The Agricultural Wage Gap: Evidence from Brazilian Micro-data*

Jorge A. Alvarez†
Princeton University

November 16, 2015

Abstract

A key feature of developing economies is that wages in the agricultural sector are significantly below those of other sectors. Using a panel data set on the universe of formal workers in Brazil, I use information on workers that switch sectors to decompose the drivers of this inter-sector gap. I find that most of the gap between sectors is explained by unobservable differences in the skill composition of workers, as opposed to differential pay of workers with similar skills. The evidence speaks against the existence of large short-term wage gains from the reallocation of workers out of agriculture and favors recently proposed Roy models of inter-sector sorting as drivers of lower average wages in agriculture. A calibrated model of worker sorting can account for 66% of the wage gap observed in 1996 Brazil and a share of both the wage gap decline and the diminishing worker participation in agriculture observed during the period between 1996 and 2013.

Key words: Wage gaps, Productivity Gaps, Inters-sector Gaps, Human Capital, Structural Transformation, Agriculture, Sorting, Brazil.

*I am grateful for the insightful and generous advice of Richard Rogerson throughout this project. I am also thankful to Stephen Redding, Elhanan Helpman, and Marc Muendler for granting me access to part of the data during the earlier stages of this project. In addition, I want to give special thanks to Niklas Engbom and Chris Moser, whose collaboration in working with Brazilian data has been invaluable. I appreciate the input of Angus Deaton, Douglas Gollin, Ilyana Kuziemko, David Lagakos, Ben Moll, Stephen Redding, Tom Vogl, Oleg Itskhoki, as well as seminar participants of the Princeton Macroeconomics lunch seminar, the Princeton Research Program for Development Studies and Center for Health and Wellbeing lunch seminar, and the Princeton Public Finance Working Group for their helpful comments on my work on both agriculture and Brazilian wage differences.

†Mail: Economics Department, Princeton University. 001 Fisher Hall. Princeton, NJ 08544. Email: jaalvare@princeton.edu. Website: scholar.princeton.edu/jalvarez.
1 Introduction

A key feature of developing economies is that wages in the agricultural sector are significantly below those of other sectors. Additionally, these economies have most of their workforce in the agricultural sector. These two observations motivate a literature dating back to Lewis (1955) and Rostow (1960) that views the exit of workers out of agriculture as a fundamental mechanism of development. The body of work on agricultural development and inter-sector differences, however, has not completely settled the question of why so many workers stay in agriculture in spite of better wages being paid in other sectors. One possibility is that some barrier prevents the movement of workers across sectors, in which case wage gaps between agriculture and other sectors indicate unexploited potential gains from the reallocation of workers out of agriculture. A second possibility is that workers in agriculture are characteristically different from those in non-agriculture, in which case wage gaps would not be evidence of potential wage gains. The objective of this paper is to shed light on which of these possibilities is a more likely explanation of the agricultural wage gap.

A challenge in exploring this question is assessing the role of unobserved worker characteristics. For instance, if an agricultural worker and a non-agricultural worker with the same observable characteristics (e.g. age and education) earn different wages, it is hard to distinguish whether the two sectors simply have differential pay for similar workers or the two workers are in fact different due to some unobserved characteristic. This paper assesses the role of unobserved characteristics by using panel micro-data on earnings and worker characteristics covering all sectors of the Brazilian economy from 1996 to 2013. The use of panel data is an improvement on the literature on agricultural wage gaps in developing countries, which has typically relied either on the estimation of structural models to match country-level moments or on the analysis of heterogenous cross-sectional surveys from a sample of countries. Specifically, the panel dimension of the data allows me to control for differences in both observable and fixed unobservable worker characteristics. Information on workers that switch between sectors (from now on referred to as ‘sector-switchers’) can therefore be used to distinguish whether the wage gap between agriculture and non-agriculture reflects differential pay of similar workers in the two sectors or, alternatively, whether the gap is due to differences in the composition of worker characteristics in each sector.

The main empirical finding of this study is that workers who transition out of agriculture

---

1In a sample of developing countries studied by Vollrath (2014), the median average wage ratio between agriculture and manufacturing was 1.6. This is 1.9 when comparing agriculture against services. In the sample of countries studied by Herrendorf and Schoellman (2014), the median ratio between agriculture and the rest of the economy is 2.0.
experience limited compensation gains when compared to the overall gap in mean earnings between agriculture and other sectors. I conclude that the agricultural wage gap does not appear to be driven by differential pay of similar workers once fixed unobservable characteristics are controlled for. Instead, the largest share of the agricultural wage gap is explained by differences in the composition of worker characteristics in each sector. In addition, I find that the wage gap between agriculture and other sectors in Brazil declined significantly from 1996 to 2013 as the economy grew richer. This reduction is similar when comparing agriculture to both services and manufacturing, and it coincided with a decline in the share of workers employed in agriculture — from 25% to 14%. Moreover, this decline does not appear to be driven by changes in educational attainment or country demographics. In fact, I find that age and education explain only a small share of the large wage gap in Brazil during the late 1990s, and that differences in the composition of these variables between sectors drove only a small share of the decline during this period. Most of the decline is driven by compositional changes in the distribution of fixed unobservable worker characteristics.

Both the limited wage gains experienced by workers who switch out of agriculture and the importance of worker composition differences between sectors pose a challenge for an agricultural wage gap model. Such a model must generate large declining wage gaps that do not result in large wage gains among sector-switchers. Building on the work of Roy (1951), a recent literature has proposed worker sorting as a possible explanation that is consistent with this pattern. In particular, Lagakos and Waugh (2013) and Young (2014) illustrate how workers with sector-specific skills can sort themselves into different sectors to generate large wage gaps. In this type of model, each worker faces a choice between two idiosyncratic wages in agriculture and non-agriculture. Workers with a comparative advantage in non-agriculture choose to work in that sector, and this generates a wage gap relative to workers who find it advantageous to stay in the agricultural sector.

To test the explanatory power of this mechanism, I build on the sorting model proposed by Lagakos and Waugh (2013) and test whether a calibrated model that targets micro-moments from sector-switchers can generate wage gaps of the magnitudes observed in Brazil in the late 1990s. I find that a large share of the gap level (66%) can be generated by this model. In addition, when calibrating the model to match moments of the data from 1996 Brazil and estimating the dynamics of the wage gap over time as the country grows, I ask whether growth can lead to a wage gap decline due to the exit of agricultural workers. I find that this mechanism can account for about a fourth of the wage gap reduction observed in Brazil during the period between 1996 and 2013.

The rest of the paper is structured as follows. Section 2 provides a literature review that
relates this paper to the labor literature on inter-sector wage gaps and the macroeconomic literature on both wage and output per worker gaps between sectors. Section 3 describes the datasets to be used in the empirical estimations. Section 4 describes the magnitude and evolution of the wage and productivity gaps in Brazil as well as the decline in the share of workers employed in the agricultural sector. Section 5 assesses the role of observables, unobservables, and differential pay of similar workers in explaining the gap. Section 6 describes the mechanics and calibration of an economy where workers sort across sectors, as well as the power of the worker sorting mechanism in explaining the agricultural wage gap magnitude and its decline. Section 7 concludes.

2 Literature Review

Most studies show that large inter-sectoral wage gaps persist even after controlling for educational attainment and other worker observables. The remaining gap stems from either differential pay of similar workers or, alternatively, sector differences in the composition of both observable and unobservable worker characteristics.

U.S. labor studies have explored this distinction with mixed results. Using matched data from the Consumer Population Survey (CPS), Krueger and Summers (1988) argue that unobservable worker characteristics cannot explain much of the difference in wages between sectors. On the other hand, Murphy et al. (1987, 1990) also use the CPS and conclude that industry switchers receive only 27% to 36% of the total industry differential and thus nearly two-thirds of inter-sector wage gaps can be attributed to differences in the composition of worker characteristics in each sector. Also using US data, Gibbons and Katz (1992) find limited evidence for differential pay of similarly-skilled workers between sectors and instead highlight the role of differences in the composition of observable and unobservable characteristics.

International studies on developing countries have also highlighted the role of differences in observable and unobservable worker characteristics in explaining the gap. Vollrath (2014) finds that large wage differences exist between workers after controlling for observed human capital in a set of 14 countries. He explores whether these wage gaps could be the result of distortions that prevent workers from being paid the value of their marginal product in each sector. Using a misallocation framework similar to Hsieh and Klenow (2009), Vollrath (2014) estimates that potential gains from eliminating distortions and eradicating human capital misallocation are less than 5% in developing countries. If misallocation is not important, this implies that differences in the composition of worker productivity are likely to be more
important drivers of the gap. Similarly, using a different sample of countries, Herrendorf and Schoellman (2014) regress wages on observables allowing for returns on observables to vary by sector. They conclude that most of the wage gap between agriculture and other sectors can be accounted for by differences in workers’ human capital — and sector-specific differential returns— present in each sector.

However, because of data constraints, these studies are limited to the comparison of a diverse collection of cross-sectional surveys. This prevents rigorous empirical testing of whether differences attributed to unobservable characteristics or differential human capital returns could in fact be the result of other forces producing differential pay of similar workers. Mobility frictions and compensating differentials, for instance, are two alternative explanations consistent with both the differential returns on observables estimated by Herrendorf and Schoellman (2014) and the residual wage differences reported by Vollrath (2014). By using a panel dataset where workers are observed as they switch across sectors, the current study overcomes the limitations of cross-sectional data and distinguishes the role of fixed unobservable characteristics from alternative stories of differential pay.

