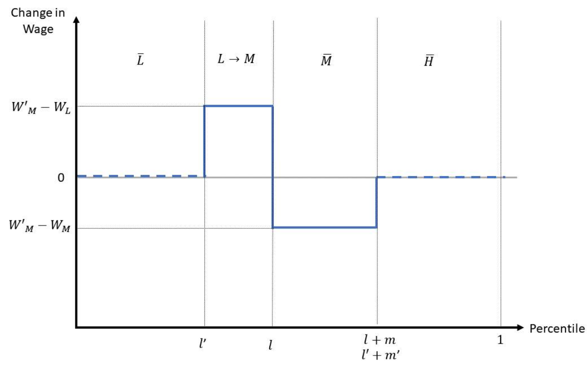
Notes: The figure shows the Cumulative Density Function (inverted) of the distribution of wages before and after the educational expansion under the different policy experiments.

Figure 8: Growth incidence curve of different type of educational expansions

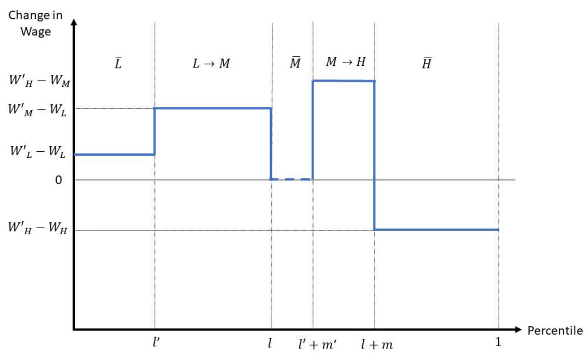
(a) Panel A: Policy experiment (i) $(\Delta h > 0)$



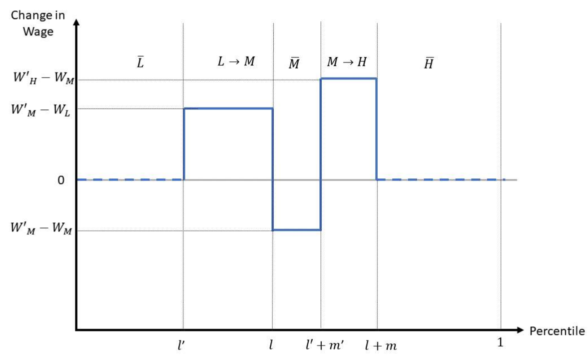
(b) Panel B: Policy experiment (ii) $(\Delta m > 0)$



(c) Panel C: Policy experiment $(iii.a)$ $(\Delta h$ dominates$)$

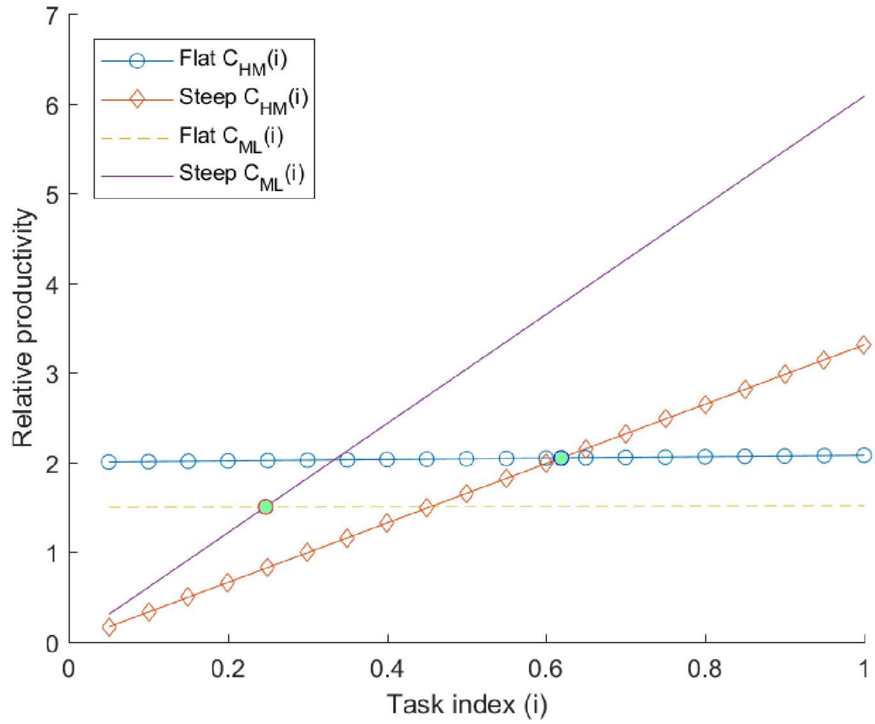


(d) Panel D: Policy experiment $(iii.b)$ $(\Delta m$ domintes$)$



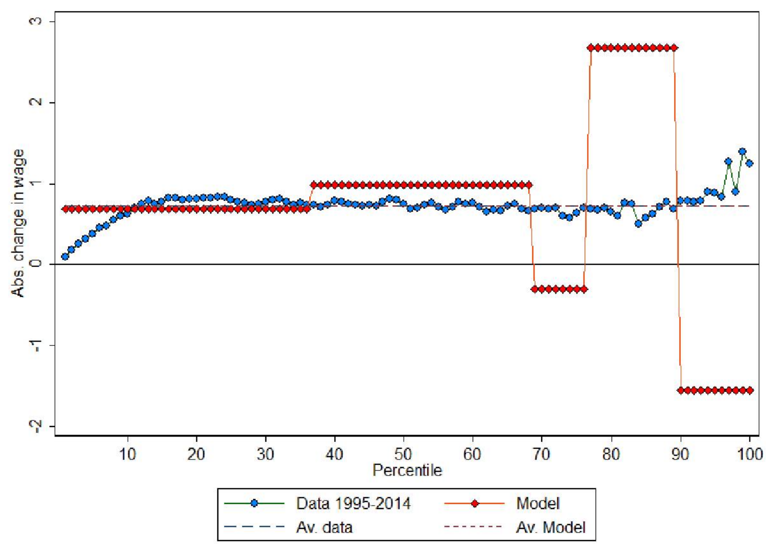
Notes: The figure shows the wage change (in absolute value) after different types of educational expansions at each percentile of the wage distribution. It is the difference between the curves *Before* and *After* from Figure 7.

Figure 9: Four different scenarios of comparative advantage schedules



Notes: The figure shows $C_{HM}(i)$ and $C_{ML}(i)$ under a flat or a steep scenario of increase in comparative advantage across tasks (i). Parameters are chosen such that initial I_H and I_L are the same scenario (depicted by a point at the intersection of the flat and steep comparative advantage).

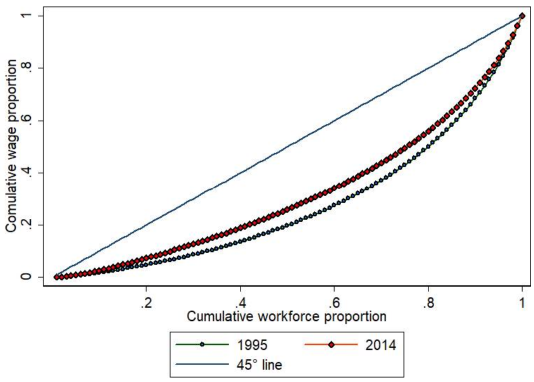
Figure 10: Growth incidence curve: Data and model



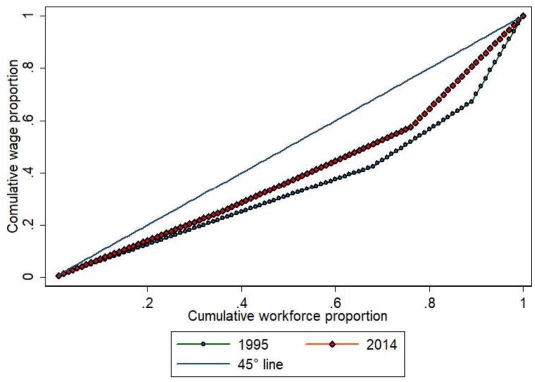
Notes: The figure shows absolute changes in wages for each percentile of the wage distribution. The blue line (circle) depicts the changes observed in the data between 1995 and 2014. The red line (diamond) shows the predicted changes in the model before and after the Brazilian educational expansion.

