Title
Essays on Labor Markets

Author
Guanziroli, Tomas

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Labor Markets

A thesis submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Tomás Guanziroli

2022

ABSTRACT OF THE DISSERTATION

Essays on Labor Markets

by

Tomás Guanziroli

Doctor of Philosophy in Economics

University of California, Los Angeles, 2022

Professor Adriana Lleras-Muney, Chair

This dissertation consists of three essays in the area of Labor Markets. In Chapter 1, I use a merger in the retail pharmacy sector in Brazil to study the effects of concentration on labor market outcomes. I observe a larger reduction in wages of salespeople than pharmacists. The results are consistent with the idea that salespeople have strong preferences for jobs or accumulate some industry-specific human capital, and that pharmacists are better organized into unions. In Chapter 2, co-authored with Ryan Boone, we use worker flows across occupations, industries, and geographies to define better approximations to labor markets. The defined markets maximize a measure of the density of links within markets. In Chapter 3, co-authored with Ariadna Jou Fuya, we investigate whether the flattening in the college premium in Brazil is due to changes in the average quality of college graduates. We show that the supply of workers with college degree has increased, but much of this increase came from newer, lower ranked and lower wage-premium universities. The college premium has actually increased when we hold constant a set of universities. There are more workers with a college degree, but with lower quality degrees, which reflects into lower average wages.

The dissertation of Tomás Guanziroli is approved.

Rodrigo Ribeiro Antunes Pinto

Till von Wachter

John William Asker

Adriana Lleras-Muney, Committee Chair

University of California, Los Angeles

2022

To my mother Diana, my sister Anna and my father Carlos for their unconditional support

To my fiancée Ariadna for her company throughout this journey

Contents

1 Does Labor Market Concentration Decrease Wages? Evidence from a Retail Pharmacy Merger **3**

1. Introduction 3

2. Merger Predictions in Oligopsony Models 8

3. Data and context 10

3.1 Employer-employee matched dataset 10

3.2 Institutional setting 11

3.3 The merger between two major retail pharmacy chains 12

4. Identification Strategy and Empirical Specification 14

4.1 Difference-in-differences with worker and establishment fixed effects . 16

4.2 Identifying assumptions 18

5. Effect of the Merger in Merging Firms 20

5.1 Effect of the merger on overall wages 20

5.2 Effect of the merger on wages, accounting for changes in composition 22

5.3 Effect of the merger on other labor market outcomes 24

6. Effect of the Merger in Competing Firms 25

7. Discussion and Magnitudes 26

7.1 Magnitudes 29

8. Conclusion 30

9. Figures and Tables 32

2 Defining Labor Markets from Worker Flows **43**

1. Introduction 43

2. Literature has different Local Labor Market Definitions 46

3. The Brazilian Matched Employer-Employee Dataset 47

4. A Method to Define Labor Market Approximations 49

4.1 Algorithm and implementation 51

5. Characterizing Labor Markets 51

6. Discussion 57

7. Figures and Tables 58

3 Has the College Premium really Flattened? The Entry of New Schools **64**

1. Introduction 64

2. Data Sources and Descriptive Statistics 67

3. Preliminary Evidence of Changes in Graduates' Skill Composition 70

4. The Adjusted College Premium Trends 71

5. Conclusion 73

6. Figures and Tables 74

A Appendix Materials for Chapter 1 **82**

1. Data Appendix 82

2. The retail pharmacy sector in Brazil 83

3. Appendix Figures and Tables 87

B Appendix Materials for Chapter 2 **101**

1. Appendix Tables 101

C Appendix Materials for Chapter 3 **104**

1. Matching the College Graduates Sample with RAIS 104

2. Appendix Figures and Tables 106

List of Figures

1.1 Presence of merging firms before the merger (2010) 32

1.2 HHI projected with mergers versus 2011 HHI 33

1.3 Trends - Ln(wage) of all workers in merging firms 34

1.4 Leads and lags - Ln(wage) of all workers in merging firms 34

1.5 Leads and lags - Ln(wage) of workers in merging firms 35

1.6 Leads and lags - Ln(wage) of pharmacists in merging firms with two-way fixed effects 36

1.7 Leads and lags - Ln(wage) of pharmacists in merging firms 37

2.1 Nationally dispersed markets 58

2.2 Markets in neighboring states 59

2.3 State dispersed markets 59

3.1 College/High School wage premium 74

3.2 Formal workers with a college degree 75

3.3 Number of graduates by major's date of foundation 76

3.4 Share of graduates from Top 25 Universities according to ENADE ranking . 77

3.5 Trends in residualized log wages (college premium) 78

A.1 Event Study of Changes in Earnings when Workers Move Between Firms . . 87
A.2 Trends - Ln(wage) of workers in merging firms 88
A.3 Residual ln(wage) of pharmacists in merging firms with two-way fixed effects 89
A.4 Job transition probabilities conditional on moving to another drugstore . . . 90
A.5 Pharmacists per capita - Brazil 91
A.6 Trends in pharmacists' gender composition - Brazil 91
A.7 Trends in pharmacists' race composition - Brazil 92
A.8 Age distribution of pharmacists - Brazil 92
A.9 Trends in pharmacists' wage distribution - Brazil 93