The study of wage gaps is also closely related to the study of output per worker gaps between agriculture and other sectors. Kuznets (1971), Caselli (2005), Restuccia, Yang and Zhu (2008), among others, have argued that a large share of income differences across countries is explained by labor productivity gaps between agriculture and other sectors. However, focusing on output per worker, even in advanced countries, risks exposure to important sources of measurement errors. For instance, Gollin, Parente, and Rogerson (2004) suggest that unaccounted home production understates agricultural output and Herrendorf and Schoellman (2014) point out that errors in value added measurement muddy comparisons of worker productivity across US States. Partially as a result of this, the role that both observed and unobserved human capital play in explaining these output per worker gaps is still an open debate. Herrendorf and Schoellman (2014) argue that human capital accounts for most of the output per worker gap between agriculture and other sectors in the US and other selected countries. In contrast, Gollin, Lagakos, and Waugh (2013) argue that human capital —along with adjustments to labor supply— account for only about a half of the gap in the developing countries they study. Focusing on wages avoids many of the problems with the measurement of differences between agriculture and the rest of the economy. Though wages and output per worker are not equivalent measures of labor productivity, the results of this paper can speak to some of the debates about the role of differences in worker composition on inter-sector gaps explored by this literature.

Beyond establishing the role of worker characteristics in explaining the inter-sector gaps,
a second objective of the literature is to establish which mechanisms are behind compensation
and output per worker differences. Two main types of mechanisms are relevant to this
study. The first are distortions that create wedges in marginal productivity of labor between
sectors. These distortions can include scale effects that impact the allocation of resources
across agricultural firms (Adamopoulos and Restuccia (2011), Donovan (2012)) or barriers
that prevent the free flow of capital and workers (Restuccia et al. (2008), Herrendorf and
Teixeira (2011)). Distortions that prevent marginal labor products to equalize have also
been studied at the firm level by Restuccia and Rogerson (2008) and Hsieh and Klenow
(2009), who highlight their greater importance in developing countries. To the extent that
these distortions are also present between sectors—and workers are not freely mobile—the
mechanisms generating productivity gaps can be related to the agricultural wage gap.

A second type of mechanism highlighted by Young (2014), and Lagakos and Waugh
(2013) portrays wage gaps as the result of sector differences in worker skill composition.
Lagakos and Waugh (2013) illustrate how such skill differences can be the result of an equi-
librium outcome. In their model, workers sort themselves to the sector where they are most
productive. This process induces differences in the composition of worker skills employed
by each sector, and this in turn generates a gap in mean wages paid in agriculture relative
to non-agriculture. Importantly, the agricultural gap in this context isn’t the result of any
additional distortions that induces differential pay of similar workers. Building on this idea,
Young (2014) uses cross-sectional surveys from developing countries to show how migration
is consistent with rural-urban consumption driven by the sorting of workers. Though his
focus is on consumption, his findings are also consistent with agricultural wage gaps gener-
ated by the sorting of workers with different unobservable skills. The mechanism proposed
by this paper—which is also supported by the empirical results to be presented—belongs
to this family of sorting models, where the agricultural wage gap is ultimately driven by
compositional differences in worker characteristics.

3 Data description

Two main databases are used. The first is the set of Brazilian household surveys from the
Pesquisa Nacional por Amostra de Domicílios (PNAD) from 1996 to 2013. This contains a
representative sample of households covering all of Brazil. The survey includes both formal
and informal workers and records demographic and employment-status characteristics as well
as monthly earnings for all members of a household. In this paper, this data is used to show
trends in sector employment shares among all workers, including both formal and informal,
during the period of study. In particular, I establish that the trends and magnitudes in inter-sector pay differences in Brazil among all workers are similar to the ones observed among formal workers. Data from PNAD is also used to compute the total number of workers in each sector and—in combination with the national accounts recorded by the *Instituto Brasileiro de Geografia e Estadística* (IBGE)—value added per worker for each year and sector. Due to the cross-sectional nature of the surveys, however, individuals cannot be followed over time in the PNAD and I am therefore unable to control for worker unobservable characteristics using data on both formal and informal workers. For this reason, most empirical decompositions in this paper focus on formal sector data which is now described in greater detail.

Data on formal workers comes from the *Relação Anual de Informações Sociais* (RAIS), which is administered by the Brazilian Ministry of Labor and Employment. This database is constructed from a mandatory annual survey filed by all formally registered firms in Brazil and contains earnings, occupation and demographic characteristics of workers as reported annually by their employers.\(^2\) Importantly, each worker in the data has a unique and time-invariant worker ID that does not change as a worker switches employers. This feature of the data allows me to follow individuals over time and create a panel of the universe of employed formal workers across all sectors. In addition, each worker is linked to their employing firm, which also has a unique and time-invariant ID. This allows me to link workers to their respective sectors, and identify transitions between sectors.\(^3\) The data covers the period from 1996 to 2013.\(^4\)

The RAIS dataset reports average monthly gross labor earnings including regular salary payments, holiday bonuses, performance-based and commission bonuses, tips, and profit-sharing agreements as well as the start and end month of the job. To account for heterogeneity in the duration of job-spells, I divide annual earnings by the number of months worked at each job within a particular firm. A worker might have multiple spells in a year if he or she switched employers during the year or worked multiple jobs, but on-the-job earnings changes within a year are not recorded. To standardize the dataset at an annual level, I restrict attention to a unique observation per worker-year by choosing the highest-paying among all employment spells in any given year. One limitation of the data is that it does
\[\text{\footnotesize \(^2\)It is common practice for businesses to hire a specialized accountant to help with the completion of the RAIS survey to avoid fines levied on late, incomplete, or inaccurate reports, which makes the quality of the data superior to household surveys.}\]
\[\text{\footnotesize \(^3\)IDs available are anonymized to protect the identity of both workers and firms.}\]
\[\text{\footnotesize \(^4\)Though earlier years are available for a large subset of Brazilian workers, the lack of universal coverage in earlier periods can be particularly problematic in studying transitions out of agriculture. Hence, the analysis is restricted to this later period.}\]
not contain hours worked for all of the period of study. Therefore, I cannot transform the measure of monthly earnings into a measure of hourly wages. The implications of interpreting results from sector earnings differences as wage gaps is discussed in section 5.2. For the rest of the paper, differences in the measure of monthly earnings described between sectors will be referred to as wage gaps.

The dataset also contains the age and educational attainment of each worker. Both of these variables are classified by groups. In particular, age groups of 18-24, 25-29, 30-39, 40-49, and 50-54 years are used.\footnote{Workers over 55 years old are excluded to avoid biases stemming from workers close to retirement age.} Similarly, educational levels are classified into less than high school, high school, some college education, and completed college education. In all regression specifications utilizing age and education as explanatory variables of the wage gap, a full set of age and education interacted dummies is used.

Finally, to identify the employment sector and occupation of workers, classification is based on categories from the IBGE. Both the industry and occupation classification system changed during the period of study. Here, I use conversion tables provided by IBGE to standardize classification between different years and choose categories for both occupations and sectors coarse enough in order to avoid potential biases arising from mechanical changes in the classification system over time. The three sectors used are Agriculture, Manufacturing (including energy and mining), and Services. Occupation categories used are at the 3-digit disaggregation level.

Due to imperfect matching of all categories within a sector and occupation classification system, I exclude firms with inconsistent sector classifications so that sector switchers are not incorrectly specified. I also exclude individual observations that have either firm IDs or worker IDs reported as invalid as well as data points with missing earnings, dates of employment, educational attainment or age. Together, these basic cleaning procedures drop less than 0.5% of the original population. The low percentage of missing responses and inconsistencies are indicative of the high quality of this administrative dataset.

Table 1 provides key summary statistics for the RAIS data for three sub-periods: 1996-2001, 2002-2007, and 2008-2013. Some features of the data are worth noting. The first is that the number of workers increases substantially over time from 49 million workers in 1996-2001 to 80 million in 2008-2013. This rise is mainly the result of two forces: population growth and an increase in formality in Brazil. A second observation is that education is quite different in agriculture in Brazil relative to other sectors. In 1996-2013, for instance, only 5% of formal workers in the agricultural sector had a high school degree and 1% had completed college, relative to 34% and 10% in other sectors. During 1996-2013, educational
attainment substantially improved partially as a result of educational reforms in the late 1990s and the rise of social programs in the 2000s. In contrast, the age distribution in each sector did not change substantially. The explanatory power of age and education will be one of the focal points of the analysis. Finally, though earnings between agriculture and other sectors are quite different, there are virtually no gaps in earnings when comparing services against manufacturing in all periods. This motivates the dual economy focus of this paper: explaining the gaps between agriculture and all other sectors in the economy.

4 The magnitude and evolution of the agricultural gap in Brazil

Differences in pay between agriculture and other sectors were significantly reduced in Brazil during the last two decades. The ratio of mean earnings between non-agriculture and agriculture among all workers (both formal and informal) in the economy—as measured by the PNAD household surveys—declined from 2.1 in 1996 to 1.7 by 2013. As discussed above, the main contributions of this paper hinge on the use of the panel structure of the data so that workers can be followed over time. Since this feature is only available for formal workers, the rest of the paper will focus on formal sector data. Similarly to the overall economy, formal workers exhibit a very similar decline in the ratio of mean earnings between non-agriculture and agriculture from 2.3 in 1996 to 1.6 in 2013 (Figure 1). Moreover, the magnitude of the gap and its decline has been similar when comparing agriculture to both services and manufacturing individually. In contrast to the differences between agriculture and non-agriculture, mean earnings in the two non-agricultural sectors were close to equal throughout this period.

Another feature of the data is that the agricultural wage gap is present throughout the earnings distribution. Figure 3 shows the ratio of earnings percentiles in agriculture and non-agriculture. Percentiles are here defined by the ranking of workers within each sector. There is a pattern, with the top earners in the agricultural and non-agricultural sectors being further apart than the bottom earners in the two sectors. The differences, however, are still significant across all percentiles and it is not the case that wage gaps are a phenomenon that is only applicable to certain parts of the earnings distribution. Furthermore, when looking at the evolution of these ratios over time, the decline in compensation differences does not appear to be driven by the catch up of only the poorest or richest parts of the distribution of agricultural workers.