Figure 11: Changes in relative inequality: Lorenz Curves

(a) Brazil: 1995 and 2015

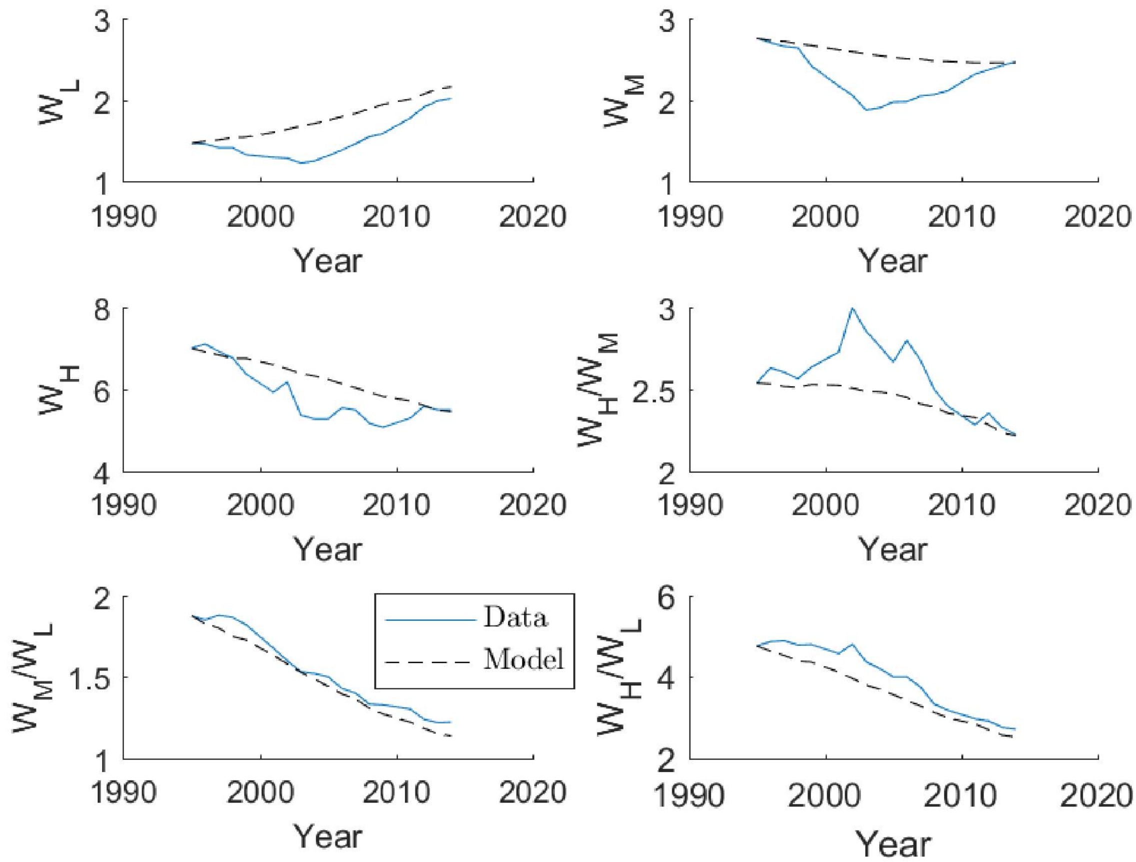


(b) Model: Before and after the educational expansion



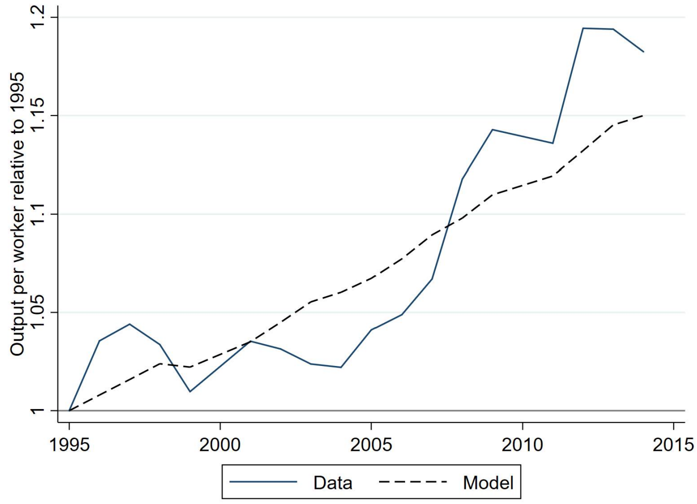
Notes: Panel (a) shows Lorenz curves for Brazil in 1995 and 2014. Panel (b) shows the Lorenz curves in the calibrated model before and after the educational expansion.

Figure 12: Data vs model predictions: wages year by year



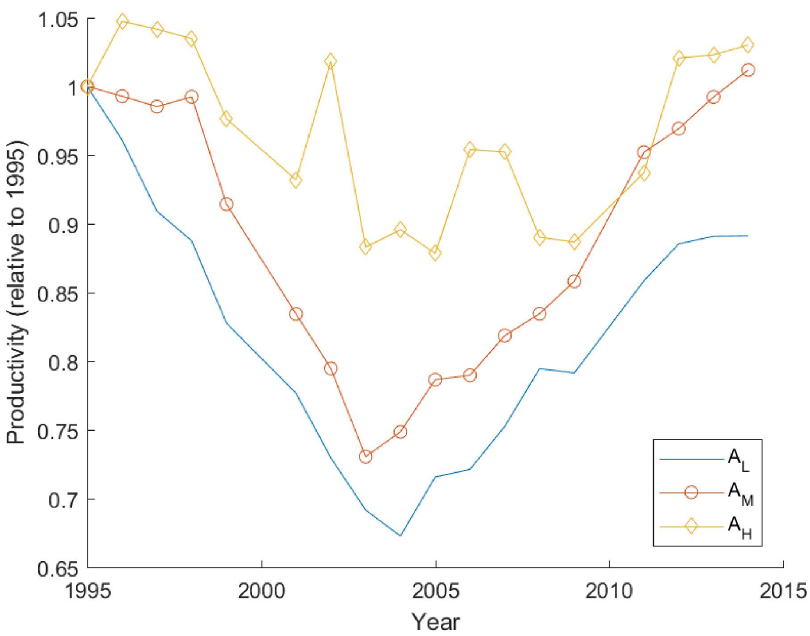
Notes: The figure shows the evolution of wages and wages gaps labor market outcomes in the data and in the model year by year. The model simulates expansions in education as the one that took place year by year in Brazil, holding all the other parameters in the model constant at the levels of 1995.

Figure 13: Evolution of output per worker in the data and in the model



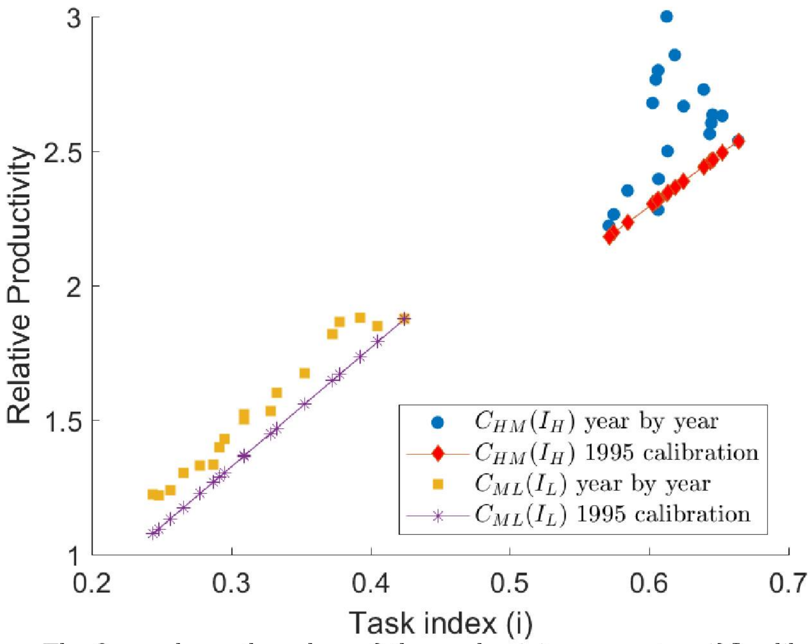
Notes: The figure shows the evolution of output per worker in the data for Brazil. Output per worker is defined as GDP over employ workers. obtained from the World Development Indicators (2017). Output per worker in the model is defined as the total amount of the final good that is produced according to equation (1) in the calibrated model and taking into account the yearly changes in skill supplies from the data.

Figure 14: Calibration of productivity parameters year by year



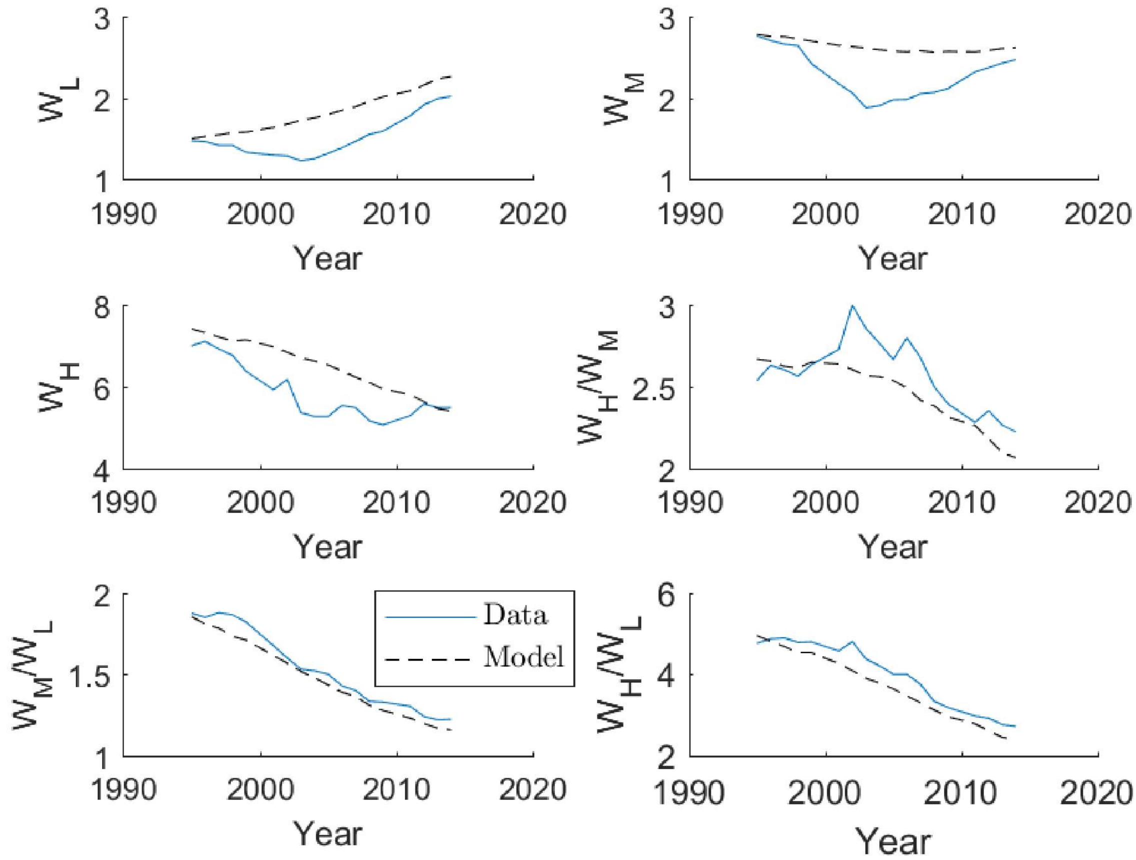
Notes: The figure shows the values of the productivity parameters if the calibration exercise is performed separately using year by year data on the share of workers with different educational level and their average wages.

Figure 15: Comparison of 1995 calibration vs year by year calibration



Notes: The figure shows the values of the productivity parameters if I calibrate the model using the average wages and educational levels of each year.

Figure 16: Data vs model predictions with alternative calibration : wages year by year



Notes: The figure shows the evolution of wages and wages gaps labor market outcomes in the data and in the model year by year as in Figure 12. In this case, the calibration of the comparative advantage parameters is obtained by the OLS regressions over the points depicted in Figure 14 instead of using the data on 1995.