C.1 Number of graduates by institution's date of foundation 106
C.2 Trends in residualized log wages 107

List of Tables

1.1 Sample size 38
1.2 Summary Statistics of treated and control groups 39
1.3 Difference-in-differences estimates. Dependent variable is Log(wage) 40
1.4 Difference-in-differences estimates. Other outcomes 41
1.5 Difference-in-differences estimates. Other outcomes - Competitors 42

2.1 Job transitions and sample size 60
2.2 Descriptive Statistics of Markets 61
2.3 Labor market concentration and number of markets by firm 62
2.4 Types of markets by geography 62
2.5 Market composition 63

3.1 University's age 78
3.2 Institutions ranking and score 79
3.3 Association between university's age and quality 80
3.4 Correlation between University ranking and wages 81

A.1 Difference-in-differences estimates. Dep. variable is ln(wage) 94
A.2 Robustness. Difference-in-differences estimates. Dep.variable is Log(wage) . 95

A.3 Robustness. Difference-in-differences estimates. Dep. variable is Log(wage) 96

A.4 Job transition for all workers 97

A.5 Job transition for movers 97

A.6 Merger-induced changes in concentration 98

A.7 Sample size 99

A.8 Descriptive statistics on the retail pharmacy sector 100

B.1 Recent Literature on Labor Market Concentration 101

B.2 Recent Literature on the Effects of Minimum Wages 102

B.3 Recent Literature on the Effects of Trade Shocks 102

B.4 Sample size 103

B.5 Labor market concentration 103

C.1 Schools with access to the data 108

C.1 Schools with access to the data (cont.) 109

C.2 Sample size by university 110

C.2 Sample size by university (cont.) 111

C.3 Comparing Matching Algorithms 112

C.4 Confusion Matrix—Out of Sample Predictions 112

ACKNOWLEDGMENTS

I am deeply grateful to many people for their support during the doctorate. I thank my advisor, Adriana Lleras-Muney, for her patience, support and humility. Next to my father, Adriana became one of my role models in the academic profession. By observing her interact with students and faculty I learned that cordial questioning is the key to a vibrant research environment. Future students of mine are allowed to call on this passage to request better treatment—which I hope to always provide. I have also benefited from the time, advice and feedback from many professors; including (but not limited to) Daniel Haanwinckel, John Asker, Till von Wachter, Rodrigo Pinto, Moshe Buchinsky, Bernardo S. da Silveira, and Martin Hackman.

During the past six years, I have benefited from the interaction with many people, including Jacob Berman, Alvaro Boitier, Ryan Boone, Alex Coblin, Emmanouil Chatzikonstantinou, Ricardo Dahis, Sepehr Ekbatani, Domenico Fabrizi, Benjamin Freyd, Diana Flores, Alex Fon, Renato Giroldo, Jonathan Gu Ariadna Jou, Ivan Lavrov, Tina Mengshan Cui, Fatih Ozturk, Mariano Palleja, Rustin Partow, Kiril Ponomarev, Brian Putininsky, Fernanda Rojas, Joaquin Serrano, Liqiang “Leo” Shi, Sumit Shinde, and many others.[1] I am glad that I met you and I hope we keep in touch.

As a resident of UCLA’s California Center for Population Research (CCPR), I thank CCPR affiliates, professors and staff for providing invaluable resources. Finally, I acknowledge financial support from the Economics Department at UCLA, and the Dissertation Year Fellowship from the Graduate Division at UCLA.

[1] Calvin, Manu and Daniel did not make the cut.

VITA

Tomás Guanziroli

Education

Pontificia Universidade Católica do Rio de Janeiro (PUC-Rio)

Master of Arts in Economics	2012-2014
Bachelor of Science in Economics	2008-2011

Relevant Positions

University of California Los Angeles (UCLA)

Research Assistant to Professor Adriana Lleras-Muney	2018-2019
Teaching Assistant	2017-2021

Pontificia Universidade Católica do Rio de Janeiro (PUC-Rio)

Research Assistant to Professors Gustavo Gonzaga and Claudio Ferraz	2015-2016
Research Assistant to Professor Pedro C.L. Souza	2015

Inter American Development Bank (IADB)

Research Fellow	2014-2015

Fellowships, Honors, and Awards

Dissertation Year Fellowship, UCLA	2021-2022
Best Paper Award, Proseminar in Industrial Organization , UCLA	2021
Outstanding teaching assistant performance (awarded to 10% of TAs)	2020

Introduction

This dissertation consists of three essays in the area of Labor Markets

In Chapter 1, I use a large merger to study the effects of a concentration on labor market outcomes. I find that concentration lowers wages, but the effects are different than prior estimates in the literature. I observe a larger reduction in wages of salespeople (low skill) than pharmacists (high skill). Two hypotheses can explain the results: (i) salespeople have strong preferences for jobs or accumulate some industry-specific human capital, and (ii) pharmacists are better organized into unions. Previous studies do not account for changes in labor force composition that are relevant after a merger, which could explain the difference in estimates.