In addition, the wage gap decline was accompanied by a similar decline in the value
<table>
<thead>
<tr>
<th>Period</th>
<th>Sector</th>
<th># Worker-years</th>
<th># Unique Workers</th>
<th>log(Earnings)</th>
<th>Education</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>1996-2001</td>
<td>Agriculture</td>
<td>9.5</td>
<td>3.1</td>
<td>6.22</td>
<td>0.60</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>34.3</td>
<td>9.4</td>
<td>6.85</td>
<td>0.82</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>131.4</td>
<td>36.5</td>
<td>6.84</td>
<td>0.84</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>176.0</td>
<td>49.2</td>
<td>6.81</td>
<td>0.84</td>
<td>1.64</td>
</tr>
<tr>
<td>2002-2007</td>
<td>Agriculture</td>
<td>13.0</td>
<td>3.9</td>
<td>6.38</td>
<td>0.51</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>43.3</td>
<td>11.0</td>
<td>6.84</td>
<td>0.75</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>172.3</td>
<td>45.0</td>
<td>6.84</td>
<td>0.77</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>229.6</td>
<td>60.2</td>
<td>6.82</td>
<td>0.77</td>
<td>1.83</td>
</tr>
<tr>
<td>2008-2013</td>
<td>Agriculture</td>
<td>13.8</td>
<td>3.8</td>
<td>6.66</td>
<td>0.47</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>Manufacturing</td>
<td>56.3</td>
<td>13.6</td>
<td>7.01</td>
<td>0.68</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>246.0</td>
<td>62.4</td>
<td>6.98</td>
<td>0.71</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>317.7</td>
<td>80.2</td>
<td>6.98</td>
<td>0.71</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Note: Number of workers and worker-years are in millions. Earnings refer to average monthly earnings in real terms (Using 2013 Reais). Education levels are defined as 1 = Primary or middle school or no education, 2 = high school 3 = some college education and 4 = college completed. Age is in years.
added per worker gap. Figure 2 shows how the between-sector difference in gross domestic product per worker as measured by the national accounts declines over the 1996-2013 period. Similar to the earnings pattern, the decline is large when comparing agriculture against both manufacturing and services. Unlike earnings, however, the differences and the magnitude of the decline is much larger when looking at the agriculture-manufacturing gap than when looking at the agriculture-services one. This is expected due in part to the natural differences in capital intensities between services and agriculture. These differences notwithstanding, the qualitative pattern of pay and value added per worker gaps is qualitatively similar. Importantly, this reduction in sectoral inequities occurred during a period where yearly real GDP growth averaged 2.7%, as the country transitioned out of a period of macroeconomic instability and hyperinflation into a period of technology modernization and growth.⁶ The interrelation of growth, productivity, and the decline in inter-sector gaps will be central to our analysis of mechanisms in section 6.

The magnitudes of both the wage and value added per worker gaps between agriculture and other sectors are large when compared with other estimates in the literature. In 1996, the magnitude of the value added per worker gap between agriculture and other sectors is 5.3, which is greater than the maximum found by Herrendorf and Schoellman (2014) in their 12 country sample and just below the mean gap reported in Gollin, Lagakos and Waugh (2013) for the poorest quartile of countries in their 151 country sample. By 2013, after a cumulative real output growth of 61%, the value added per worker gap is 2.4. This estimate is similar to the median of 2.3 in the Herrendorf and Schoellman (2014) sample and closer to the 2.0 mean of the richest 25% of countries in the Gollin, Lagakos and Waugh (2013) sample. When compared to the cross-sectional evidence, Brazil appears to have endured a significant transformation during the period of study.

In terms of the wage gap between agriculture and other sectors, Brazil’s 1996 wage gap of 2.3 is above the median of 2.0 from the Herrendorf and Schoellman (2014) sample. By 2013, it falls below the sample’s mean to 1.6. Figure 4 shows how these gaps compare to the list of 15 developing countries studied by Vollrath (2014). Brazil’s 1996 gap between agriculture and manufacturing would rank third highest, just below Ecuador in that sample. When comparing agriculture vs services, the rank would be 5th, just above Indonesia. In contrast, Brazil’s 2013 gap levels with respect to manufacturing and services would rank 8th and 11th, respectively. Though the data on Brazil is not entirely comparable to the wage data from other countries’ surveys, the significant move down the ranking of countries

⁶Bustos and Bruno (2015) explain some of the agricultural modernization of the agricultural sector in Brazil.
suggest that Brazil’s decline cannot be described as an insignificant change.

In parallel to the closing of both output per worker and wage gaps, Brazil also endured a substantial transformation of the employment structure. The workforce composition based on household surveys is shown in Figure 5. The economy employed 25% of the labor force in agriculture in 1996, which declined to 14% by 2013. Manufacturing employed 13-15% throughout this same period, and services increased from 61% to 72%. Among formal workers, a similar pattern is observed and the share of workers in agriculture has declined from 5% to 3.6% since 1996. Though the population of formal workers is much smaller than the universe of workers in agriculture, the magnitude of the wage gap also shows a declining pattern in the share of labor employed in agriculture. The interrelation between the movement of workers out of agriculture and the agricultural wage gap will be considered in section 6, when mechanisms behind the gap’s decline are discussed. First, a statistical decomposition of the agricultural wage gap is conducted using the panel structure of the data on formal workers.

Figure 1: Wage gap in Brazil

(a) Formal workers

(b) All workers

Note: The wage gap is calculated as the ratio in average labor monthly earnings between agriculture, manufacturing and services as classified by the IBGE. Data on formal workers comes from the Relação Anual de Informações (RAIS). Data on all workers (both formal and informal) comes from the PNAD household surveys.
Figure 2: **Value added per worker gap in Brazil**

![Value added per worker gap in Brazil](image)

Note: Value added per worker gaps are constructed from national accounts available from IBGE and labor statistics from the Pesquisa Nacional por Amostra de Domicílios (PNAD).

Figure 3: **Gaps in Brazil by percentile**

(a) Agriculture vs manufacturing  
(b) Agriculture vs services

![Gaps in Brazil by percentile](image)

Note: Difference in the means of log earnings between sectors for formal workers are presented. Each line corresponds to the difference between each percentile group in the two sectors.
Figure 4: Wage gaps in Brazil vs other countries

Note: Data for Brazil comes from PNAD and national accounts from IBGE. For other countries, estimates are constructed based on cross-country data from Vollrath (2014).

Figure 5: Workers by sector

(a) Formal workers  (b) All workers

Note: Share of total employed workers. Formal worker data is from RAIS. Data on all workers is from PNAD household surveys.
5 Sources of the agricultural gap

We now turn to explore what drives the wage gap between agriculture and other sectors. Three possible alternatives are considered. The first are differences in the composition of observable human capital as measured by age and education. The second are differences in the distribution of fixed unobserved worker characteristics between sectors. Finally, the third alternative is the presence of mechanisms that induce differential pay of similar workers employed by different sectors. Inter-sector mobility frictions, sector-specific rent-sharing agreements, and compensating differentials are some of the mechanisms that fit this third category. This section argues that the first two alternatives, where the gap is driven by compositional differences in worker characteristics, explain most of the agricultural wage gap and its decline.

5.1 Human capital

Differences in human capital introduce heterogeneity in the productivity of workers which, in a standard competitive environment, should translate into wage differences. Table 1 indeed shows differences in education between sectors, with agriculture workers being on average less educated than their peers in services and manufacturing. To the extent that these characteristics determine human capital, these differences can potentially explain part of the agricultural wage gap.

There are two margins on which human capital influences the wage gap. On one hand, human capital can be lower in one sector than the other. On the other hand, even if the composition of human capital is the same in the two sectors, the returns to human capital might be different in the two sectors. I first assess whether compositional differences in human capital, as measured by age and education, can account for a substantial share of the gap by estimating the following model for each year $t$.

$$\log(w_{ist}) = F_t(\text{education}_{ist}, \text{age}_{ist}) + \epsilon_{ist}$$

Here, $\text{education}_{ist}, \text{age}_{ist}, w_{ist}$ are the wage, education level, and age of worker $i$ in sector $s$ in year $t$. To impose minimal restrictions on how age and education influence wages, the mapping of education and age to wages is specified as $F_t(\text{education}_{ist}, \text{age}_{ist}) = \sum_{a,e} 1(\text{age}_{ist} = a, \text{edu}_{ist} = e) \times \beta_{aet}$. Thus, the specification allows full flexibility in terms of both age and education, and this relationship can vary in every year of the sample.

For the rest of the paper, I will define the wage gap as the mean difference of log earnings.
with respect to agriculture. Specifically, the gap between sector $s'$ and agriculture is defined as

$$\Delta_{s'} E(\log(w_{ist})) \equiv E(\log(w_{ist}) | s = s') - E(\log(w_{ist}) | s = a)$$

where the possible values for sector $s'$, $\{a, m, s\}$, refer to agriculture, manufacturing and services respectively. The focus on additively separable mean log-wage gaps is done to simplify the presentation of the log-linear models to be studied.

Figure 6 shows the decomposition of the mean log difference into two parts: a component due to age and education and another due to the residual. There, we can see that the effect on wages from age and education differences between agriculture and other sectors have remained roughly constant throughout 1996-2013. When comparing agriculture vs manufacturing, these observable characteristics explain a nearly constant .09 log points of the gap. When comparing agriculture vs services, observables matter more and wage gaps due to age and education have averaged .24 log points. Overall, age and education differences accounted for 10% to 26% of the wage gap level during the period. The results show that most of the wage gap level is largely driven by factors not accounted by compositional differences in age and education alone.