Table 1: Changes in the educational attainment of the workforce and the Brazilian wage distribution between 19995 and 2014

	Brazil		
	1995	2014	Change
Educational attainment (share of the workforce)			
Low educated (less than secondary)	0.68	0.36	-0.32
Medium educated (some secondary)	0.21	0.40	0.20
High educated (some university)	0.11	0.24	0.12
Average Wages			
Total workers	2.36	3.02	27.8%
Low educated	1.47	2.02	37.3%
Medium educated	2.76	2.47	-10.5%
High educated	7.00	5.49	-21.6%
Wage gaps			
W_M/W_L	1.88	1.22	-34.8%
W_H/W_M	2.54	2.22	-12.4%
W_H/W_L	4.76	2.72	-42.9%
Inequality			
Gini	0.445	0.360	-0.085
Theil			
Total	0.327	0.219	-0.109
Between educ. groups	0.071	0.051	-0.020
Atkinson (Inequality aversion = 2)			
Total	0.572	0.440	-0.132
Between educ. groups	0.140	0.062	-0.078
Wage Poverty (wage line: 2.4)			
FGT(0)	0.626	0.462	-0.164
FGT(1)	0.305	0.144	-0.162
FGT(2)	0.183	0.065	-0.118

Notes: The table shows the levels and changes in the educational attainment of the workforce and in different outcomes of the wage distribution estimated from PNAD 1995 and 2014.

Table 2: Changes in the occupational structure of employment in Brazil between 19995 and 2014

	Average	**Total workforce**		**Low educated**		**Medium educated**		**High educated**	
	wage	**1995**	**1995-2014**	**1995**	**1995-2014**	**1995**	**1995-2014**	**1995**	**1995-2014**
Occupations ISCO-88 1 digit									
Proffessionals	6.3	5.1	5.3	0.6	0.2	3.9	-2.1	33.9	3.2
Legislators and managers	5.5	7.9	-2.8	4.4	-2.5	12.5	-8.8	20.4	-10.1
Military	4.5	2.1	-1.2	1.3	-1.2	4.5	-3.5	2.5	-0.7
Technicians	3.3	7.8	-0.2	3.4	-2.0	17.1	-8.9	17.3	-3.8
Clerks	2.4	8.2	3.6	2.9	0.1	21.9	-7.0	15.6	2.1
Plant and machine operators	2.2	8.1	-0.9	10.5	-1.0	4.3	4.4	0.7	1.3
Craft and related trades	2.0	20.5	-2.1	25.8	1.8	12.5	7.6	2.7	2.0
Service workers and salers	1.8	24.2	6.7	28.8	9.0	19.0	17.8	5.8	6.0
Agriculture and elementary	1.6	16.1	-8.2	22.2	-4.3	4.2	0.5	1.2	0.1
Total		**100.0**	0.0	**100.0**	**0.0**	**100.0**	**0.0**	**100.0**	**0.0**
Total share change			**15.5**		**11.1**		**30.2**		**14.7**
Occupations ISCO-88 3 digit									
Average ranking		0.335	0.007	0.240	-0.037	0.449	-0.129	0.670	-0.058
Median ranking		0.268	0.000	0.183	-0.085	0.439	-0.134	0.732	-0.037
Sd		0.264	0.001	0.211	-0.034	0.238	-0.021	0.233	0.025
Three categories ISCO-88 3 digit									
Share bottom-third		0.562	-0.011	0.718	0.070	0.305	0.249	0.087	0.089
Share middle-third		0.279	-0.005	0.216	-0.033	0.460	-0.110	0.335	-0.045
Share top-third		0.159	0.015	0.066	-0.037	0.235	-0.139	0.579	-0.044

Notes: Occupations ISCO-88 1 digit: Occupations are classified according to ISCO-88 at 1-digit level except for agriculture and elementary occupations that are shown together. Occupations ISCO-88 3-digit: I construct a ranking of occupations classified by ISCO-88 3 digit (a total of 82 occupations) based on their median wage over the period 1995-2014.Three categories ISCO-88 3 digit: 82 occupations are divided into 3 groups of an equal number of occupations according to their average wage. Bottom-third refers to the 27 occupations with lower average wage, medium-third are the next 27, and the remaining 28 are classified as top-third occupations. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university).

Table 3: Summary of the labor market effects of different educational expansion

	(i) $\Delta h > 0$	**(ii)** $\Delta m > 0$	**(iii)** $\Delta m > 0$ **and** $\Delta h > 0$	
			(iii.a) $\downarrow I_H$	**(iii.b)** $\uparrow I_H$
Thresholds				
I_L	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
I_H	$\downarrow$	$\uparrow$	$\downarrow^*$	$\uparrow^*$
Average task within each group				
Low educated	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Medium educated	$\downarrow$	$\uparrow\downarrow$	$\downarrow$	$\uparrow\downarrow$
High educated	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
Occupational composition (workforce)				
Bottom-level (E_B)	$\Delta E_B > -\Delta h$	$\Delta E_B > -\Delta m$	$\Delta E_B > -(\Delta m + \Delta h)$	$\Delta E_B > -(\Delta m + \Delta h)$
Medium-level (E_M)	$\downarrow$	$\Delta E_M < \Delta m$	$\Delta E_M < \Delta m$	$\Delta E_M < \Delta m$
Top-level (E_T)	$\Delta E_T < \Delta h$	$\uparrow$	$\Delta E_T < \Delta h$	$\Delta E_T > \Delta h$
Wages				
Low educated (W_L)	$\uparrow$	$\Delta W_L > \Delta W_M$	$\uparrow$	$\Delta W_L > \Delta W_M$
Medium educated (W_M)	$\Delta W_H < \Delta W_M < \Delta W_L$	$\downarrow$	$\Delta W_H < \Delta W_M < \Delta W_L$	$\downarrow$
High educated (W_H)	$\downarrow$	$\Delta W_H > \Delta W_M$	$\downarrow$	$\Delta W_H > \Delta W_M$
Relative Wages				
High/Medium	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
Medium/Low	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
High/Low	$\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$

Notes: The table shows the theoretical predictions of the model under four different policy experiments. Policy experiment (i) consists of an increase in high education. Policy experiment (ii) refers to increase in medium education. Policy experiment (iii) refers to an increase in both medium and high, and it is further separated between case (iii.a) when the threshold I_H declines, and case (iii.b) when the threshold I_H increases. The share of low educated workers is reduced by the same amount in all policy experiments. $\uparrow$ denotes an increase for any value in the parameters of the model; $\downarrow$ denotes a decrease; $\uparrow\downarrow$ denotes that it can increase or decrease. Other cells are filled with lower or upper bound to changes in the outcome variable. * Given by assumption.

Table 4: A numerical example on the differential effect of Δh for different scenarios of comparative advantage across tasks.

	Case 1	**Case 2**	**Case 3**	**Case 4**
Comparative adv. schedule				
CA_H	*Flat*	*Steep*	*Steep*	*Flat*
CA_L	*Flat*	*Steep*	*Flat*	*Steep*
Thresholds				
Initial I_L	0.304	0.304	0.304	0.304
Initial I_H	0.686	0.686	0.686	0.686
Final I_L	0.212	0.255	0.224	0.224
Final I_H	0.566	0.610	0.598	0.586
Δ mean occup. ranking				
Low educated	-0.046	-0.025	-0.040	-0.028
Medium educated	-0.107	-0.063	-0.084	-0.079
High Educated	-0.060	-0.038	-0.044	-0.050
Δ % in wages				
Low educated	0.2	16.3	5.1	16.2
Medium educated	0.2	-2.8	5.0	-5.4
High Educated	-0.3	-13.6	-8.4	-5.7

Notes: The table shows the theoretical predictions of the model under four different scenarios of comparative advantage across tasks. All scenarios start with the same thresholds I_L and I_H, and simulate the effect of an increase in 10 percentage points in the share of high educated (and a corresponding reduction in the share of low educated). Column one shows the scenario when both comparative advantage schedules are flat, that is, they barely increase with the complexity of the task denoted by i. Column two displays the scenario of steep comparative advantage schedules. Columns 3 and 4 show a combination of steep and flat comparative advantage schedules. Parameters are calibrated so that initial wages and thresholds are the same in each column.

Table 5: Initial moments in the data (1995) and the model

	Brazil 1995	**Model**
Thresholds		
Initial I_L		0.424
Initial I_H		0.664
Wages 1995 (targeted)		
Low educated	1.47	1.47
Medium educated	2.76	2.76
High Educated	7.00	7.00
Average Occp Ranking 1995 (not-targeted)		
Low educated	0.243	0.212
Medium educated	0.454	0.544
High Educated	0.669	0.832
Occup. Composition of Employment in 1995 (not-targeted)		
Bottom-third	0.562	0.529
Middle-third	0.279	0.359
Top-third	0.159	0.112
Inequality (not-targeted)		
Gini	0.445	0.271
Theil		
Total	0.327	0.179
Between educ. groups	0.071	
Atkinson (Inequality aversion = 2)		
Total	0.572	0.205
Between educ. groups	0.140	
Wage Poverty- wage line $2.4 (not-targeted)		
FGT(0)	0.626	0.709
FGT(1)	0.305	0.275
FGT(2)	0.183	0.106

Notes: Estimations for Brazil comes from PNAD 1995. The model is calibrated for the year 1995 according to Section 6. 82 occupations at ISCO-88 3 digit level are ranked from 0 to 1 according to its median wage for the period 1995-2014. The average ranking is computed for each educational group. Bottom-third refers to the 27 occupations with lower average wage, medium-third are next 27, and the remaining 28 are classified as top-third occupations. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university).

Table 6: Qualitative observed changes vs model predictions

	Brazil	**Model (iii.a)**
	1995-2014	$\Delta m > 0$ **and** $\Delta h > 0$
Thresholds		
I_L		$\downarrow$
I_H		$\downarrow$
Average task within each group		
Low	$\downarrow$	$\downarrow$
Medium	$\downarrow$	$\downarrow$
High	$\downarrow$	$\downarrow$
% Δ in overall occupations 1995-2014		
Bottom-third	$(\downarrow)\Delta O_B > -(\Delta m + \Delta h)$	$\Delta O_B > -(\Delta m + \Delta h)$
Middle-third	$(\downarrow)\ \Delta O_M < \Delta m$	$\Delta O_M < \Delta m$
Top-third	$(\uparrow)\Delta O_H < \Delta h$	$\Delta O_H < \Delta h$
Wages		
Low	$\uparrow$	$\uparrow$
Medium	$(\downarrow)\ \Delta W_H < \Delta W_M < \Delta W_L$	$\Delta W_H < \Delta W_M < \Delta W_L$
High	$\downarrow$	$\downarrow$
Relative Wages		
High/Medium	$\downarrow$	$\downarrow$
Medium/Low	$\downarrow$	$\downarrow$
High/Low	$\downarrow$	$\downarrow$

Notes: The table shows the qualitative changes in the labor market in Brazil during the years 1995-2014 and the changes predicted by the model under policy experiment (iii.a). The table contents should be interpreted as follows: $\uparrow$ denotes an increase for any value in the parameters of the model; $\downarrow$ denotes a decrease; $\uparrow\downarrow$ denotes that it can increase or decrease. Other cells are filled with lower or upper bound to changes in the outcome variable.