In Chapter 2, co-authored with Ryan Boone, we use worker flows to define better approximations to labor markets. Labor market definitions are important to predict the effects of government policies and other events on labor market outcomes. Researchers have typically relied on ad-hoc approximations such as geographic boundaries, occupational codes, or industry codes at differing levels of granularity. For any given job, it is not clear which definition is most appropriate. The chapter follows two steps. First, we use an approach from the network literature to identify labor markets. We use worker flows across occupations, industries, and geographies to define N markets. The defined markets maximize a measure of the density of links within markets. Second, we describe the resulting markets. For example, we show that labor markets often cross regions, industries and occupations.

In Chapter 3, co-authored with Ariadna Jou Fuya, we investigate whether the flattening in the college premium in Brazil is due to changes in the average quality of college graduates. We do this by matching novel data with the names of around one million college graduates from 42 schools and 20 different cohorts with the RAIS, a Brazilian employer-employee matched dataset. First, we show that the supply of workers with college degree has increased, but much of this increase came from newer, lower ranked and lower wage-premium universities. Second, we show that the college premium has actually increased when we hold constant a set of universities. There are more workers with a college degree, but with lower quality degrees, which reflects into lower average wages. The findings in this paper are relevant for any study that uses college premium proxies in countries, or periods of time, with increasing access to lower quality colleges.

Chapter 1

Does Labor Market Concentration Decrease Wages? Evidence from a Retail Pharmacy Merger

1. Introduction

The fall in the labor share and recent increase in inequality (Elsby et al., 2013; Karabarbounis and Neiman, 2013; Piketty and Zucman, 2014) have led economists and policymakers to consider whether increasing labor market concentration has decreased wages.[1] This argument is supported by indirect evidence of the monopsonistic behavior of firms, such as the small employment effects of minimum wage increases (Card and Krueger, 1994; Manning, 2011; Azar et al., 2019); the small labor supply elasticity estimates (Falch, 2010; Staiger et al., 2010; Matsudaira, 2014); and the presence of anticompetitive agreements (Krueger and Ashenfelter, 2018; Starr et al., 2021). However, direct evidence that increased concentration reduces

[1] For example, the following passages from the Executive Order by President Joseph Biden of July 2021 state: "It is the policy of my Administration to enforce the antitrust laws to combat the excessive concentration of industry, the abuses of market power, and the harmful effects of monopoly and monopsony—especially as these issues arise in labor markets" and "Consolidation has increased the power of corporate employers, making it harder for workers to bargain for higher wages and better work conditions."

wages is scarce.

This paper estimates the effect of concentration on labor market outcomes by analyzing a large merger in Brazil's retail pharmacy sector. Mergers can lead to greater concentration but can also generate synergies. To separately identify the effect of concentration, I chose to study a merger that increased concentration in some markets but not others. The effect of market power is estimated using a difference-in-differences (DiD) approach that compares the labor market effects of the merger in regions in which concentration increased relative to those in which it didn't. To account for composition effects, I implement a novel DiD estimator that includes establishment and worker fixed effects. Finally, I analyze how competitors respond to the merger, which is another important determinant of overall labor market outcomes.

The data used in this analysis have many advantages. I use a matched employer-employee database that covers the universe of formal employees in Brazilian companies. The data are structured as a panel between 2007 and 2018, which allows me to follow establishments over time. Furthermore, it is possible to assess how the merger affected the wages of workers in the firms that merged, including both incumbents and new entrants. The data covers all firms in Brazil, such that I can also assess the responses of competing firms and use all the other firms to identify and estimate worker fixed effects. Lastly, because informality levels in the retail pharmacy level are much lower than in other sectors (partly because it is a highly regulated sector), the results capture the relevant population effects.

I find that labor market concentration lowers wages, but the effects are different than prior estimates in the literature. This difference is due to composition effects. Without accounting for composition, I find that the wages of pharmacists fell by 7.9% and those of salespeople by 0.7%. These estimates are very similar to those of Prager and Schmitt (2021) in the US, despite the differences in setting and identification strategy.[2] However, when firm and

[2]Prager and Schmitt (2021) analyze the effects of many mergers, not just one. They analyze hospital mergers and find that the wages of nursing and pharmacy workers fall by 6.8% in locations in which mergers increased concentration. They do not find evidence of differences in wage growth for unskilled workers.

worker fixed effects are included, the wages of pharmacists only fall by 2.6%. This suggests that the overall decline in wages was partly due to churning among pharmacists: After the merger, newly hired pharmacists were of lower quality, as measured by their estimated fixed effects. In contrast, the wages of salespeople fall by 3.7%. These results show that failing to account for composition effects biases estimates of the effects of concentration.