Moreover, the decline in the wage gap cannot be entirely attributed to changes in education and the distribution of age in each sector. When comparing manufacturing and agriculture, the stability of the gap due to age and education shown in Figure 6 contrasts the decline in the overall wage gap. When comparing services and agriculture, age and education explain some of the decline, but the flatter pattern of this component relative to the total gap decline indicates that this reduction is not sufficient to explain the entire decline in the gap between these two sectors. Similarly to the wage gap level, the decline does not appear to be explained by either demographic changes or educational attainment.
Figure 6: Gap in mean log earnings between agriculture and other sectors due to age and education

(a) Agriculture vs Manufacturing
(b) Agriculture vs Services

Figure 7: Mean difference in log earnings relative to agriculture by educational attainment and age

(a) By education
(b) By age

The approach above, however, does not allow for differences in pay within each education-age group across sectors. Figure 7 shows that large average wage differences by education and age groups exist, with older workers gaining significantly less in agriculture relative to other sectors and workers in each age and education group being paid less than their comparable peers in non-agriculture. The difference in average pay for worker characteristics in each sector may reflect differential returns to education and experience by sector. For instance, worker with a high school degree might be more productive in manufacturing and services
than in agriculture due to the availability of jobs that require this level of educational attainment.

The question is then to what extent do composition vs differential pay of each education-age group can explain the overall gap. In order to separate these components, I conduct a Oaxaca decomposition with agricultural workers as the reference group (Oaxaca (1973)). For notation simplicity, let \( F_{st}(\text{education}_{ist}, \text{age}_{ist}) = \beta_s^t X_{ist}^s \), where \( X_{ist}^s \) is a vector of dummies for each age-education group in sector \( s \). We can then decompose the wage gap in each year as follows:

\[
\Delta_s'(E(\log(w_{ist}))) = \beta_s^t(E(X_{ist}^s)) - E(X_{it}^a) + (\beta_s^t - \beta_a^t)E(X_{it}^a) + (E(X_{ist}^s) - E(X_{it}^a))(\beta_s^t - \beta_a^t)
\]

The first term is entirely due to composition effects due to age and education differences in workers employed by sector \( s' \) relative agriculture. In other words, this component reflects the mean wage gap if all education-age groups were equally paid in both agriculture and sector \( s' \). The second term reflects the wage gap due to differential pay of each age and education pair, weighted by the distribution of observable characteristics present in agricultural workers. Unlike the first term, this second component is solely affected by differential returns to age and education, and not by differences in composition. The third term accounts for the interaction between the the composition and return effects.

Figure 8 shows the result of this decomposition. In 1996, composition effects explain only a small share of the agriculture vs manufacturing gap throughout the sample period, and they explain a larger share, but not all, of the services vs agriculture gap. This means differences in the age composition and educational attainment in each sector cannot account for most of the agricultural wage gap in the earlier period. In contrast, differential pay to these characteristics in each sector explains a larger share of the total gap level. Moreover, when looking at the evolution of this decomposition over time, most of the decline in the gap between agriculture and both manufacturing and services is driven by the steeper decline in the gap due to the estimated return coefficients. This implies that changes in the age distribution of the population or educational attainment do not drive the decline in the gap in Brazil. As seen in Table 1, educational attainment differences between agriculture and the rest of the economy were reduced over 1996-2013 in Brazil; nevertheless, this reduction alone explains little of the overall decline.

This limited role of age and education is present in spite of the lack of control for
unobservable skill differences between education-age groups. It is likely that this omission overstates the role of compositional differences implied by the decomposition described. For instance, if workers with higher education are paid more not because of their education, but rather because of unobservable skills that are correlated with their education level, this correlation biases upward the share of the wage gap explained by these observable characteristics. Hence, to the extent that more highly paid age-education groups possess more highly valued unobservable skills, the share of the gap explained by observables above is an upper bound on the role of these characteristics. In appendix A, the role of observables after controlling for worker fixed effects is estimated. Since individual workers’ changes in age and education have little impact on their wages, controlling for unobservable fixed characteristics erases most of the role of observables in explaining the gap.

Therefore, to understand the magnitude of the wage gap and its decline, one must go beyond measuring observable differences in worker characteristics. That is, one must ask why workers with similar age and education are paid differently across sectors, why this difference is part of an equilibrium, and why has this difference declined over time as the economy has developed.

Figure 8: Oaxaca decomposition

(a) Agriculture vs Manufacturing
(b) Agriculture vs Services

Note: Gap refers to the difference in mean log earnings between two sectors. Returns refer to the term \((\beta^s_t - \beta^a_t)E(X^a_t)\) and composition refers to term \(\beta^a_t(E(X^s_t) - E(X^a_t))\) of the Oaxaca decomposition.

5.2 Unobservable characteristics

The role of differential returns emphasized above does not necessarily imply that workers in agriculture are intrinsically less productive or skilled. It might still be the case that their
unobservable characteristics are similar to that of workers in other sectors but that these traits are simply paid differently in each sector.

Under perfectly competitive labor markets with fully mobile workers, every worker should move to the sector where he or she is paid the most. This process would eliminate any differences in pay among workers with similar —observed and unobserved— characteristics. This result is independent of any capital or technological limitations that are particular to each sector. However, compensating differential stories —where workers value sector-specific non-pay characteristics and are therefore willing to receive lower pay in some sectors— or mobility frictions can break this pattern. For instance, one can imagine a situation in which workers are unwilling to pay a mobility cost from moving to industrial areas or one in which workers are unwilling to sacrifice the perks of employment conditions in agriculture. These stories are able to generate wage gaps within each age-education groups that are consistent with the differential returns observed in the previous section. There are therefore two types of competing types of stories in explaining the Oaxaca decomposition above. On one hand, agricultural workers may have a different composition of unobservable characteristics which makes them less valuable in the market. On the other hand, workers may be similar in the two sectors, but mobility frictions or compensating differentials may induce differential pay for each worker type.

In order to distinguish differential pay from compositional differences in unobservable characteristics, it is necessary to use the panel dimension of the dataset. In particular, I analyze the inter-sector flows of workers and test whether wages change substantially when a worker switches sectors. The two types of stories described above have different hypotheses about the behavior of wages among sector-switchers. If the wage gap is entirely due to differences in unobservable characteristics, then we should not expect wages to change substantially among workers that switch sectors. In contrast, if mobility frictions or compensating differentials are driving most of the wage gap between agriculture and the rest of the economy, we would expect workers that switch out of agriculture to experience large gains. In addition, since we do not observe hours, but only monthly earnings in each sector, a third possibility is that differences in hours worked are driving the gap. This can be considered as an extension of the compensating differential story where the hours worked are an additional non-pay component of work that is specific to each sector. Similarly to the compensating differential story, if the measured earnings gap in Brazil is driven by non-agriculture workers laboring for longer hours, we would expect these to be reflected in large earnings jumps among the workers that switch out of agriculture into other sectors.
Figure 9: **Number of sector-switchers from and into agriculture**

![Graph showing the share of total workers switching between sectors from 1995 to 2015.](image)

**Note:** Share of total employed workers that switch out of or into agriculture in any given year.

Using information on sector-switchers, I estimate the magnitude of wage changes from sector transitions, after controlling for time trends and changes in workers’ observable characteristics. In order to study these switches, however, enough sector-switchers are needed to estimate these changes precisely. Figure 9 shows the share of workers that switch across sectors throughout the sample period. The small share of sector switchers (<1% of employed workers in a given year) would usually complicate the study of sector wage jumps using a small-sample panel dataset. However, because of the large number of workers covered the Brazilian data, this is not a problem. In any given year, there are over 180,000 formal workers who switch into and out of agriculture in Brazil.

To assess the magnitude of wage changes after controlling for differences in unobserved characteristics, the following worker fixed effect model is estimated

\[
\log(w_{it}) = \beta^t_m \cdot M_{it}\phi_t + \beta^t_s \cdot S_{it}\phi_t \\
+ \phi_t + \phi_i + F_t(education_{it}, age_{it}) + \epsilon_{it}
\]

(1)

where \(M_{it}\) and \(S_{it}\) are indicators for working in the Manufacturing and Services sectors, respectively; \(\phi_t\) and \(\phi_i\) are time and individual fixed effects; and \(F_t(education_{it}, age_{it})\) is
set of interacted year, age and education dummies. Most importantly, sector indicators are interacted with time; therefore, the coefficients $\beta^s_t$ and $\beta^m_t$ reflect average earnings change from switching sectors from agriculture to both manufacturing and services in each year $t$. I will refer to these coefficients as sector premiums with respect to agriculture, of which there are $2 \times T$ in the model, where $T$ is the number of years in the sample. The model is estimated using all formal workers in Brazil from 1996 to 2013. In the baseline estimation of the model, the sector premiums are identified by workers who switch sectors during this period, and controls are estimated using information from all formal workers in the data.

The time series of both services premiums ($\beta^s_t$) and manufacturing premiums ($\beta^m_t$) are shown in Figure 10. A first takeaway from the figure is that wage differences estimated from switchers are much smaller than the overall wage gap. This is true throughout 1996-2013. For manufacturing, the average sector premium during 1996-2013 is 0.11 compared to the overall wage gap of 0.48 log points relative to agriculture. Similarly, for services, the average jump in wages is 0.08 compared to the mean total gap of 0.46 log points. Hence, sector premiums as a percentage of the total gap in a given year averaged 21% when comparing agriculture vs manufacturing and 16% when comparing agriculture to services. Potential gains from switching out of agriculture are significantly smaller than the overall agricultural gap. These shares suggests that the potential role of theories producing differential pay of similar workers across sectors is limited.