Table 7: The labor market effects of an educational expansion: data vs model

	Data-Brazil	**Model**	**%Predicted**
Δ Thresholds			
I_L		-0.167	
I_H		0.084	
Δ Mean Occup. Ranking 1995-2014			**71.6**
Low educated	-0.037	-0.083	44.6
Medium educated	-0.129	-0.126	97.7
High Educated	-0.058	-0.042	72.4
% Δ Occup. Composition of Employment 1995-2014			**38.3**
Bottom-third	-0.011	-0.072	15.3
Middle-third	-0.005	-0.004	80.0
Top-third	0.015	0.076	19.7
% Δ in Wages 1995-2014			**91.1**
Total workforce	27.8	29.3	94.9
Low educated	37.3	46.9	79.5
Medium educated	-10.5	-10.9	96.3
High Educated	-21.6	-22.2	97.3
% Δ Wage Gap 1995-2014			**92.5**
W_H/W_M	-12.44	-12.72	97.8
W_M/W_L	-34.81	-39.34	88.5
W_H/W_L	-42.92	-47.05	91.2
Δ Welfare and Inequality			**84.6**
First order dominance	Yes	No	
Gini	-0.085	-0.082	96.5
Theil	-0.109	-0.099	90.8
Between component	-0.020	-0.099	20.2
Atkinson(2)	-0.132	-0.088	66.7
Between component	-0.078	-0.088	88.6
Δ Wage Poverty (wage line 2.4)			**68.4**
FGT(0)	-0.164	-0.331	49.5
FGT(1)	-0.162	-0.237	68.3
FGT(2)	-0.118	-0.103	87.3

Notes: Estimations for Brazil comes from PNAD 1995-2014. The estimations for the model comes from simulating the Brazilian educational expansion between 1995-2014 on the calibrated model, holding constant the rests of the parameters. The Brazilian educational expansion consists of: an increase in the share of medium educated from 20.5 to 40.0 percent, an increase in the share of high educated workers from 11.3 to 23.6 percent, and a decline in the share of low educated workers from 68.1 to 36.4 percent.

Table 8: Differential effect of an educational expansion in 1995 and 2014

	Effect of 1 p.p increase in m		Effect of 1 p.p increase in h	
	1995	2014	1995	2014
Productivity	0.1	-0.1	1.8	0.7
Wages				
Av. wage	0.5	0.1	2.3	1.1
W_L	0.7	1.2	1.8	2.0
W_M	-0.7	-0.5	-0.1	0.1
W_H	-0.3	-0.3	-2.2	-1.3
Relative wages				
W_M/W_L	-1.4	-1.7	-1.9	-1.9
W_H/W_M	0.4	0.3	-2.1	-1.3

Notes: The table presents the results of simulating an increase of 0.01 in m or h in the model under the different equilibrium conditions arising from a low educated workforce in 1995 and a higher educated workforce in 2014. The results have to be interpreted as the predictions of the model of increasing the supply of skills by 1 percentage point while decreasing the share of low educated workers in the same magnitude.

Table 9: Model results with different calibrations

	Brazil	Model calibration			
		1995	**mean 1995-2014**	**min 1995-2014**	**max 1995-2014**
Thresholds					
Initial I_L		0.424	0.420	0.419	0.421
Initial I_H		0.664	0.661	0.652	0.669
Final I_L		0.257	0.254	0.254	0.255
Final I_H		0.580	0.578	0.567	0.586
Average occp ranking 1995					
Low educated	0.243	0.212	0.210	0.209	0.210
Medium educated	0.454	0.544	0.541	0.535	0.545
High Educated	0.669	0.832	0.831	0.826	0.834
Delta mean occup. Ranking 1995-2014					
Low educated	-0.046	-0.083	-0.083	-0.082	-0.083
Medium educated	-0.135	-0.126	-0.125	-0.125	-0.125
High Educated	-0.064	-0.042	-0.042	-0.043	-0.041
% Change in overall occuaptions 1995-2014					
Bottom-third	-0.011	-0.072	-0.074	-0.072	-0.076
Middle-third	-0.005	-0.004	-0.003	-0.009	0.001
Top-third	0.015	0.076	0.077	0.081	0.075
Average wage 1995					
Low educated	1.47	1.47	1.31	1.11	1.53
Medium educated	2.76	2.76	2.51	2.05	2.99
High Educated	7.00	7.00	6.37	5.54	7.25
% Changes in wages 1995-2014					
Low educated	49.1	46.9	47.3	47.8	47.0
Medium educated	-4.5	-10.9	-10.8	-10.4	-11.0
High Educated	-26.0	-22.2	-22.1	-22.1	-22.0
% Change Wage gap 1995-2014					
W_H/W_M	-22.43	-12.72	-12.67	-13.06	-12.38
W_M/W_L	-36.00	-39.34	-39.45	-39.38	-39.49
W_H/W_L	-50.35	-47.05	-47.12	-47.30	-46.98

Notes: Estimations for Brazil comes from PNAD 1995-2014. The productivity parameters of the model are calibrated in different ways, According to the values in Figure 14. The model simulates an educational expansion between 1995-2014 following Table 1. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university).

A Appendix 1

A.1 Proof of Lemma 1

The profit of a competitive firm producing tasks i is:[49]

$$\Pi(i) = p(i)\left(L\alpha_L(i)L(i) + A_M\alpha_M(i)M(i) + A_H\alpha_H(i)H(i)\right) - w_L L(i) - w_M M(i) - w_H H(i),$$

where the price of the task $p(i)$ and the wage of each type of worker are given for the firm. The first order conditions with respect to $L(i)$, $M(i)$, and $H(i)$ imply

$$J(i) > 0 \qquad if \qquad \pi_J(i) = p(i)A_J\alpha_J(i) - w_J \geq 0 \qquad \text{for } J = \{L, M, H\}.$$

Competition assures that $\Pi(i) \leq 0$. Moreover, the Cobb-Douglas production function for the final good implies that production of all task is positive. So $\forall i$ it must be that either $\pi_L(i) = 0$, or $\pi_M(i) = 0$, or $\pi_H(i) = 0$ or two of these conditions has to hold. Unemployment of any type of worker is not possible in this economy because it would imply that wages fall to zero and firms could make a profit in producing task i with that type. Then, a positive amount has to be produce of each task using either low, medium, or high educated workers. Therefore, given the assumption of increasing comparative advantage on i for more educated workers, there must exist two thresholds levels I_L and I_H (where $0 < I_L < I_H < 1$) such that: a) $\pi_L(i) - \pi_M(i) > 0$ and $\pi_L(i) - \pi_H(i) > 0$ for all $i < I_L$, $\pi_L(I_L) - \pi_M(I_L) = 0$; b) $\pi_M(i) - \pi_L(i) > 0$ and $\pi_M(i) - \pi_H(i) > 0$ for all $i = (I_L, I_H)$; c) $\pi_M(I_H) - \pi_H(I_H) = 0$; and d) $\pi_H(i) - \pi_L(i) > 0$ $\pi_H(i) - \pi_M(i) > 0$ for all $i > I_H$.

A.2 Derivations of the Equilibrium Conditions in the Model

The competitive equilibrium in the economy described in Section 5 consist in an assignment of workers' type to tasks and real wages for each type of worker such that producers maximize profits and labor markets clear, given the relative supply of skills and the productivity across tasks of each type of worker. The equilibrium conditions in this model are similar to those in A&A. I keep the notation to a minimum and refer to A&A for a more detailed discussion.

The characterization of the equilibrium is simple given the positive assortative matching that arises from the structure of the model (supermodular production function). In particular, there exists two thresholds I_L and I_H that determines what tasks are performed by low($i < I_L$), medium ($I_L < i < I_H$) and high educated workers ($i > I_H$). See Lemma 1 in Section 5 for a formal definition of the thresholds.

These thresholds naturally arise from the assumption that comparative advantage is increasing in (i), generating positive assertive matching. Intuitively, it is optimal for employers to use the most productive workers (high educated) to the more complex task (closer to the index 1), where they have a larger comparative advantage. On the opposite, it is in their profit-

[49]The proof presented here follows the analisys in Acemoglu & Zilbotti (1999).

maximizing interest to employ low educated workers in the tasks of low complexity (closer to 0). And finally, medium educated workers will be employed in the remaining task with a medium level of complexity. The location of these thresholds depends on the relative supply of workers' types and on their relative productivities in each task, which are assumed to be exogenous.

Prices

The final good is the numeraire of this economy, that is:

$$exp\left[\int_0^1 ln\, p(i)\, di\right] = 1. \tag{16}$$

Workers with the same skills perform different tasks, but they will receive the same wage because of the competitive equilibrium assumption. Let w_L, w_M w_H be the wage of low, medium and high educated respectively, it is possible to state:

$$w_L = p(i)A_L\alpha_L(i) \qquad \forall i < I_L.$$

$$w_H = p(i)A_M\alpha_M(i) \qquad \forall I_L < i > I_H.$$

$$w_H = p(i)A_H\alpha_H(i) \qquad \forall i > I_H.$$

It implies that for two tasks i and i' that are produced using the same type of labor it must be the case that price differences exactly offset productivity differences. That is:

$$p(i)\alpha_L(i) = p(i')\alpha_L(i') \equiv P_L \qquad \forall i, i' \in (0, I_L). \tag{17}$$

$$p(i)\alpha_M(i) = p(i')\alpha_M(i') \equiv P_M \qquad \forall i, i' \in (I_L, I_H). \tag{18}$$

$$p(i)\alpha_H(i) = p(i')\alpha_H(i') \equiv P_H \qquad \forall i, i' \in (I_H, 1). \tag{19}$$

Because the technology of the production function is Cobb-Douglas, the expenditure across all task should be equalized ($p(i)y(i) = p(i')y(i')$ for any i,i'). It implies that for any two task produced by low educated workers, that is

$$p(i)\alpha_L(i)L(i) = p(i')\alpha_L(i')L(i')$$

Given equation (17), $L(i) = L(i')$ for all i,$i' < I_L$. Using the market clearing condition (3),

$$E_i = \begin{cases} L(i) = \frac{l}{I_L}, & if\ 0 < i < I_L \\ M(i) = \frac{m}{(I_H - I_L)}, & if\ I_L < i < I_H \\ H(i) = \frac{h}{(1 - I_H)}, & if\ I_H < i < 1. \end{cases} \tag{20}$$