These wage results are not negligible when considering merging firm's market share. Overall, my estimates suggest that the elasticity of wages with respect to labor market concentration ranges from -0.34 to -0.1 for pharmacists and from -17.5 and -0.13 for salespeople, depending on the labor market definition used to compute concentration measures. The large variation in elasticity for salespeople demonstrates the relevance of properly defining labor markets. The effects of the merger on employment are not precisely estimated, and I cannot rule out a significant decline in employment for both occupations.

The effects of mergers on competing firms are theoretically ambiguous and depend on whether firms are strategic complements or substitutes. I find that in markets in which concentration increased and wages in merging firms decreased, competing firms responded by increasing employment. Even though competing firms hire one more salesperson after the merger, the wages of salespeople at competing firms still fall by 3.3%. This suggests that firms are strategic substitutes—i.e., they respond to other firms' hiring strategies by doing the opposite of what their competitors do. This new finding helps explain why the estimated effects are modest, despite the fact that concentration increased: the labor market effects of a merger are limited by competitors' responses. If firms were instead strategic complements, we would likely observe a more sizeable wage decrease for all workers after the merger.

The results are counterintuitive, in that ex ante we expect pharmacists to have fewer outside options relative to salespeople, and thus face larger wage declines when concentration increases. To explain this, I estimate the effects of the merger on incumbents and newly hired workers separately. I find that the wages of incumbent pharmacists do not fall, but the wages of newly hired pharmacists do. These results are consistent with the higher unionization rate of pharmacists in Brazil.[3] Unions can protect incumbent workers by negotiating wage floors and annual wage increases. However, in a scenario in which wage floors are not binding—as is the case for merging firms—unions do less for newly hired workers. This explains why the wages of pharmacists don't fall as much as expected.

The finding that wages of salespeople fall suggests that the preconceived idea that low-skill workers have greater outside options is inaccurate; salespeople working at drugstores may have some degree of firm- or industry-specific human capital or stronger preferences to work at a drugstore.[4] To assess this, I analyze data on mobility and show that 24% of salespeople employed in drugstores move to another drugstore when they switch jobs. If they were moving randomly across jobs with the same occupation title, this number should be closer to 5%. Labor markets for low-skilled workers might be finer and more concentrated than they appear to be.

This paper makes two contributions to the literature. The first and most important contribution is implementing an estimator that accounts for composition effects. To my knowledge, I am the first to combine two widely used empirical approaches: DiD regression and two-way fixed-effects regressions (usually referred to as AKM to reflect the work of Abowd, Kramarz, and Margolis, 1999). The AKM approach consists of estimating a log wage regression with worker and firm fixed effects and is generally used to measure the contributions of workers and firms to earnings dispersion (Card, Heining, and Kline, 2013; Lamadon et al., 2021; Song et al., 2018, among many others). The method requires the use of large datasets in which we observe the same workers across firms, since this is the only

[3]In the 2013 Brazilian household survey, 34% of pharmacists working in drugstores reported being unionized. Only 9% of salespeople working in drugstores reported being unionized (PNAD, 2013).

[4]Neal (1995); Parent (2000); Poletaev and Robinson (2008); Kambourov and Manovskii (2009); Sullivan (2010) and many others present evidence that workers accumulate firm-, occupation-, and industry specific human capital. With upward an slopping firm-specific labor supply, firms may act as monopsonists. The models in Card et al. (2018) and Lamadon et al. (2021) have workers with heterogeneous preferences over non-wage job characteristics that view firms as imperfect substitutes.

consistent way to separately identify worker and firm fixed effects. I show that it is feasible to use this method to control for changes in the unobserved characteristics of workers in a DiD empirical strategy.

A growing literature examines the overall labor market effects of mergers and acquisitions (Brown and Medoff, 1988; McGuckin and Nguyen, 2001; Li, 2012; DePasquale, 2018; Todd and Heining, 2020; He and Maire, 2020; Gehrke et al., 2021; Lagaras, 2020). While these papers find mixed results on wages, they mostly agree that merging firms restructure their labor force. For example, Gehrke et al. (2021) show that acquiring firms hire younger workers after the merger. I contribute to this literature by showing that merging firms may also restructure their labor force in unobservable ways, as captured by workers' fixed effects. I find that accounting for these changes is important in understanding the effects of mergers and concentration on wages. This is true when following establishments over time (as in this paper and in Prager and Schmitt, 2021) but also when following a cohort of workers over time (as in Arnold, 2021).