A key identification assumption of the model is that the error term must be orthogonal to the manufacturing and services dummies. This is violated if workers that switch out of agriculture are precisely the ones who would experience the largest wage jump from switching out of agriculture, which may certainly be the case. In a mobility frictions story, for example, it is precisely the workers who stand to gain the most from transitioning the ones who are willing to overcome this friction and move out of agriculture. Similarly, in a compensation differential story, workers only accept to move out of agriculture if compensated for the loss of non-pay benefits enjoyed in their original sector. These mechanisms, however, would bias our sector premium estimates upwards, so that $\beta^m_t$ and $\beta^s_t$ are upper bounds on the potential wage gains to be obtained from switching out of agriculture. To the extent that sector-switchers are the ones who stand to gain the most, this further depresses the role of differential pay stories in explaining the overall wage gap.

Another related concern is that the estimates might be affected by the inclusion of all workers in the estimation rather than just sector-switchers. Table 2 shows the average sector premium coefficients by period when the model in equation (1) is estimated using only sector-switchers and only transitions out of agriculture. If anything, a focus on switchers
further lowers the estimates of sector premiums estimated in the baseline. Moreover, results
do not appear to be driven by asymmetries from sector-switches. This might be a concern
if switchers into agriculture are solely driven by improving job offers and these positive job
changes counterweight large potential premiums from workers switching out of agriculture.
This is no the case, as the model estimated solely on workers who switch out of agriculture
yields lower results relative to the baseline. In fact, when performing an event-study of
workers that exit agriculture (appendix B), it is not entirely clear whether the wage jump
from exiting agriculture is drastically different that the average gain expected from an extra
year of experience working in any given sector. Potential gains from switching sectors appear
to be even smaller when focusing exclusively on transitions out of agriculture.

The small sector-premiums described above may seem surprising. At first glance, they
appear at odds with a literature that highlights the role of working environments —and
firms in particular— in determining wage differences. For example, Card, Heining and
Kline (2013), Menezes-Filho, Muendler and Ramey (2008), Barth et al. (2014) and others
have highlighted the existence of a firm component of wages that is orthogonal to both
observable and unobservable worker characteristics. Alvarez et al. (2015) analyze this
pattern in Brazil and conclude that a reduction in wage dispersion has been accompanied
by a compression of firm-specific components. The reason for this apparent discrepancy is
that firm premium differences do exist between firms in Brazil, but the differences in average
 premiums between agriculture and other sectors are small when compared to the overall gap.
Appendix C addresses the role of within sector heterogeneity in this context.

Altogether, the limited pay gains associated with the exit out of agriculture into other
sectors limits the scope of stories that predict large wage jumps among workers who switch
sectors. Stories generating differential pay of similar workers do not appear to be the main
drivers of the gap. On the contrary, most of the wage gap stems from differences in the
composition of unobservable —and to a lesser degree observable— worker characteristics in
each sector.
Figure 10: Sector gaps relative to agriculture controlling for individual fixed effects

Note: Total refers to the difference in mean log earnings between each non-agricultural sector and agricultures. Sector premiums for services ($\beta_t^s$) and manufacturing($\beta_t^m$) are defined by equation (1). With the exception of 2008, coefficients are all statistically different from zero ($p < .01$).

6 How are compositional differences in characteristics generated?

The analysis above suggests that most of the wage gap — particularly in the late 1990s — is not due to differential pay of equally skilled workers between agriculture and non-agriculture. Instead, the wage gap appears to be largely driven by compositional differences in educational attainment and fixed unobservable characteristics between sectors. According to the results presented, a plausible mechanism for generating wage gaps must therefore achieve a very particular goal. It must generate wage gaps driven by large differences in worker characteristics in each sector without giving rise to large differences in pay for similar workers in the two sectors.

Following the work of Roy (1951), recent papers have proposed the sorting of workers with sector-specific skills as a possible explanation of wage and productivity differences
6.1 Motivating a Roy Model: The existence of sector-specific skills

A basic premise of Roy models is the existence of occupation or sector-specific skills. In the context of the agricultural wage gap, a worker under this view has agriculture-specific skills and non-agriculture-specific skills, which determine the productivity of the worker when performing sector-specific tasks. By influencing labor productivity, sector-specific skills also

The mechanism can generate inter-sector gaps driven by compositional differences in worker characteristics in a manner that is consistent with the empirical observations described. In this section, I test the explanatory power of worker sorting in explaining the wage gap level and its decline. I first assess the existence of sector-specific skills, which is a key assumption of these models. Motivated by this exercise, I then describe and calibrate a sorting model to show how large differences in mean wages between sectors can be generated as an equilibrium outcome of heterogeneous workers with sector-specific skills freely choosing sectors. Finally, I show how growth in such an economy can explain part of the wage gap decline.

Note: Average of sector premiums, $\bar{\beta}_s$ and $\bar{\beta}_m$, over each 6-year period are presented. These are defined by equation (1), and they are estimated for three samples. All workers category comprise all formal workers between 18 and 55 years old in the RAIS. Sector switchers restrict the sample to workers that have switched into or out of agriculture at least once in each 6-year interval. Exiters from agriculture are defined as workers that have switched from agriculture to another sector at least once in each 6-year interval.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Services ($\bar{\beta}_s$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All workers</td>
<td>0.14</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>Sector-switchers</td>
<td>0.08</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Exiters from agriculture</td>
<td>0.09</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td><strong>Manufacturing ($\bar{\beta}_m$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All workers</td>
<td>0.17</td>
<td>0.09</td>
<td>0.06</td>
</tr>
<tr>
<td>Sector-switchers</td>
<td>0.17</td>
<td>0.10</td>
<td>0.06</td>
</tr>
<tr>
<td>Exiters from agriculture</td>
<td>0.13</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

between countries, urban vs rural areas, and sectors. This mechanism can generate inter-sector gaps driven by compositional differences in worker characteristics in a manner that is consistent with the empirical observations described. In this section, I test the explanatory power of worker sorting in explaining the wage gap level and its decline. I first assess the existence of sector-specific skills, which is a key assumption of these models. Motivated by this exercise, I then describe and calibrate a sorting model to show how large differences in mean wages between sectors can be generated as an equilibrium outcome of heterogeneous workers with sector-specific skills freely choosing sectors. Finally, I show how growth in such an economy can explain part of the wage gap decline.
determine the potential wage of the worker in each sector and influence his labor allocation decision between sectors. It is not clear, however, whether sector-specific skills exist at all.

Though I cannot test the presence of sector-specific skills directly, I can test an implication of sector-specific skills on wage changes among sector-switchers and the sector premiums described in section 5.2. In particular, I study to what extent are sector premiums driven by workers performing different occupations after switching sectors. In a world where workers have sector-specific skills, wage gains from transitioning out of agriculture into another sector should be more prominent when workers perform a different task in their new sector of employment. If, on the contrary, workers are equally productive regardless of the task performed, then wage changes from transitioning out of agriculture must be driven by other forces that are not necessarily related to an increase in labor productivity.

For example, consider a member of the cleaning staff of an agricultural firm who is considering switching out of agriculture. In a world where sector-specific skills exist, he has the potential to achieve a different level of productivity in the non-agricultural sector. That is, the possibility of performing new tasks (e.g. machinery operation, human-capital intensive tasks) that are fundamentally different from the ones performed originally can enable the worker to exhibit sector-specific skills, and therefore improve his productivity of his labor. This change in productivity can in turn induce a wage gain from transitioning sectors. In contrast, if the worker transitions out of agriculture but performs the same set of tasks related to his original cleaning job, we would expect gains to be more limited. Switching sectors without switching occupations limit the realization of sector-specific skills and, therefore, potential wage gains under this view.

A measurable implication of this is that, to the extent that sector-specific skills are important in explaining wage changes among sector-switchers, we would expect sector premiums from 5.2 to be significantly reduced once we control for changes in occupation. If, on the contrary, wage changes are driven by differential pay for similar work, a cleaning worker switching between sectors but remaining a cleaning worker, would experience a gain in wages that is similar to the overall sector premium. In this second case, sector premiums should not be significantly reduced by controlling for occupations. To test which story is a more accurate description of sector wage premiums, I calculate premiums before and after controlling for occupations. The following equation is estimated

\[
\log(w_{it}) = \gamma_m^t * M_{it} \phi_t + \gamma_s^t * S_{it} \phi_t + \phi_{occupation} \\
+ \phi_t + \phi_i + F(education_{it}, age_{it}) + \varepsilon_{it}
\] (2)
where coefficients $\gamma_{tm}$ and $\gamma_{ts}$ reflect the average differential pay of workers performing the same occupation in both pre and post-transition sectors, $\phi_{occupation}$ are occupation fixed effects at the three-digit classification level, and the rest of variables are defined as described in section 5.2. Similarly to the model outlined in the previous section, this model is identified by workers who switch sectors. The main difference of this approach, however, is that the coefficients $\gamma_{tm}$ and $\gamma_{ts}$ are identified using sector-switchers that do not switch occupations after they transition. In the data, several occupations are common to all sectors (e.g. cleaning, security services, drivers/messengers) and the model is therefore identified.

The specification above allows the decomposition of the sector premiums estimated in section 5.2 into two components: one stemming from occupation switches and another reflecting differential pay for workers performing the same occupation in the two sectors.\(^8\) Figure 11 shows the evolution of this decomposition over time. The decomposition is depicted comparing agriculture to both services and manufacturing separately with similar results. Throughout the period, most of the sector premium was explained by occupation differentials across sectors. In contrast, the differential pay component was relatively low throughout the period. In fact, the sector premium practically disappears by the end of the period when comparing services vs agriculture, and it is below .03 log points when looking at manufacturing vs agriculture.