Comparing now tasks performed by low educated workers and those of medium educated workers, and using the fact that total expenditure of employers in each task has to be the same to maximize profits (because of the production is Cobb-Douglas), that is:

$$p(i)A_L\alpha_L(i)L(i) = p(i')A_M\alpha_M(i')M(i'), \qquad \text{for any } i < I_L, I_L < i' > I_H.$$

$$p(i)A_M\alpha_M(i)M(i) = p(i')A_M\alpha_M(i')M(i'), \qquad \text{for any } I_L < i > I_H, i' > I_H.$$

Using equations (17), (19), (18), (20),

$$\frac{P_M}{P_L} = \left(\frac{I_H - I_L}{I_L}\right)\left(\frac{A_L l}{A_H h}\right). \tag{21}$$

$$\frac{P_H}{P_M} = \left(\frac{1 - I_H}{I_H - I_L}\right)\left(\frac{A_M m}{A_H h}\right). \tag{22}$$

No arbitrage across skills

The cost of producing task I_L with low or medium educated workers must be the same or employers could increase profits by using the less expensive worker in this task. Similar, to produce task I_H with medium and high educated workers must be the same. We can express these conditions as follows:

$$p(I_L)A_L\alpha_L(I_L)L(I_L) = p(I_L)A_M\alpha_M(I_L)M(I_L).$$

$$p(I_H)A_M\alpha_L(I_H)M(I_H) = p(I_H)A_H\alpha_H(I_H)H(I_H).$$

By using (20),

$$\frac{(I_H - I_L)}{I_L}\frac{\alpha_L(I_L)}{\alpha_M(I_L)} = \frac{A_M}{A_L}\frac{m}{(1 - m - h)}. \tag{23}$$

$$\frac{(1 - I_H)}{I_H - I_L}\frac{\alpha_M(I_H)}{\alpha_H(I_H)} = \frac{A_H}{A_M}\frac{h}{m}. \tag{24}$$

These equations provide a unique mapping between the thresholds I_H and I_L, and the relative supply and relative productivities of high and low educated workers across different tasks.

Equilibrium wages

Wages are equal to the value of the marginal products of different type of workers, that is:

$$W_L = P_L A_L. \tag{25}$$

$$W_M = P_M A_M. \tag{26}$$

$$W_H = P_H A_H. \tag{27}$$

By equations 21 and 22, I get the following expression for relative wages:

$$\frac{W_M}{W_L} = \frac{P_M A_M}{P_L A_L} = \frac{(I_H - I_L)}{I_L}\frac{(1-h-m)}{m}. \tag{28}$$

$$\frac{W_H}{W_M} = \frac{P_H A_H}{P_M A_M} = \frac{(1 - I_H)}{I_H}\frac{(m)}{h}. \tag{29}$$

To finalize characterizing the equilibrium, it is necessary to compute price levels. Due to the choose of the numeraire, $\left[\int_0^1 ln\, p(i)\, di\right] = 0$ (from equation (16)). By using equations (17), (18) and (19),

$$\int_0^{I_L} (ln\, P_L\ - ln\, \alpha_L(i))\, di + \int_{I_L}^{I_H} (ln\, P_M\ - ln\, \alpha_M(i))\, di + \int_{I_H}^{1} (ln\, P_H\ - ln\, \alpha_H(i))\, di = 0. \tag{30}$$

This equations, together with (28) and (29), solves for the price levels P_L, P_M and P_H. Having estimated prices, real wages are computed using (25), (26) and (26), and are equal to:

$$ln\, W_H = AVP + I_H C_{HM}(I_H) + I_L C_{ML}(I_L) \tag{31}$$

$$ln\, W_M = AVP - (1 - I_H)C_{HM}(I_H) + I_L C_{ML}(I_L) \tag{32}$$

$$ln\, W_L = AVP - (1 - I_H)C_{HM}(I_H) - (1 - I_L)C_{ML}(I_L) \tag{33}$$

where,

$$AVP = \int_0^{I_L} ln\, A_L \alpha_L(i)\, di + \int_{I_L}^{I_H} ln\, A_M \alpha_M(i)\, di + \int_{I_H}^{1} ln\, A_H \alpha_H(i)\, di.$$

and

$$C_{HM}(j) = ln\, A_H \alpha_H(j) - ln\, A_M \alpha_M(j)$$

$$C_{ML}(j) = ln\, A_M \alpha_M(j) - ln\, A_L \alpha_M(j).$$

By taking the difference og the log wage equations, the wages gaps are then equal to:

$$ln\, W_H - ln\, W_M = C_{HM}(I_H) \tag{34}$$

$$ln\, W_M - ln\, W_L = C_{ML}(I_L) \tag{35}$$

$$ln\, W_H - ln\, W_L = C_{HM}(I_H) + C_{ML}(I_L) \tag{36}$$

A.3 Proof of propositions 1-3

The proof of propositions (1), (2) and (3) proceeds as follows. First, I compute the thresholds' changes that determine the distributions of workers' types to tasks. Then, the effects on wages and the overall distribution of occupations follow directly from the new thresholds.

Changes in task composition within each educational level

The task composition within each educational level is determined by the thresholds levels I_L and I_H. I use comparative statics to estimate the effects of an educational expansion on the thresholds levels.

Proof of $(i.1)$. I follow A&A by expressing (5) and (6) in logs:

$$ln\,(1 - I_H) - ln\,(I_H - I_L) - C_{HM}(I_H) - ln\,h + ln\,m = 0 \tag{37}$$

$$ln\,(I_H - I_L) - ln\,(I_L) - C_{ML}(I_L) - ln\,m + ln\,(1 - m - h) = 0. \tag{38}$$

Now consider the effect of a change in h by totally differentiating these equations. We thus obtain:

$$\begin{bmatrix} -\frac{1}{(1-I_H)} - \frac{1}{(I_H-I_L)} - C'_{HM}(I_H) & \frac{1}{(I_H-I_L)} \\ \frac{1}{(I_H-I_L)} & -\frac{1}{(I_H-I_L)} - \frac{1}{I_L} - C'_{ML}(I_L) \end{bmatrix} \begin{bmatrix} dI_H \\ dI_L \end{bmatrix} = \begin{bmatrix} \frac{1}{h} \\ \frac{1}{(1-m-h)} \end{bmatrix} dh \tag{39}$$

The determinant of the first matrix, Δ is positive (see A&A), therefore by Cramerś rule:

$$\frac{dI_H}{dh} = \frac{\frac{1}{h}\left(-\frac{1}{(I_H-I_L)} - \frac{1}{I_L} - C'_{ML}(I_L)\right) - \frac{1}{(1-m-h)}\frac{1}{(I_H-I_L)}}{\Delta} < 0 \tag{40}$$

$$\frac{dI_L}{dh} = \frac{\frac{1}{(1-m-h)}\left(-\frac{1}{(1-I_H)} - \frac{1}{(I_H-I_L)} - C'_{HM}(I_H)\right) - \frac{1}{h}\frac{1}{(I_H-I_L)}}{\Delta} < 0 \tag{41}$$

The inequality in the first equation is straightforward because all its terms are negative. For the second equation to be positive it is necessary that

$$-C'_{HM}(I_H) > \frac{(1-m-h)}{h}\frac{1}{(I_H - I_L)} + \frac{1}{1 - I_H} + \frac{1}{(I_H - I_L)}$$

Which leads to a contradiction since $C'_{HM}(I_H) > 0$ and the RHS is positive. Therefore $\frac{dI_L}{dh} < 0$. ■

An increase in h leads to a reduction in both thresholds, given that higher educated workers start performing medium-level tasks, and medium educated workers are pushed towards lower-level tasks displacing low educated workers. With an increase in h, the average task performed by each educational level diminishes leading to conditional occupational downgrading.

Proof of $(ii.1)$. Following a similar procedure it is possible to write:

$$\frac{dI_L}{dm} = \frac{\left(\frac{1}{m} + \frac{1}{(1-m-h)}\right)\left(-\frac{1}{(1-I_H)} - \frac{1}{(I_H-I_L)} - C'_{HM}(I_H)\right) + \frac{1}{m}\frac{1}{(I_H-I_L)}}{\Delta} < 0 \tag{42}$$

$$\frac{dI_H}{dm} = \frac{-\frac{1}{m}\left(-\frac{1}{(I_H-I_L)} - \frac{1}{I_L} - C'_{ML}(I_L)\right) - \left(\frac{1}{m} + \frac{1}{(1-m-h)}\right)\frac{1}{(I_H-I_L)}}{\Delta} > 0 \tag{43}$$

The inequality in the first equation is straightforward. For the second equation to be positive it is necessary that:

$$C'_{ML}(I_L) > \frac{m}{(1-m-h)}\frac{1}{(I_H-I_L)} - \frac{1}{I_L}$$

From (38) the RHS is < 0, because

$$\frac{m}{(1-m-h)}\frac{1}{(I_H-I_L)} = \frac{1}{I_L}\frac{A_L\alpha_L(I_L)}{A_M\alpha_M(I_L)} < \frac{1}{I_L}.$$

The result $\frac{dI_H}{dm} > 0$ follows by the comparative advantage schedule, given that $C'_{ML}(I_L) > 0$. ■

It is easy to see that with an increase in m the average task performed by low educated workers ($I_L/2$) and that of high educated workers ($(1-I_H)/2$) will increase. The change in the average task performed by medium educated workers depends on the parameters of the model. Moreover, an increase in m leads to a larger share of tasks performed by medium educated workers, by increasing its participation in low-level and high-level tasks. We are also interested in $\frac{d(I_H+I_L)}{dm}$ since the average task performed by medium educated workers can be defined as $I_L+(I_H-I_L)/2 = (I_H+I_L)/2$. Depending on which effect is stronger, $\frac{dI_H}{dm}\frac{dI_L}{dm} \lesseqgtr 0$. The average task performed by medium educated workers will increase or decrease depending on whether this inequality is positive or negative. We can write it as follows:

$$\frac{d(I_H+I_L)}{dm} = \frac{-\frac{1}{m}\left(-\frac{1}{I_L} - C'_{ML}(I_L)\right) - \frac{1}{(1-m-h)}\frac{1}{(I_H-I_L)}}{\Delta} + \frac{\left(\frac{1}{m} + \frac{1}{(1-m-h)}\right)\left(-\frac{1}{(1-I_H)} - \frac{1}{(I_H-I_L)} - C'_{HM}(I_H)\right) + \frac{1}{m}\frac{1}{(I_H-I_L)}}{\Delta} \lesseqgtr 0$$

When m and h increase separately, I_L always declines and I_H increases in the first case and decreases in the second case. When m and h increase simultaneously there is a mixture of these effects. As a consequence, I_L will diminish but I_H can increase or decrease depending on what effect is stronger.