As a second contribution, I provide causal evidence that concentration affects wages by studying a single merger and comparing establishments from the same firms. I identify the market power effect with the assumption that productivity gains are the same for establishments within merging firms regardless of whether they are in a market in which concentration increases. Azar et al. (2022); Qiu and Sojourner (2019); and Rinz (2020) show that labor market concentration is associated with lower wages. A few studies have used mergers as an instrument for the increase in concentration. Arnold (2021); Benmelech et al. (2022); and Prager and Schmitt (2021) use variation from many mergers and highlight the market power mechanism that allows firms to reduce wages.[5] These papers rely on the assumption that mergers between different companies will have similar productivity gains, regardless of their effect on labor market concentration. I add to this literature by studying a single merger, in

[5]A recent literature also shows that firms with market power pay lower prices for agricultural inputs (Giroldo and Hollenbeck, 2021; Rubens, 2021) and manufacturing inputs (Morlacco, 2020).

which the assumption that synergies are identical across establishments within the merging firm is more likely to hold.[6]

2. Merger Predictions in Oligopsony Models

The purpose of this section is to explain why, as a matter of theory, the impact of a merger on labor market outcomes is ambiguous. Mergers may change the structure of product and labor markets, and changes in both markets can spillover to workers.[7] In this paper, I refer to market power as the combined effect of product and labor market power. Below, I explain how the prediction on wages and employment, at merging firms and competing firms may differ between (i) monopsony, (ii) search friction and bargaining, and (iii) collective bargaining models.[8] Which model better represents the reality is an empirical question.

In **modern monopsony models** (Card et al., 2018;Berger et al., 2022), a merger between two firms will decrease wages and employment at the **merging firms**. This can occur both due to labor market power and product market power. In the labor market, merging firms face upward slopping labor supply curves and internalize that their hiring decisions will affect both firms' labor supply. Therefore, each firm hires less workers such that both firms can pay lower wages. In the product market, merging firms face downward slopping product demand and internalize that by decreasing their production both firms can receive higher prices. The decrease in production means that firms will have to decrease their labor inputs, which leads to a decrease in wages.[9]

[6]Nevo and Whinston (2010) and Berry et al. (2019) review the empirical industrial organization literature in the last 30 years and highlight the benefits of studying a single industry or even a single event.

[7]Mergers may also have synergies that affect productivity, wages, and employment. The theoretical framework and the empirical analysis in this paper exclude such effects because synergies might be specific to each merger.

[8]See Card (2022) for a discussion of the evolution of these models.

[9]Note that this is a partial equilibrium effect since it does not take competing firms' response into account.

In this model, a merger will reduce wages at **competing firms** but the effect on employment is ambiguous. There are two mechanisms at work: income and competition. The income mechanism states that labor is now cheaper, so competing firms hire more workers. The competition mechanism depends on whether firms are strategic substitutes (Cournot) or strategic complements (Bertrand). Firms that are strategic substitutes will hire more workers when their competitors hire less workers. Firms that are strategic complements will lower their wage offer and hire less workers when competing firms lower their wage offer. In equilibrium, employment could either increase or decrease at competing firms.

In **search models**, a merger is predicted to reduce wages and keep employment unchanged. In Jarosch et al. (2019), a merger reduces the number of firms but not the number of job vacancies. As a consequence, workers are matched to the same number of jobs but to a lower number of firms. By having fewer outside options, workers have a weaker bargaining position. In this scenario, a merger reduces wages paid by merging and competing firms, but unlike the monopsony model, employment remains unchanged.

When considering product market concentration within a search model, a merger could potentially increase wages. For example, suppose that product demand is perfectly inelastic. In that case, merging firms can increase prices without decreasing production. Employment does not change, and the surplus created by the job match increases. Even though workers have lower bargaining power, total surplus may increase to the point that wages also increase.

In **collective bargaining models**, a merger might not have any effect on the labor market. This could happen when workers at both merging firms are represented by the same union before the merger and wages were set jointly—like it is for some unions in Brazil. In this case, a merger would not change the bargaining relation between unions and employers and would not change wages nor employment. In other countries and institutional settings, the model can have different predictions.

The results of this paper suggest that the labor market of salespeople is consistent with modern monopsony models and the labor market of pharmacists is consistent with collective bargaining models. Such heterogeneity highlights the benefits of studying a particular industry.

3. Data and context

3.1 Employer-employee matched dataset

I use the Brazilian employer-employee matched dataset (RAIS—Relação Anual de Informações Sociais). RAIS is a restricted-access longitudinal dataset of administrative records collected by the Brazilian Ministry of Employment and Labor that covers all formal workers and firms in Brazil.[10]

The dataset includes information on firms (legal form of the company); establishments (industry sector and county); workers (age, gender, race, and education); and job characteristics (occupation, date of hiring, date of separation, hours worked, tenure, average wage, and wage in December). A unit of observation in the data is a job relation between a worker and an establishment in a given year. Using this information, I am able to track workers throughout their formal job history.