The fact that sector premiums are much diminished once occupation differences are controlled for imply that sector premiums are attributable to the changes in occupation when transitioning sectors. This favors the existence of sector-specific skills, for the difference between pre and post transition can be nearly entirely attributed to the change in tasks performed by the worker after switching sectors. In terms of the example described, the results suggest that a worker in cleaning services in the agricultural sector is not expected to experience a significant gain in earnings when switching to a different sector unless he performs a different occupation. Instead, workers switching out of agriculture appear to get access to new types of jobs, and this accounts for most of the average jump in their wages. To the extent that sector-specific jobs imply the demonstration of sector-specific skills, this jump in wages among sector-switchers supports a Roy view of the world where workers have sector-specific abilities.

\(^8\)The sector premiums estimated in section 5.2 can be written as $\beta_{s'} = \gamma_{s'} + (E_{s'}(\phi_{occupation}) - E_{agriculture}(\phi_{occupation}))$ where $s$ is each non-agricultural sector and $E_{s'}(\phi_{occupation})$ is the average of occupation fixed effects in each sector.
Figure 11: Sector premiums after controlling for occupational changes

<table>
<thead>
<tr>
<th>Year</th>
<th>Occupations (Manuf.)</th>
<th>Premium (Manuf.)</th>
<th>Occupations (Serv.)</th>
<th>Premium (Serv.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Occupations refer to the difference in mean occupation fixed effects between sectors. Sector premiums for services ($\gamma_t^s$) and manufacturing ($\gamma_t^m$) are defined by equation (2). Premium coefficients are all statistically different from zero ($p < .01$).

6.2 A Roy model of selection with mobility frictions

Motivated by the empirical results shown above, I construct a Roy model to assess the explanatory power of worker sorting in explaining the agricultural wage gap. The following model borrows heavily from Lagakos and Waugh’s (2012) framework, deviating from it in two ways. The first is the introduction of a friction that allows for differential pay of similarly skilled workers. The second is the use of earnings information from workers that switch sectors to calibrate the distribution of idiosyncratic and unobservable productivity parameters of workers. In contrast to their framework, the level of the inter sector wage gap is not a calibration target of the model but rather an outcome that I evaluate the model against. In particular, I ask whether a wage gap driven by compositional differences in worker characteristics of the magnitude seen in Brazil can be explained by the worker sorting mechanism. I then assess whether growth in this environment can generate a wage gap decline similar to one observed in Brazil during 1996-2013. The model’s components are now described in detail.
6.2.1 Preferences and endowments

There is a unit continuum of workers with unit mass, where each worker has identical preferences over agricultural and non-agricultural goods. Following other studies on structural change and the dual economy, workers have a subsistence requirement of $\bar{a}$ agricultural goods. Hence, preferences are non-homothetic and the share of expenditure on agricultural goods increases as income grows. Preferences for each worker $i$ are given by

$$U(c^i_a, c^i_n) = \log(c^i_a - \bar{a}) + \phi \log(c^i_n)$$

where $c^i_a$ and $c^i_n$ refer to consumption of agricultural and non-agricultural goods respectively and $\phi$ is a weight parameter that determines the relative importance of non-agricultural goods in consumption.

In addition, each worker is endowed with one unit of labor and sector-specific individual productivities $\{z^i_a, z^i_n\}$ drawn from a distribution $G(z_a, z_n)$ with support $[z, \infty)^2$. A worker can freely choose to work in one of the two sectors but faces a cost $k$ if he decides to work in the non-agriculture sector. This distortionary mobility friction will introduce compensation differences for workers that switch sectors as observed in the data. In this environment, workers maximize their income $y^i$, so that $y^i = \max\{w_a(z^i_a), w_n(z^i_n) - k\}$ where $w_a(z^i_a)$ and $w_n(z^i_n)$ are the wages offered to worker $i$ in the two sectors. Thus, each worker faces the following budget constraint

$$p_a c^i_a + c^i_n \leq y^i$$

where the non-agricultural good is set to be the numeraire and $p_a$ is the relative price of the agricultural good.

6.2.2 Technology

There is a profit-maximizing representative firm in each sector with production functions given by

$$Y_a = AZ_a, Y_n = AZ_n$$

---


<sup>10</sup>Since the focus of the quantitative exercise is analyzing the growth-induced exit of agricultural workers into other sectors, I focus on the mobility cost of going from agriculture to other sectors and not vice-versa.
where $A$ is an economy-wide productivity parameter and $Z_s$ are the total effective units of labor employed by sector $s$. This second term is equivalent to the sum of individual worker productivities hired by the firm, or $Z_s = \int_{z \in \Gamma_s} z_s^i dG_s$, where $\Gamma_s$ is the set of workers hired by sector $s$. Similarly, the number of workers employed by sector $s$ is given by $L_s = \int_{z \in \Gamma_s} dG$.

Labor productivity of a sector is therefore a function of the integral over the individual productivity parameters of workers employed in that sector.

### 6.2.3 Competitive equilibrium

An equilibrium is determined by a relative price of the agricultural good $p_a$, wage functions $w_a(z_a), w_n(z_n)$, consumption decisions $c_a^i, c_n^i$ and labor allocations such that:

1. Firms maximize profits in the two sectors given their technology.
2. Workers’ labor allocations maximize their income.
3. Consumption allocations maximize utility subject to the budget constraint.
4. Labor, agricultural, and non-agricultural goods markets clear.

In a competitive market, the first condition requires firms to offer workers a wage equal to the value of their marginal product in their respective sector. These wages vary by both the aggregate productivity factor, relative prices, and the idiosyncratic productivity parameter of each worker.

$$w_a(z_a) = p_aAz_a, w_n(z_n) = Az_n$$

Taking these wages as given, each worker decides to allocate their unit of labor to one of the two sectors. The second condition implies that a worker will chose to work in agriculture as long as the value of their marginal product in agriculture is more than his potential production in non-agriculture minus the mobility cost, or

$$Ap_a z_a^i \geq Az_n^i - k$$

An implication of this is that higher relative agricultural prices, for a given non-agriculture productivity, lowers the minimum agricultural productivity required to stay in agriculture. Moreover, income is given by $y^i = \max\{Ap_a z_a^i, Az_n^i - k\}$ and consumption demand for both goods are given by,
The above holds as long as \( y^i \geq \bar{a}p_a \). Otherwise, \( c^i_a = y^i/p_a \) and \( c^i_n = 0 \). Intuitively, the consumption rule consists of allocating resources on agricultural goods until the minimum subsistence requirement is met and then distributing the remainder among the two goods according to the weight parameter \( \phi \). Thus, as income grows, a lower proportion of income is allocated to the consumption agricultural goods.

Finally, the market clearing conditions require that good markets clear and that the labor employed in each sector is consistent with workers’ labor allocations. This is

\[
\begin{align*}
\int c^i_n dG &= \int_{i \in \Gamma^n} (A z^i_n - k) dG \\
\int c^i_a dG &= \int_{i \in \Gamma^a} A z^i_a dG \\
\Gamma^n &= \{ i : A p_a z^i_n \leq A z^i_n - k \} \\
\Gamma^a &= \{ i : A p_a z^i_a > A z^i_n - k \}
\end{align*}
\]

Our main subject of study is the wage gap between agriculture and non-agriculture. In the model, this is determined by

\[
\frac{E(w_n)}{E(w_a)} = \frac{\int_{i \in \Gamma^n} w_n(z^i_n) dG}{\int_{i \in \Gamma^a} w_a(z^i_a) dG} = \frac{\int_{i \in \Gamma^n} z^i_n dG}{p_a \int_{i \in \Gamma^a} z^i_a dG}
\]

The wage gap is therefore the result of two main mechanisms. The first is the direct effect of the relative price, which affects the relative valuation of efficiency units, \((z^i_a, z^i_n)\), for the output produced in the two sectors. The lower is the relative price of agriculture, the lower is the relative value of agricultural output and hence the greater is the wage gap, holding composition of workers constant. The second mechanism is selection, which affects the productivity distribution of the sets of workers working in the two sectors \((\Gamma^a, \Gamma^n)\). The lower is the relative price of the agricultural good, the more people exit agriculture to work in the other sector. This process changes the composition of workers in each sector which can increase or decrease the gap in mean worker productivity between sectors. The wage gap level is therefore larger or smaller depending on the equilibrium effect of these two mechanisms. Moreover, as a country grows richer and the relative price of agriculture
declines, the net effect of both price and composition effects on the wage gap over time is undetermined. Whether wage gaps decline or rise in this environment depends on the parameters of the economy.

### 6.2.4 Calibration strategy

I now proceed to calibrate the model to answer two questions. The first is whether worker sorting can generate wage gaps that are of the same order of magnitude of the ones seen in the data. The second is whether sector-neutral growth can reduce the gap over time.

To do this, I calibrate the economy to Brazil in the earliest period of 1996-1997, when the gap was the largest. I then introduce growth in parameter $A$ that is consistent with real output growth rates observed in 1996-2013. Preference, production, friction and productivity distribution parameters are jointly estimated to match different moments of the data. Although all of these parameters interact in the model, each of them has stronger implications for particular moments. Below, I describe the relationship of each parameter to each moment and how these are calibrated.

**Preference parameters** Consistent with the literature using dual economy models with minimum subsistence requirements, preference parameters $\phi$ and $\bar{a}$ are calibrated to match two moments of the data that relate to labor and output shares. The first is the share of workers in agriculture of 25% observed in 1996. The second is a long-run agriculture output share of 0.5%, which is the standard parameter used by Lagakos and Waugh (2013), Restuccia, Yang and Zhu (2008) and other studies of structural change. Once calibrated, the minimum subsistence requirement is 23% of the average wage.