Proof of ($iii.3$). Suppose that $\frac{m'}{h'} < \frac{m}{h}$ and $I'_H > I_H$. By equations (41) and (42) $I_L < I'_L$. From (6), $\frac{m}{h} = c_{HM}(I_H)(I_H - I_L)$, with $c_{HM} = exp(C_{HM})$ and where $c'_{HM}(I_H) > 0$. After the educational expansion $\frac{m'}{h'} = c_{HM}(I'_H)(I'_H - I'_L) > \frac{m}{h}$, which is a contradiction.■

It means that I_H cannot increase unless the relative supply of medium with respect to high educated workers also increases. However, this condition does not work in the other direction.

Changes task composition of the economy

In only a few cases the model provides a sharp prediction of the changes in the task composition of the economy that does not depend on the parameters of the model.

Proof of ($i.2$) *and* ($ii.2$). When there is an increase in I_H, some medium educated workers will be added to tasks originally performed only by high educated workers. Therefore, the share of employment in high-level tasks in the economy will increase.

$$E_T = h$$

$$E'_T = h' + \int_{I_H}^{I'_H} \frac{m'}{I'_H - I'_L} di$$

$$\Delta E_T = \Delta h + \int_{I_H}^{I'_H} \frac{m'}{I'_H - I'_L} di > 0 \blacksquare$$

This inequality is written for the more general case in which m and h increases. Note that the inequality stills holds for proposition 2 when $\Delta h = 0$.

In the rest of the cases, when I_H declines, it is only possible to define boundaries for the changes in the occupational composition of employment. For example,

$$E_B = l$$

$$E'_B = l' + \int_{I'_L}^{I_L} \frac{m'}{I'_H - I'_L} di$$

$$\Delta E_B = (l' - l) + (I_L - I'_L) \frac{m'}{I'_H - I'_L} > \Delta l.$$

The inequality establishes that when the share of low educated workers is reduced by Δl, the share in bottom-level tasks will decline less than Δl since some medium educated workers will start performing some of these tasks. The extent to which medium educated workers will start performing those tasks depends on the extent of the decline in I_L, which at the same time depends on the comparative advantage schedules and whether the decline in l was due to an increase in m, h or both.

A similar result arises when looking at changes in employment for top-level occupations when I_H declines. Formally,

$$\Delta E_T = (h' - h) - (I_H - I'_H) \frac{h'}{1 - I'_H} < \Delta h.$$

A decline in I_H is only possible if $\Delta h > 0$. The inequality establishes that there is less increase in employment in top-level occupations than the increase in the share of high educated workers. Some high educated workers start performing middle-level tasks. To study the determinants for an increase in employment in top-level tasks, it is possible to state that:

$$\Delta E_T > 0 \Leftrightarrow \frac{\Delta h}{h'} > \frac{(I_H - I'_H)}{(1 - I'_H)}$$

that is, the increase in the supply of high educated workers has to be large with respect to the change in the threshold I_H.

Changes in the wages

We now turn to the case of changes in the wage levels. From the wage equations (7), (8) and (9) the threshold levels determine the wage levels. After an educational expansion I_L always falls, while I_H can increase or decrease depending on the case. I consider these two cases separately.

Case 1: $\Delta I_L < 0$ *and* $\Delta I_H < 0$

Proof of proposition $(i.3)$.. After an educational expansion, wages for high educated workers can be expressed as:

$$ln\, W'_H = AVP' + I'_H C_{HM}(I'_H) + I'_L C_{ML}(I'_L).$$

Therefore, the wage increase $\Delta ln\, W_H = ln\, W'_H - ln\, W_H$ can be expressed as:

$$\Delta ln\, W_H = \underbrace{\int_{I'_L}^{I_L} C_{ML}(i)\, di + I'_L C_{ML}(I'_L) - I_L C_{ML}(I_L)}_{\Delta I_L \text{ in } W_H \text{ effect} < 0} + \underbrace{\int_{I'_H}^{I_H} C_{HM}(i)\, di + I'_H C_{HM}(I'_H) - I_H C_{HM}(I_H)}_{\Delta I_H \text{ in } W_H \text{ effect } < 0} < 0,$$

where the inequalities for each term comes from the assumption about the comparative advantage schedules. Given the comparative and absolute advantage schedule, $C_{ML}(I_L) > C_{ML}(i) > C_{ML}(I'_L)$ for all $i = (I'_L, I_L)$ and $C_{HM}(I_H) > C_{HM}(i) > C_{HM}(I'_H)$ for all $i = (I'_H, I_H)$. I can establish the following inequalities:

$$\int_{I'_L}^{I_L} C_{ML}(i)\, di < I_L C_{ML}(I_L) - I'_L C_{ML}(I_L) < I_L C_{ML}(I_L) - I'_H C_{ML}(I'_L)$$

$$\int_{I'_H}^{I_H} C_{HM}(i)\, di < I_H C_{HM}(I_H) - I'_H C_{HM}(I_H) < I_H C_{HM}(I_H) - I'_H C_{HM}(I'_H)$$

Therefore, $\Delta ln\, W_H < 0$.

Similarly, define $\Delta ln\, W_L = ln\, W_L - ln\, W'_L$. The change in wages for low educated workers can be expressed as

$$\Delta ln\, W_L = \underbrace{\int_{I'_L}^{I_L} C_{ML}(i)\, di - (1 - I'_L) C_{ML}(I'_L) + (1 - I_L) C_{ML}(I_L)}_{\Delta I_L \text{ in } W_L \text{ effect } > 0} + \underbrace{\int_{I'_H}^{I_H} C_{HM}(i)\, di - (1 - I'_H) C_{HM}(I'_H) + (1 - I_H) C_{HM}(I_H)}_{\Delta I_H \text{ in } W_L \text{ effect } > 0} > 0,$$

where the inequalities arise from considering

$$\int_{I'_L}^{I_L} C_{ML}(i) > I_L C_{ML}(I'_L) - I'_L C_{ML}(I'_L)$$

and,

$$\int_{I'_H}^{I_H} C_{HM}(i) > I_H C_{HM}(I'_H) - I'_H C_{HM}(I'_H)$$

then,

$$\begin{aligned}\Delta ln\, W_L &> I_L C_{ML}(I'_L) - I_L C_{ML}(I_L) + C_{ML}(I_L) - C_{ML}(I'_L) \\ &+ I_H C_{HM}(I'_H) - I_H C_{HM}(I_H) + C_{HM}(I_H) - C_{HM}(I_H L') \\ &= (1 - I_L)(C_{ML}(I_L) - C_{ML}(I'_L)) + (1 - I_H)(C_{HM}(I_H) - C_{HM}(I'_H)) > 0.\end{aligned}$$

Finally, wages of medium educated workers can be expressed as:

$$\begin{aligned}\Delta ln\, W_M = &\underbrace{\int_{I'_L}^{I_L} C_{ML}(i)\, di + I'_L C_{ML}(I'_L) - I_L C_{ML}(I_L) +}_{\Delta I_L \text{ in } W_M \text{ effect} < 0} \\ &\underbrace{\int_{I'_H}^{I_H} C_{HM}(i)\, di - (1 - I'_H) C_{HM}(I'_H) + (1 - I_H) C_{HM}(I_H)}_{\Delta I_H \text{ in } W_M \text{ effect } > 0} \lesseqgtr 0,\end{aligned}$$

where the results follow from the inequalities in opposite directions derived above for $\Delta ln W_H$ and $\Delta ln\, W_L$.

When $\Delta I_L < 0$ and $\Delta I_H < 0$, W_H increases, W_L decreases, and W_L can increase or decrease depending on the parameters of the model.

■

Case 2: $\Delta I_L < 0$ *and* $\Delta I_H > 0$

Proof of proposition $(ii.3)$.. Similarly, in the case$I_H > 0$ it is possible to express:

$$\begin{aligned}\Delta ln\, W_H = &\underbrace{\int_{I'_L}^{I_L} C_{ML}(i)\, di + I'_L C_{ML}(I'_L) - I_L C_{ML}(I_L) +}_{\Delta I_L \text{ in } W_H \text{ effect} < 0} \\ &\underbrace{- \int_{I_H}^{I'_H} C_{HM}(i)\, di + I'_H C_{HM}(I'_H) - I_H C_{HM}(I_H)}_{\Delta I_H \text{ in } W_H \text{ effect } > 0} \lesseqgtr 0,\end{aligned}$$

$$\begin{aligned}\Delta ln\, W_M = &\underbrace{\int_{I'_L}^{I_L} C_{ML}(i)\, di + + I'_L C_{ML}(I'_L) - I_L C_{ML}(I_L) +}_{\Delta I_L \text{ in } W_M \text{ effect} < 0} \\ &\underbrace{- \int_{I_H}^{I'_H} C_{HM}(i)\, di - (1 - I'_H) C_{HM}(I'_H) + (1 - I_H) C_{HM}(I_H)}_{\Delta I_H \text{ in } W_M \text{ effect } < 0} < 0,\end{aligned}$$

$$\Delta ln\, W_L = \underbrace{\int_{I'_L}^{I_L} C_{ML}(i)\, di - (1 - I'_L)C_{ML}(I'_L) + (1 - I_L)C_{ML}(I_L) +}_{\Delta I_L \text{ in } W_L \text{ effect} > 0}$$

$$\underbrace{- \int_{I_H}^{I'_H} C_{HM}(i)\, di - (1 - I'_H)C_{HM}(I'_H) + (1 - I_H)C_{HM}(I_H)}_{\Delta I_H \text{ in } W_L \text{ effect} < 0} \lesseqgtr 0,$$

■

Changes in relative wages

Relative wages are easy to estimate in the model. Their level depends on the changes in the thresholds and are equal to the difference of the equations stated above (alternatively, they can be obtained directly from equations (10)-(12)) .

Proof of (*i*.4) *and* (*ii*.4). Relative wages can be expressed as follows.

$$ln\, W'_H - ln\, W'_M - (ln\, W_H - ln\, W_M) = C_{HM}(I'_H) - C_{HM}(I_H) \begin{cases} > 0, & \text{if } I'_H > I_H. \\ < 0, & \text{if } I'_H < I_H. \end{cases}$$

$$ln\, W'_M - ln\, W'_L - (ln\, W_M - ln\, W_L) = C_{ML}(I'_L) - C_{ML}(I'_L) < 0$$

$$lnW'_H - lnW'_L - (lnW_H - lnW_L) = C_{HM}(I'_H) - C_{HM}(I_H) + C_{ML}(I'_L) - C_{ML}(I'_L) \begin{cases} \lesseqgtr 0, & \text{if } I'_H > I_H. \\ < 0, & \text{if } I'_H < I_H. \end{cases}$$

■

I find that in any educational expansion the wage gap between medium and low educated workers always fall according to the model. This is the case because medium educated workers will always start performing bottom-level tasks when there is an educational expansion (decline in I_L), and the productivity differential of medium with respect to low educated workers always decline, and this is what defines the wage gaps.