The empirical strategy used in this paper requires two different data samples. The first sample—the drugstores sample —is restricted to firms in the retail pharmacy sector that had at least one pharmacist working between 2007 and 2018. I further restrict the sample to a balanced panel of establishments—i.e, establishments that had at least one employee every year between 2007 and 2018. In the second sample—the fully connected sample—I keep the establishments from all sectors in the economy that are connected to establishments in the drugstore sample through worker mobility. The sample is restricted to workers who switched

[10]The main purpose of RAIS is to administer a federal wage supplement (Abono Salarial) to formal employees. There are incentives for truthful reporting. In principle, an employer's failure to report the information can result in fines proportional to the firm size.

jobs at least once between 2007 and 2018.[11] See Appendix 1. for a more detailed description of the data restrictions.

Table 1.1 presents sample size numbers by merging firms and competitors. The balanced panel of drugstores is spread among 91 counties and contains 392 establishments. Between 2007 and 2018, 51,380 individuals worked in one of these establishments, resulting in 134,216 person-year observations. The second column of Table 1.1 presents the sample size for competing firms in the same counties: 9,798 firms that employed 497,658 different individuals during the period. The fully connected set consists of 189 million observations from 29 million individuals working in 4.2 million establishments.

3.2 Institutional setting

In this section, I present some key similarities and differences between the Brazilian and international settings that are relevant to this study of the retail pharmacy sector.

Retail pharmacies, or drugstores, are facilities that sell medication, cosmetics, and pharmaceutical products. Pharmacies may also administer vaccines and compound medication, and can sell some food products for special purposes (Sebrae-SP, 2015). While drugstores in Brazil are similar to U.S. drugstores in terms of the products sold, they tend to differ in size. In Brazil, drugstores are smaller in size and number of employees. An average Brazilian drugstore has around 8 employees, with one or two pharmacists, while an average American drugstore has around 14 employees. In Appendix 2., I present more detailed information on the retail pharmacy sector in Brazil.

In Brazil, the relation between employers and employees is mediated through unions, with some key differences with the rest of the world. Workers are automatically associated with a union that represents their category. For example, most states have a pharmacists' union and a union representing other workers at drugstores. The smallest representing unit

[11] Due to this restriction, the drugstore sample is not a subset of the fully connected sample.

is the city, with most unions representing workers in a group of counties or within the state (Menezes-Filho et al., 2011). In contrast, U.S. workers tend to organize at the firm or plant level.[12] Another singularity of the Brazilian case is the existence of employer unions that represent groups of employers in a region. The unions of employers and employees bargain for wage floors and wage increases, among other conditions of employment, in regions in which they overlap. Hence, unions might have to sign multiple collective agreements within a year.

The Brazilian analytical process for mergers and anti-competitive agreements is closely related to the conduct of the U.S. Department of Justice and Federal Trade Commission. Firms above a certain revenue threshold that wish to merge must pass through the scrutiny of CADE (Conselho Administrativo de Defesa Economica), the antitrust agency. CADE evaluates whether the merger significantly increases concentration in the relevant market and decides whether to approve the merger, propose remedies or deny the merger. The law that regulates CADE is ambiguous regarding whether the agency can stop a merger based on labor markets concerns.[13]

3.3 The merger between two major retail pharmacy chains

I study a merger between two large retail pharmacy chains in Brazil.[14] This was a horizontal merger in the sense that the firms operated in the same markets and sold substitute goods. Firms announced the merger in 2011 and the Brazilian antitrust agency unanimously

[12] In the U.S., union membership is voluntary, with unions being able to represent workers at the plant level. Membership is also free in signatory countries of Convention 87 of the ILO (Freedom of Association and Protection of the Right to Organise Convention, 1948).

[13] Law number 8,884, from 1994, and law number 12,529, from 2011, establish that the agency must intervene when companies harm competition in the "relevant market". That said, in 2021, for the first time ever, CADE opened a case to investigate a cartel between human resource departments from the pharmaceutical industry and medical product suppliers (the cartel did not operate in the retail pharmacy sector). http://valor.globo.com/legislacao/noticia/2021/03/24/cade-investiga-formacao-de-cartel-entre-departamentos-de-recursos-humanos.ghtml

[14] For confidentiality reasons, I do not reveal the firms' names.

approved the merger in 2012, nine months after the announcement.

An attractive feature of the merger is that firms overlapped in some counties and did not overlap in others, which provides plausible exogenous changes in concentration. Although other mergers and acquisitions occurred during the time frame of the analysis, the merger I study is the only one in which firms' operations overlapped in many counties. Figure 1.1 presents a map of the nine southern states from Brazil to show the presence of the merging firms in each county before the merger. Counties in which both merging firms had an establishment in 2010 are drawn in red. Counties in which only one of the merging firms had an establishment in 2010 are drawn in yellow and orange. The white area represents counties in which none of the firms had an establishment in 2010.