**Production and friction parameters** The technology parameter $A$ is set to 1 for the initial calibration in 1996. Later, when studying the effect of growth on the wage gap, changes in $A$ are calibrated to match total yearly real output growth in Brazil during 1996-2013. This implies a cumulative growth in $A$ of 105% by 2013. The sector-neutral nature of the productivity parameter in this economy implies that the wage gap is solely dependent on endogenous price and selection effects in the model.12

---

11 As a country grows richer, non-homothetic preferences imply that a lower share of income is allocated to agricultural goods consumption and a lower relative price of agricultural goods. See Lagakos and Waugh (2013) for a detailed discussion.

12 Alternatively, one could introduce distinct growth rates by sector. The impact of differential growth on the wage gap, however, is similar to the one described in this paper. Sector-specific growth rates change the relative price of agriculture which induces the exit of workers out of agriculture and changes worker
In addition, the mobility friction $k$ is calibrated to match the average wage gain, in logs, from workers that switch out of agriculture. This was at its highest point (.13 log points) in 1996. The relative small transition wage gains relative to the overall gap described in section 5.2 imply that the magnitude of $k$ is relatively small at 20% of the minimum subsistence requirement and 4.6% of the average wage in the economy. This confirms the conclusion in the previous section: micro-data is not consistent with mobility frictions generating large pay differentials in sector-switching workers.

**The joint distribution of sector-specific worker productivities** The wage gap is affected by the exogenous distribution of individual productivity parameters. I calibrate these parameters using information from wage dispersion within each sector as well as information from workers that have worked in two sectors. First, I restrict $\tilde{z}$ to match $\tilde{a}$, so that every worker in the economy has a labor productivity endowment that is sufficient to both afford minimum subsistence and transition out of agriculture. This is, workers close to $\tilde{z}$ will freely choose to stay in agriculture because of preferences, and not because of the lack of sufficient income potential to pay the cost $k$. In this way, the wage gap is entirely due to endogenous selection and prices, and not to the distribution’s support parameter.

Since I do not observe all workers at all sectors, it is impossible to calibrate the distribution $G$ without imposing some structure. Non-parametric estimation is therefore not an option. Instead, following Lagakos and Waugh (2013), I allow workers to have dependent draws from sector-specific Fréchet distributions $X(z_a)$ and $Y(z_n)$ and restrict the joint distribution $G(z_a, z_n)$ to be a Frank copula resulting from the two primary distributions. This is

$$
G(z_a, z_n) = C[X(z_a), Y(z_n)]
$$

$$
C[u, v] = -1/\rho \ast \log(1 + \frac{(e^{-\rho u} - 1)(e^{-\rho v} - 1)}{e^{-\rho} - 1})
$$

$$
X(z_a) = e^{-z_a - \theta_a}
$$

$$
Y(z_n) = e^{-z_n - \theta_n}
$$

The sector-specific distributions have dispersion parameters $\theta_a$ and $\theta_n$, which control the within sector variance of the productivity distribution in agriculture and non-agriculture. Productivity composition in each sector. Regardless of whether growth is sector-neutral or not, wage gaps are driven solely by the sorting of workers across sectors.
These are calibrated to match the variances of log earnings in agriculture and non-agriculture of .33 and .63 respectively. The calibrated parameters that match these log variances are $\theta_a = 1.2$ and $\theta_n = .51$.

Besides the transparent mapping that exists between these parameters and wage dispersion, there are two other reasons why a Fréchet shape is a sensible choice to model sector-specific distributions. First, the Fréchet distribution is a special case of the extreme value distribution; therefore, the marginal Fréchet distribution of a particular sector can be interpreted as the distribution of the maximum draw from a set of productivity distributions within that sector. For example, this may represent the maximum productivity draw out of a series of jobs that are available within the manufacturing or agricultural sectors. Second, the shape of the distribution, with greater mass at lower productivity parameters and fat tails, resembles the within sector distribution of both raw wages and unobservable worker characteristics observed in the data.

To form a joint distribution out of the two sector-specific marginal distributions, a Frank copula is used. The advantage of using this copula is that it allows the degree of dependence in the two distributions to be controlled by a single parameter $\rho$. Along with $\theta_a$ and $\theta_n$, this parameter is calibrated to match the fraction of workers that exit agriculture during 1996-1997 (4% of agricultural workers) when the economy grew 2.2% in real terms. Intuitively, for given dispersion parameters $(\theta_a, \theta_n)$, $\rho$ controls the amount of workers close to the labor allocation indifference condition ($A_p a z^i a = A z^i n - k$). A growth-induced change in prices pushes a larger or smaller share of workers out of agriculture depending on the mass of workers that are close to indifferent in the base year. The resulting parameter from this calibration is $\rho = 12$, which implies a linear correlation of 35% between the sector-specific productivity parameters $z_a$ and $z_n$. Importantly, no difference in mean productivity between sectors is assumed in the calibration of the joint distribution. The agricultural wage gap is therefore not a calibration target but an outcome of the model.

6.3 The explanatory power of the sorting mechanism

The model generates both cross-sectional and inter-temporal predictions. In the 1996 cross-section, the predicted wage gap by the calibrated model is .43 log points. This magnitude is 66% of the wage gap observed. The calibration using data on sector switchers and within sector variances is, therefore, consistent with large compositional differences in worker skills arising from sorting, but working sorting by itself does not explain all of the gap.

\footnote{By the extreme value theorem, the maximum of independent draws from any distribution converges to an extreme value distribution. The Fréchet is an example of these distributions.}

34
The reason behind this shortcoming is related to the magnitude of $\rho$. In the model, a higher correlation in the sector-specific productivity parameters generates larger wage gaps between agriculture and non-agriculture. Intuitively, under perfect correlation of productivity draws $z_n$ and $z_a$, the most skilled workers would work in non-agriculture, while the worst skilled workers would work in non-agriculture. In that setting, the lowest paid worker in non-agriculture earns more than the highest paid worker in agriculture. Thus, the perfect correlation produces a large inter-sector wage gap. In contrast, with imperfect correlation, some workers are much more productive in agriculture than in non-agriculture. These workers sort into agriculture. In equilibrium, imperfect correlation produces some agricultural workers that earn more than their peers in non-agriculture. A lower correlation parameter $\rho$, therefore, reduces the wage gap.

In the model, the correlation parameter $\rho$ required to match the total wage gap between sectors would generate too few workers that are close to indifferent between the two sectors and, therefore, would underestimate the amount of workers who choose to switch. In order to match the wage gap perfectly, an additional source of wage heterogeneity would be required. For instance, the inclusion of transitory shocks in a dynamic setting would increase the share of sector switchers and ease the effects required from productivity dispersion parameters. Larger wage gaps can be generated if the target share of switchers to be produced by the sorting mechanism is lowered. This additional uncertainty is beyond the scope of this paper, as we are focused on sorting stemming from fixed unobservable skill differences between workers. However, it is worth noticing that, even when restricting the sources of wage heterogeneity in the model, this mechanism can generate a wage gap that is of a similar order of magnitude to the one observed in the data.

Once the economy is calibrated to 1996, I examine whether growth —here introduced as an increase in the sector-neutral productivity parameter $A$— generates a wage gap decline. Growth has two effects. On one hand, the increased income decreases the demand for agricultural goods consumed in the economy relative to non-agricultural goods. This is a direct consequence of the subsistence requirement present in preferences. The reduction in relative demand induces a lowering of the relative price of the agricultural good, which decreases the market value of the marginal product of agricultural workers. This in turn depresses relative wages in agriculture which widens the gap. On the other hand, the price changes cause the exit of workers out of agriculture into the manufacturing sector. The transitioning workers lower the average productivity of the non-agriculture sector, which reduces the gap in average wages between the two sectors. As mentioned before, the net effect of these two forces depends on the economy’s parameters.
Figure 12 shows the comparison of the wage gap decline, in absolute value and relative to the 1996 wage gap level, produced by the model and the data. Qualitatively, the model is indeed successful in generating a decline. Over the 18 year period, the wage gap in the model declines by 15% relative to the start of the period. This relative decline is 58% in the data. In absolute magnitude, the understated gap in 1996 declines .06 log points, which is small relative to the overall decline in the data of .38 log points. The worker sorting mechanism is therefore consistent with the wage gap decline; nonetheless, the limited magnitude of the predicted wage gap reduction leaves a substantial share of the decline unexplained.

One concern is whether the magnitudes of the results are driven by the parametrization assumptions behind the joint distribution of sector-specific productivities. Parametrization is necessary since we are not able to observe the productivity of all workers across all sectors, but it is hard to determine whether this is a reasonable approach to study the sorting of agricultural workers into other sectors as an economy grows. Though far from being a test of validity from the parametrization assumptions, one can assess whether the shape of the distribution is reasonable by observing the rate of exit of workers out of agriculture. For instance, if the joint distribution worker productivities $G$ contains discontinuities or a radically different shape from the calibrated distribution, we would expect the relationship between growth and the exit of workers out of agriculture to behave different in the data and the model. The evolution of the share of workers employed in agriculture in the data and in the model are shown in Figure 13. Similar to the data, the model generates a gradual exit of agricultural workers that has a similar pattern to the one observed in Brazil. The model predicts a decrease of 14% in the share of workers in agriculture so that it reaches 11% in 2013. This is slightly lower than the 14% share observed in the data by the end of the period. Remarkably, however, the qualitative direction and linear relationship between growth and the share of workers in agriculture is similar in both the data and the model. The pattern resemblance is partially due to the calibration strategy used to match growth and the exit of workers, since the parameters are matched to the share of workers exiting agriculture in 1996-1997. However, the relatively good fit of this relationship over the entire 18 year pattern is somewhat surprising. The sensible magnitudes in the number of workers that exit agriculture speak in favor of the validity of the shape assumptions and calibration of the joint productivity distribution.