B Online Appendix

B.1 Estimations of Changes in Quality of Human Capital

When the access to education increases rapidly in a country, it is possible that the quality of education or the unobserved ability of people being educated diminished. If that is the case, the quantity of human capital after the educational expansion may be lower for a given educational attainment.

In this section, I evaluate whether the observed changes in wages of medium and high educated workers are driven by market forces or by a change in the human capital that these educational levels represent. I follow Bowlus & Robinson (2012) and Heckman et al. (1998) by computing changes in the prices of human capital by looking into yearly changes in median wages of cohorts for which the quantity of human capital should have remained constant. I conclude that most of the changes in wages are due to changes in the price of human capital, providing support to the hypothesis that the decline of wages were driven by the market forces described in the previous sections.

In the theory and the model presented above, an educational level is equivalent to a quantity of human capital that the workers possess. In another strand of the literature related to human capital models, wages are defined as the product of efficiency units of human capital that each worker poses and its price. Let w_{it} be the wage of worker i in time i, λ_t be the market level price for human capital, and E_{it} be the quantity of human capital. It is possible to define

$$w_{t,i} = \lambda_t E_{t,i}.$$

Changes on median wages of workers between t and t' with educational level s from cohort c can be defined as:

$$ln\, Median\, w_{t',s,c} - ln\, Median\, w_{t,s} = ln\, \lambda_{t',s,c} - ln\, \lambda_{t,s,c} + ln\, Median\, E_{t',s,c} - ln\, Median\, E_{t,s,c}.$$

This equation shows that change in median wages can be decomposed into a market price effect given by the first term and a human capital or quantity effect reflected in the second term. To estimate the incidence of each of these two terms in changes in wages, Heckman et al. (1998) look into changes in wages in cohorts for which the level of human capital does not change in time. In other words, they look into the age groups for which the quantity of human capital is assumed to remain constant so that the second term in the previous equation is equal to zero. Let define the cohorts of educational level s for which human capital remain constant from t to t'. The changes in prices for an educational group s can be defined as an average of the changes in prices for each cohort in the flat point like follows

$$ln\, \lambda_{s,t'} - ln\, \lambda_{s,t} = \frac{1}{n}\sum_c ln\, \lambda_{t',s,c} - ln\, \lambda_{t,s,c} = \frac{1}{n}\sum_c ln\, Median\, w_{t',s,c} - ln\, Median\, w_{t,s,c}, \quad (44)$$

where n is the number of cohorts consider to be in the flat spot.

To find the ages groups, I follow Bowlus & Robinson (2012) who define it for similar years of potential experience for each educational groups, finding the "flat" point by looking into the cross-sectional life earning cycle of wages. I find that the life earnings cycle is flat in most years when looking into low educated between 43-52 years old, between 48-57 years for medium educated workers, and 50-61 for high educated workers.[50]. Table 10 replicates Table 1 in Bowlus & Robinson (2012). A flat stop is found if earnings do not change with age. I found that to be the case for most of the years, except for the first four years for low educated workers. As indicated by Bowlus & Robinson (2012), this could generate a downward bias in the price series of low educated workers, since the decline in prices for those years might be partly due to decline in the quantity of human capital. In this case, the price series I estimate can be interpreted as a lower bound.

The price series estimated with this method is displayed in Figure 17. Note that the evolution of prices is similar to the patters funded for median wages depicted in Figure 17. Once the price series is obtained, the quantity effect is estimated as the residual of wage effects minus the price effect.

Table 11 summarizes the results of this decomposition exercise for the period 1995-2014. I found that the price effects explain most of the changes in wages between 1995 and 2014. Most of the increase in wages for low educated workers and the decrease for high educated workers (97% and 84% respectively) is explained by the price effect. For medium educated workers, price effects are larger than the changes in wages, suggesting that the quantity of education embedded in medium educated workers is actually higher in 2014 than in 1995.

The evidence presented in this section is consistent with the theoretical model presented above where the changes are originated by market level effects as a response to changes in the supply of skills, and not due to education miss measuring the levels of skills. This evidence is also consistent with results from international standardize tests that find an increase in test scores during the 2000s.[51]. It has been a remarkable achievement for Brazil to have one of the largest educational expansions in history without decreasing the quality of education. However, this quality is still low to international standards and one of the main challenges that the educational system faces today is on increasing the quality.

[50] I perform this analysis only for men, working more than 35 hours a week. The reason to restrict the sample to men is that at this age range participation in the labor force decline rapidly for women, and price effect may be biased due to selection effects into the labor force

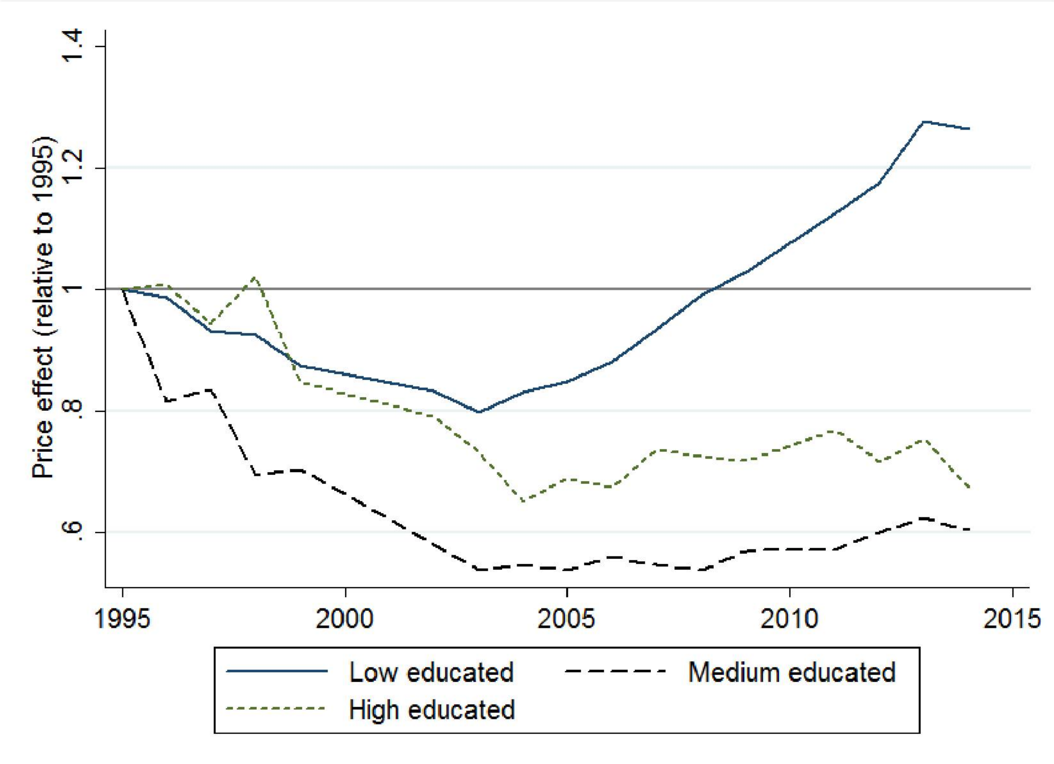
[51] Bruns et al. (2011) contain a detailed analysis of the evolution of test scores in Brazil when compared to other countries as a measure of the evolution of quality of education. The authors conclude that, although the scores are still far below OCDE levels, Brazil has made a fast and sustained progress during the 2000s.

Table 10: Flat point on earnings: Cross sectional evidence

	Low educated 43-52 years old	Mediun Educated 48-57 years old	High educated 51-60 years old
1995	-0.00965**	0.00746	0.0300*
1996	-0.0115***	-0.0133	0.00505
1997	-0.0125***	-0.00230	-0.0178
1998	-0.00683*	-0.00235	0.000176
1999	0.00116	-0.00728	-0.0307*
2002	-0.00200	0.00943	0.00564
2003	0.00261	0.00323	-0.00511
2004	0.00508	0.00702	-0.00456
2005	0.00389	0.0118	-0.00290
2006	0.00303	0.00248	-0.00344
2007	0.00242	0.0113*	0.0104
2008	0.00211	0.00570	0.00191
2009	0.00389	0.00354	0.00122
2011	-0.00125	-0.0000646	0.00331
2012	0.000809	0.00264	0.0131
2013	0.00269	0.00763	-0.00329
2014	-0.00121	0.00523	0.0166*

Note: Each cell in the table shows the returns to an extra year of experience in a mincer equation of low wages. The sample is restricted only to employed males. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university). * Significant at 10 percent. ** Significant at 5 percent. *** Significant at 1 percent.

Figure 17: Evolution of prices relative to 1995



Note: The figure shows the evolution of the price effect for low, medium and high educated workers. The price effect is estimated for male working more than 30 hours a week according to equation (44). The flat spot age range is 43-52 for low educated, 48-57 for medium educated, and 51-60 for high educated. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university).

Table 11: Decomposition of the changes in wages on price and quantity effects

	Low	Medium	High
Median log wages 1995	0.504	1.197	2.209
Median log wages 2014	0.776	1.038	1.820
Change in wages 1995-2014	0.272	-0.160	-0.389
Prices effect	0.264	-0.396	-0.326
Quantity effect	0.008	0.237	-0.063

Note: Median wages are estimated for males working more than 30 hours per week. Price effect is estimated from Figure 17. Quantity effect is the residual of the total change in median wages minus the price effect. Workers are classified into low (less than secondary), medium (some secondary), and high educated (some university).

B.2 Oaxaca Decomposition of Average Wages

This section evaluates the relationship between changes between 1995 and 2014 in average wages of each educational group and their occupational composition of employment. Wages for a given educational group may change in part because workers switched to occupations with a different wage, and in part because there was an increase or decrease in wages of occupations they were already performing. The Oaxaca decomposition disentangles the importance of these two effects on changes in average wages.