The merger changed labor market concentration in some regions. Panel A of Figure 1.2 shows that the merger increases the pharmacists' labor market Herfindal-Hirschman index (HHI) in many counties.[15] Panel B shows that the merger did not change concentration in the labor market of salespeople, where the market is also defined at the county-occupation intersection. Given these pictures, it is reasonable to assume that a merger between two retail pharmacy chains will affect pharmacists and salespeople in different ways. In theory, pharmacists have fewer outside options than salespeople, with the job opportunities of pharmacists being restricted to drugstores, hospitals and pharmaceutical companies. Salespeople, on the other hand, may work in a similar position at drugstores, supermarkets, general stores, and any other retail shop. That said, Figure 1.2 uses ad-hoc labor market definitions. In section 7., I discuss how finer labor market definitions change our ex-ante hypothesis regarding

[15]I measure concentration in the labor market using the HHI, as described in the equation below. There is an HHI index for each occupation O in county C at time t. The index is constructed by summing the square of firm f's occupation shares. Note that the set of firms $f \in F$ is restricted to firms that had at least one employee from occupation o in the years previous to the merger. Hence, the number of firms that employed pharmacists is smaller than the number of firms that employed salespeople. The HHI varies between zero and 10,000, with higher values indicating a more concentrated labor market.

$$HHI^{o}_{ct} = \sum_{f} (\text{Share of occupation o workers})^2_{fct}.$$

the effect of the merger.

Table 1.2 presents some descriptive statistics of counties and workers from merging firms in the treated and control groups. On average, counties in the control group are smaller in population and have lower per capita government revenue than counties in the treated group. Counties in the overall sample had lower informality rates than the nationwide average (27% versus 40%) and HDI levels comparable to countries such as Spain or Greece (measured in 2000). There are no significant differences in informality or HDI between treated and control groups.

While baseline characteristics or workers and counties in treated and control groups may differ, this does not trigger great concern about identification. The empirical strategy in this paper adopts a differences-in-differences approach and the identifying assumption requires that the outcomes of interest in both groups follow similar trends—and not levels—which I discuss now in detail.

4. Identification Strategy and Empirical Specification

Section 2. argues that theory yields ambiguous predictions regarding the effects of a merger. Hence, whether merger-induced concentration reduces wages is an empirical question. To identify the effects of concentration on labor market outcomes, I use a difference-in-differences (DiD) design applied to the study of a single merger in the retail pharmacy sector in Brazil.

The DiD compares labor market outcomes before and after the merger in treated versus control counties. Merging firms are present in many locations, so I use the variation in the firms' overlap within locations to define the treatment and control group. I denote as treated counties in which both merging firms had at least one establishment in 2010, one year before the merger was announced. I denote as control counties in which only one of the firms had an establishment in 2010.

I separate the analysis into two samples: establishments of merging firms and establishments of competing firms. In both cases, I restrict the sample to establishments that had at least one employee in all years between 2007 and 2018. Besides separating the sample by type of firm, I estimate the regressions for pharmacists and salespeople separately. The empirical specification, at the establishment level, is of the form

$$y_{jt} = \delta_0 + \delta_1 post_t + \delta_2 treat_{c(j)} + \delta_3 post_t \times treat_{c(j)} + \varepsilon_{jt}, \tag{1.1}$$

where y_{jt} is an outcome such as average wage or employment in establishment j in year t. The indicator $post_t$ equals one for observations after 2011; $treat_{c(j)}$ is a variable that indicates whether establishment j is located in a treated county c; and ε_{jt} is an error term. Additionally, I estimate a leads and lags equation, presented below. In the equation, $\mathbb{1}[t = k]$ indicates the year relative to the merger; λ_t are year fixed effects; and ε_{jt} is an error term.

$$y_{jt} = \beta_1 treat_{c(j)} + \sum_{\substack{k=-4 \\ k\neq 0}}^{7} \delta_k \mathbb{1}[t = k] \times treat_{c(j)} + \lambda_t + \varepsilon_{jt}. \tag{1.2}$$

Using establishment-level data to analyze changes in wages has a major limitation: As establishments hire and fire workers, their labor force composition changes over time. Furthermore, the merger may also induce changes in labor force composition. If, for example, firms hired younger employees after the merger, we would naturally observe a fall in average wages. In this case, the DiD estimator above is not informative regarding how the merger changes market wages—i.e., we cannot use this estimator to infer wages in the counterfactual scenario in which the merger does not occur. To identify this effect, I develop the identification strategy at the individual level, which accounts for changes in labor force composition in observable and unobservable characteristics, next.

4.1 Difference-in-differences with worker and establishment fixed effects

The main empirical specification in this paper is an extension of the DiD model that adds worker and establishment fixed effects and uses the fully connected sample. The inclusion of worker fixed effects allows me to control for all of time-invariant characteristics of workers, including unobserved skill. In this way, I can estimate the actual effect of the merger on market wages, net of observable and unobservable changes in labor force composition.