Overall, the sorting mechanism of the model captures relatively well the movement of workers out of agriculture, the wage gap levels, and the qualitative decline in the agricultural wage gap. However, the calibration using micro-data suggests that other complimentary
mechanisms must also be at play to fully explain the magnitude of these patterns.

**Figure 12: Wage Gap**

(a) Wage gap level

(b) Relative to 1996

Note: Panel (a) shows the difference in the mean of log wages between sectors. Panel (b) shows the relative decline of the wage gap relative to the level observed in 1996 in both the data and model output.

**Figure 13: Share of Workers in Agriculture**

Note: Share of workers from data includes formal and informal workers.
7 Conclusion

The large wage gaps between agriculture and other sectors are mostly driven by compositional differences in unobservable worker characteristics. In accordance with cross-country patterns on inter-sector productivity and wage gaps, these differences declined gradually in Brazil since the late 1990s as the country became richer. This paper argues that a Roy model of worker sorting based on sector-specific skills is consistent both with the magnitude of the gap and its decline as a country grows. Moreover, the mechanism suggests that the fall in the wage gap may be an integral part of the development process, where compositional differences between sectors are reduced over time as workers transition out of agriculture.

Nonetheless, a significant share of the gap remains unexplained, and the model does not speak to the steeper decline in value added per worker observed both in Brazil and in cross-country studies. Exploration of additional mechanisms is therefore needed, and the evidence presented in this paper can guide the design of future wage gap models. Specifically, the results show that pay differences for workers with similar skills are relatively small when compared with the total wage gap. This finding discourages models that generate large pay differences for similar work in different sectors. For instance, mobility frictions or compensation differentials that induce large gaps in wages per efficiency units between agriculture and non-agriculture would predict large wage gains from switching sectors that are at odds with the data. Therefore, complementary mechanisms to worker sorting that attempt to rationalize the wage gap level and its decline must, at the very least, produce large differences in average pay between sectors without producing relatively large wage gains for workers that switch out of agriculture.

References


A The gap explained by observable differences after controlling for unobservable characteristics

Differences in pay across sectors for each age and education group can potentially be explained by differences in the composition of unobservables. It might be the case that, for a given age and education level, workers in one sector are fundamentally more productive than others. Workers certainly differ in dimensions other than age and education, and the literature has emphasized the importance of unobservable skill heterogeneity in explaining wage differences. Thus, an added challenge when comparing wages in different sectors is understanding how these unobserved individual skills and human capital play out differently in each sector. In order to analyze the relative importance of observed and unobserved worker characteristics, I estimate the following model using the panel dimension of the dataset.

\[
\log(w_{ist}) = \phi_{is} + F_{st}(\text{education}_{ist}, \text{age}_{ist}) + \varepsilon_{ist}
\]  

(3)

where \(\phi_{is}\) is the fixed effect of a worker in sector \(s\).\(^{14}\) This specification controls for worker unobservables that are fixed over time when estimating differential compensation of age and education levels. As before, \(F_{st}(\text{education}_{ist}, \text{age}_{ist})\) is a sum of education, age and year interacted dummies. Unlike the baseline specification in section 5, the identification of age and education returns comes from the few workers that switch age and education categories throughout the period of study. I then conduct a modified Oaxaca decomposition\(^{15}\) using this age and education compensation function. This decomposition is given by

\[
\Delta(E(\log(w_{ist}))) = \Delta E(\phi_{is} + \varepsilon_{ist}) \\
+ \beta_s^s(E(X_{ist}^s) - E(X_{ist}^a)) + (\beta_s^s - \beta_t^a)E(X_{ist}^a) \\
+ (E(X_{ist}^s) - E(X_{ist}^a))(\beta_s^s - \beta_t^a)
\]

The first term reflects the wage gap between sectors due to unobservables, the second term accounts for compositional effects, the third term measures wage gaps due to differential pay to education characteristics, and the last term is the interaction effect. Figure 14 shows the results of this decomposition for both manufacturing and services. Once we control for fixed unobserved characteristics, differential pay across sectors of each education-age group explains little of the gap levels.

\(^{14}\)Since there are relatively few sector switchers these results hold even when restricting \(\phi_{is} = \phi_i\).

\(^{15}\)From Oaxaca (1973).
Figure 14: **Oaxaca decomposition controlling for unobservables**

(a) Agriculture vs manufacturing  
(b) Agriculture vs services

Note: Gap reference to the difference in mean log earnings between two sectors. Returns refer to the term \((\beta_s^t - \beta_a^t)E(X_{it}^a)\) and composition refers to term \(\beta_a^t(E(X_{s}^{s'}) - E(X_{a}^{s}))\) of the Oaxaca decomposition. Unobservables refer to the difference in mean worker fixed effects, \(\Delta E(\phi_{is})\).

### B Event-study of transitions into and out of agriculture

This section adopts an event-study framework focusing only on workers that switch out of agriculture, restricting the analysis to workers that were employed in the pre and post transition sector for at least two years. The following equation is estimated for transitions out agriculture into both services and manufacturing.

\[
\log(w_{ist}) = \sum_{j=-2}^{1} \gamma_j + \phi_t + F_t(education_{ist}, age_{ist}) + \epsilon_{ist}
\]

Figure 15 shows the results of the transition impact coefficients, \(\gamma_j\), as well as the pre and post-transition raw wages. As before, the wage jumps from transitions are much smaller than the magnitude of the overall gap. The average wage boost in the year after switching out of agriculture ranges from 0.01 to 0.1 compared to the overall gap of 0.3-0.7. These small magnitudes are persistent for workers that switch out of agriculture into both services and manufacturing. Moreover, when looking at the yearly changes in the pre and post-transition years, it is not entirely clear the whether wage jumps in transition years are of substantially different magnitudes than from the typical within sector gains in any given year. It can be concluded that the focus on any particular switches does not alter the conclusion of limited differential pay of similar workers across sectors.
Figure 15: Transitions out of agriculture

(a) From agriculture to manufacturing

<table>
<thead>
<tr>
<th>Year</th>
<th>Log Difference</th>
<th>Wage Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996_2001</td>
<td>-0.05</td>
<td>-1</td>
</tr>
<tr>
<td>2002_2007</td>
<td>-0.5</td>
<td>-2</td>
</tr>
<tr>
<td>2008_2013</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

(b) From agriculture to services

<table>
<thead>
<tr>
<th>Year</th>
<th>Log Difference</th>
<th>Wage Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996_2001</td>
<td>-0.1</td>
<td>-1</td>
</tr>
<tr>
<td>2002_2007</td>
<td>-0.05</td>
<td>-2</td>
</tr>
<tr>
<td>2008_2013</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Raw mean wages before and after transitions out of agriculture are shown. Year 0 refers to the last year worked in the pre-transition sector and Year 1 refers to the first year in the post-transition sector. Year 1 refers to the first year in the post-transition sector. Coefficients $\gamma_j$ reflect wage changes from transitions after controlling for time fixed effects and changes in age and education.

C Within sector heterogeneity

In order to assess within sector heterogeneity, at the firm level, in explaining the agricultural gap, I follow the methodology described in Alvarez, Engbom and Moser (2015) to decompose wages into firm, individual and worker characteristic components, or
\[
\log(w_{ist}) = \phi_{J(i,t)} + \phi_i + F_t(\text{education}_{ist}, \text{age}_{ist}) + \epsilon_{ist}
\]

where \(\phi_{J(i,t)}\) is a time varying firm-effect for firm \(J\) where worker \(i\) is employed at time \(t\), and the other variables are defined as described in section 5. This estimation parallels the one conducted in equation (1) but it estimates average differences at the firm-level rather than at the sector level. As shown by AKM (1999), worker and firm effects can only be separately identified within a set of firms and workers connected through the mobility of workers. Following a similar approach to Alvarez, Engbom and Moser (2015) and Card, Heining and Kline (2013), I find the largest set of firms connected through mobility of workers. I compute this set using a graph theoretical algorithm included in the MatlabBGL library. Because of the high mobility of workers in each of the sub-periods, these connected sets contain more than 95 percent of all workers and more than 70 percent of all firms.

To estimate the model, I have to solve the normal equations. However, the model contains hundreds of millions of worker and firm effects in every period, which poses a computational challenge. I follow Card et. al. (2013) in approximating the system of normal equations without actually inverting the matrix using an iterative gradient method. By doing this, I obtain a vector of estimated firm effects for each period.

Figure 16 shows the distribution of firm-level premiums \(\phi_{J(i,t)}\) for the three sectors in the economy for the 1996-2001 and 2008-2013 periods. As mentioned before, agricultural firms pay lower average premiums relative to agriculture and services. However, there is great heterogeneity and overlap between the different sectors which cause the mean effect between sectors to be small. In fact, even in the early period when sector premiums used to be higher, there is great overlap in the distribution of firm level premiums between agriculture and other sectors. For example, manufacturing firms did pay a premium relative to agriculture in the early periods, but there were also better agriculture firms that paid better premiums than manufacturing firms. By 2008-2013, the differences were greatly reduced but a significant wage gap remained.

Thus, consistent with the literature on the importance of firms in wage setting, firm premiums exist in Brazil, but mean differences in mean firm effects in relation to other sectors are much smaller than the overall wage gap. Since we are concerned with inter-sector differences, this observation motivates the study of sectors as uniform entities and a focus on worker-side mechanisms in explaining the between-sector wage gap magnitude and its decline.
Figure 16: DISTRIBUTION OF FIRM EFFECTS ($\phi_{J(i,t)}$) BY SECTOR

(a) 1996-2001

(b) 2008-2013