The changes in log wages between 1995 and 2014 can be decomposed by the method proposed by Blinder (1973) and Oaxaca (1973), and the role of occupations is obtained following Ferreira et al. (2016). Let's W_{it}^s be the wage of individual i from the educational group s in time t and X be a vector of occupational dummies and other personal characteristics (gender and age group dummies). Then we can express:

$$log(W_{it}^s) = \beta_t^s X_{it}^s + \epsilon_{it}^s$$

Consider $t = 1995, 2015$, it is possible to write a model with the two periods pooled together as follows:

$$log(W_i^s) = \beta^s X_i^s + \epsilon_i^s$$

The difference in average log wages between 1995 and 2014 can be written as:

$$\begin{aligned} E(log(W_{i2014}^s)) - E(log(W_{i1995}^s)) &= E(X_{i2014}^s)'(\beta_{2014}^s - \beta^s) + E(X_{i1995}^s)'(\beta^s - \beta_{1995}^s) \\ &+ (E(X_{i2014}^s) - E(X_{i1995}^s))'\beta \end{aligned}$$

Let $\bar{j}$ denote sample average of variable j, $\hat{\beta}$ denote ordinary last squares estimates of parameter β, 1 and 2 be 1995 and 2014 respectively, and w be $log(W)$. We can write the sample estimate of the above expression as:

$$(\bar{w}_2^s - \bar{w}_1^s) = \underbrace{\bar{X}_2^s(\hat{\beta}_2^s - \hat{\beta}^s) + \bar{X}_1^s(\hat{\beta}^s - \hat{\beta}_1^s)}_{\hat{\Delta}_p^s\text{: pay structure effect}} + \underbrace{(\bar{X}_2^s - \bar{X}_1^s)\hat{\beta}}_{\hat{\Delta}_c^s\text{:composition effect}} .$$

where $\hat{\Delta}_p^s$ and $\hat{\Delta}_c^s$ are, respectively, the estimate of the pay structure effect and the composition effect for group s. The first term, the composition effect, reflects the changes in average wages of the group s due to changes in the return to covariates Xs. The second term, the composition effect, estimates the change in average wages that is due to changes in the distribution of covariates Xs.

We are interested in studying how changes in returns and composition of occupations are related to changes in average wages. Let X_j be composed by 82 occupational dummies from ISCO-88-3 digit level classification, and X_g be other covariates. Given the linearity of both the pay structure and the composition effect, it is easy to get the part of the effect driven by occupations. It is possible to write:

$$\hat{\Delta}_p^s = \underbrace{\sum_{j=1}^{j=82} \bar{X}_{2,j}^s(\hat{\beta}_{2,j}^s - \hat{\beta}_j^s) + \bar{X}_{1,j}^s(\hat{\beta}_j^s - \hat{\beta}_{1,j}^s)}_{\textbf{Occupations}} + \underbrace{\sum_g \bar{X}_{2,g}^s(\hat{\beta}_{2,g}^s - \hat{\beta}_g^s) + \bar{X}_{1,g}^s(\hat{\beta}_g{}^s - \hat{\beta}_{1,g}^s)}_{\textbf{Other covariates}}$$

$$\hat{\Delta}_c^s = \underbrace{\sum_{j=1}^{j=82} (\bar{X}_{2,j}^s - \bar{X}_{1,j}^s)\hat{\beta}_j}_{\textbf{Occupations}} + \underbrace{\sum_g (\bar{X}_{2,g}^s - \bar{X}_{1,g}^s)\hat{\beta}_g}_{\textbf{Other covariates}}$$

Table 12 shows the results of the decomposition. The decomposition is performed for the entire workforce and for each educational group separately. In column (1) the vector of personal characteristics X only incorporates occupational dummies. In Column (2) X also contains dummies for combinations of gender, 4 age groups, and 6 levels of education, to account for changes in these characteristics for the entire workforce and within each educational group.

The results of the decomposition are stated as follows. For the entire workforce, the average log wage increased 0.382 between 1995 and 2014, mainly due to a general raise in wages within each occupation. The composition effect due to a change in occupations is small (only explains 11 percent of the increase in wages), while most of the increase in wages comes from the pay structure effect. This is consistent with Section 4 where we stated that the occupational composition of the Brazilian economy did not change much during the period 1995-2014. The increase in average wages is mainly due to a generalized increase in wages across all occupations and a decline in the occupational premiums, reflected in the negative effects of occupational premiums in the pay-structure effect and the large positive change in the constant of the regression. These results are robust to incorporate other covariates into the decomposition (column (2)).

The decomposition exercise shows heterogeneous results for workers with different educational level. For low educated workers, the average wages increased due to small but negative composition effect and a large and positive pay structure effect. This indicates that they occupational composition deteriorated (they are more concentrated in low wage in 2014 when compared to 1995), but the wage level had a generalized increase (reflected in the constant) that more than compensate for the occupational downgrading. Note that the sign of the occupation term in the pay structure effect is negative, indicating that the occupational premiums (with respect to the constant) declined. For medium educated workers, the composition effect was large and negative, and it was not compensated by a small and positive pay structure effect. It reflects that the occupational composition of medium educated workers deteriorated, and even when there was an increase in wages within each occupation it was not enough to compensate for the occupational downgrading. In the case of high educated workers, both the composition and the pay structure effect are negative, indicating that not only there was an occupational downgrading during this period but also that wages within each occupation decline as well. The fact that the occupation term in the pay structure effect is positive shows that the occupational premiums increase (with respect to the constant that largely diminished). This is because wages fall for high educated workers in low wage occupations. The results for each educational level practically do not change when other covariates are incorporated into the decomposition (column (2)).

A better intuition of these results can be obtained from Figure 18. The figure displays a locally weighted smoothing regression of the changes in log wages between 1995 and 2004 on the ranking of occupations for all workers in panel (a) and for workers with different educational level in panel (b). From the first panel, it is clear that wages increased more in low wage occupations and occupational premiums declined. While the smoothing regression in the second panel shows that wages increased for all occupations in the case of low educated workers, only increased in occupations with a low ranking for medium educated workers, and wages declined for almost all occupations for high educated workers (except those at the top), and this decline was more pronounced in low ranked occupations.

In summary, this section decomposes the changes in wages during the period 1995-2014 between an occupational composition effect (changes in the occupational structure of employment) and a pay structure effect (changes in wages within occupations). We find that for all workers the composition effect is small, consistent with small changes in the occupational composition of the economy; while the increase in wages is driven by higher wages at each occupation, especially those that started with a lower wage. We also find that all educational levels experienced a negative occupational composition effect, consistent with conditional occupation downgrading. For low educated workers wages within each occupation largely increase, compensating for the negative composition effect, and average wages increased. For medium educated workers, wages increase only in a small number of occupations and it was not enough to compensate for the composition effect, so average wages decreased. For high educated workers, wages declined in each occupation, reinforcing the negative composition effect, and average wages fell.

These results are in line with to those of the model presented on Section 5 for the case that I_H declines, as was the case in the calibrated estimation in Section 7.1.[52] In the model, an educational expansion generates a deterioration in occupations conditional on education, which is larger for medium educated workers; this is exactly what the composition effect indicates in the data. The model predicts a large increase in the price of occupations where low educated workers are employed, an mild change in prices of occupations performed by medium educated workers (can increase or decrease, but always between the changes in prices of low and high educated workers), and a sharp decline in the prices of occupations performed by high educated workers. These predictions are consistent with the heterogeneous pay structure effect that is observed in the data.

[52]In this discussion, I assume the tasks in the model can be directly interpreted as occupations. See Section 3 for a detail discussion on occupations being interpreted as tasks

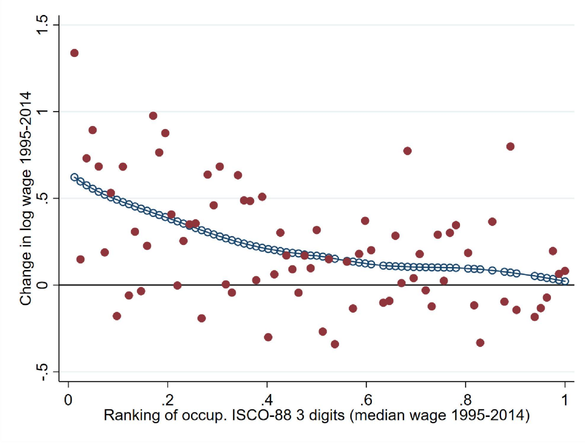
Table 12: Oaxaca decomposition of changes in mean wages

	All workers		Low educated		Medium Educated		Highly Educated	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Total change in mean ln wages	0.382	0.382	0.333	0.333	-0.038	-0.038	-0.151	-0.151
Decomposition								
Composition effect	0.041	0.200	-0.045	0.016	-0.123	-0.084	-0.055	-0.049
Occupations	0.041	0.023	-0.045	-0.031	-0.123	-0.117	-0.055	-0.052
Gender x edu x age		0.178		0.048		0.032		0.003
Pay structure effect	0.341	0.181	0.379	0.317	0.084	0.046	-0.096	-0.101
Occupations	-0.568	-0.363	-0.487	-0.338	-0.585	-0.616	0.400	0.442
Gender x edu x age		0.302		-0.399		0.169		0.068
Constant	0.909	0.242	0.865	1.054	0.670	0.493	-0.496	-0.611
Relative importance								
Composition/total	0.11	0.53	-0.14	0.05	3.21	2.21	0.36	0.33
Structure/total	0.89	0.47	1.14	0.95	-2.21	-1.21	0.64	0.67
Occupations compositions/total	0.11	0.06	-0.14	-0.09	3.21	3.06	0.36	0.35
Occupations structure/total	0.89	-0.32	1.14	2.15	-2.21	3.21	0.64	1.12

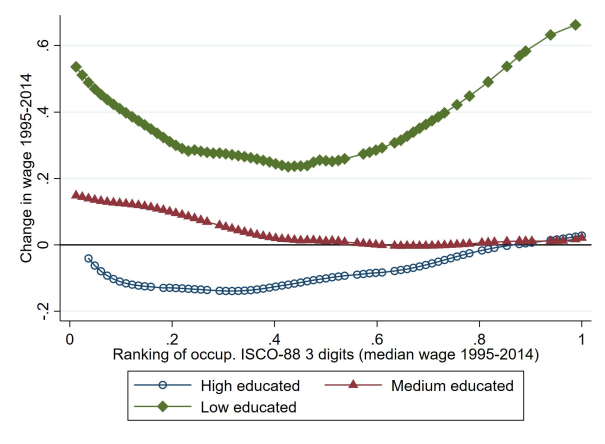
Note: The table presents the Oaxaca decomposition of average log wages between a composition effect and a pay structure effect, which are in turn decompose into the effect of occupations and those of other covariates. Column (1) only considers occupational dummies according to ISCO-88-3 digit level. Column (2) incorporates group dummies for combinations of gender, 4 age groups and 6 levels of education (GenderxEduxAge).

Figure 18: Changes in log wages by occupation. Period 1995-2014

(a) All Workers



(b) By Educational Level



Note: The figure plots a locally weighted smoothing regression of the changes in log wages between 1995 and 2004. Occupations are ranked as in Figure 3. Source: Author's calculation based on PNAD 1995 and 2004.