The equation below incorporates worker and establishment fixed effects in the DiD model.[16]

$$\begin{aligned} \ln(wage)_{it} = \theta_i + \psi_{J(i,t)} &+ \sum_{k \neq o,o'} \delta_0^k \mathbb{1}[Group_k] \\ &+ \sum_k \delta_1^k \mathbb{1}[Group_k] \times Post_t \\ &+ \sum_k \delta_2^k \mathbb{1}[Group_k] \times Treat_c \\ &+ \sum_k \delta_3^k \mathbb{1}[Group_k] \times Treat_c \times Post_t \\ &+ X_{it}'\beta + \lambda_t + \varepsilon_{it}, \end{aligned} \tag{1.3}$$

where log wages of individual i at time t are separable into worker fixed effects, θ_i; establishment fixed effects, $\psi_{J(i,t)}$, with the subscript $J(i,t)$ referring to establishment J in which individual i was working at time t; time-varying individual characteristics, X_{it}; and year fixed effects λ_t. The equation also includes four sums over the groups k. The terms inside the sums correspond to the DiD terms for each group. For example, the parameter δ_0^1 is analogous to the parameter δ_0 from Equation 1.1. There are six groups from the interactions between two types of firms (merging firms and competing drugstores) and three occupation categories (pharmacists, salespeople, and other occupations). I omit δ_0^k for groups o and o' to avoid collinearity with establishment fixed effects, where o and o' refer to the groups "other

[16] In Appendix **??**, I rewrite Equation 15 in its extensive form for easier exposition.

occupations in merging firms" and "other occupations in competitors" As in the previous section, the indicator $Post_t$ equals one for observations after 2011, and $Treat_c$ is an indicator for observations in treated counties. I am interested in the parameters δ_3^k, which recover the effect of the merger on wages for each group k, excluding composition effects.

The purpose of including establishment fixed effects in Equation 1.3 is to prevent other parameters from being biased. There are several theories regarding why different establishments have different levels of $\psi_{J(i,t)}$: Establishments may differ in their productivity levels, in amenities, or in market power in the labor market. Given that, individual fixed effects and other parameters will be biased if we do not include establishment fixed effects. For example, a worker who switches from a low-productivity establishment to a high-productivity establishment is likely to receive a wage increase. I would mistakenly assign the establishment productivity effect to that individual (or to individual time-varying characteristics) if I did not include an establishment fixed effect.

As a consequence, I have to extend the drugstore estimation sample to the fully connected sample. The fully connected sample is necessary in order to jointly identify worker and establishment fixed effects—a well-known problem in the literature.[17] For example, I cannot separately identify worker and establishment effects if an establishment is composed of workers who never switched establishments. In my setting, workers in the retail pharmacy sector have also worked in firms from other sectors. Hence, I keep the set of establishments from all sectors in the economy that are connected to establishments in the drugstore sample through worker mobility.

Equation 1.3 includes all six groups, such that I must estimate 22 δ^k parameters jointly. I explain why this is the case with the following example. Suppose the merger reduces the wages of pharmacists at merging firms and competing firms by the same amount. Then suppose that a pharmacist working at a merging firm switches firms after the merger and

[17]For example, see Abowd et al. (1999) and Card et al. (2013).

starts working at a competing drugstore, still as a pharmacist. We would expect a reduction in the wage of this pharmacist that is attributable to the merger and not the worker. However, if I did not include the DiD parameters for the competing firms, I would estimate a lower establishment fixed effect for the competing firm, which in turn would bias the estimate of that worker's fixed effect, which in turn would bias the δ_3^k parameter for pharmacists in merging firms. Thus, all DiD parameters must be jointly estimated.

4.2 Identifying assumptions

There are three main challenges when trying to identify the effects of a merger on wages through market power: (i) other events might occur at the same time as the merger; (ii) firms may change their labor force composition after the merger; and (iii) merger-induced synergies may also affect wages, which makes it difficult to identify the market power effect. Note that antitrust agencies such as the DOJ, FTC, and CADE are interested in the total effect of mergers on prices and wages, which includes the effect through synergies. I am only interested in identifying the effects through market power, excluding synergies.

The DiD approach solves the first challenge by taking differences with a control group. The underlying assumption is the parallel trends assumption: Conditional on controls, wages in the control group would have followed the same trends as wages in the treated group, if they had been treated. In the results section, I provide suggesting evidence that this assumption holds by showing that the pre-trends are parallel. Yet, counties may have different trends associated with characteristics that are not balanced between treated and control groups (see Table 1.2), which violates the parallel trends assumption. To address this concern, I control for interactions between county characteristics and year fixed effects in the main specification. Results do not change when I do this.

The second challenge—accounting for changes in the labor force composition—is addressed with Equation 1.3. Identification relies on three assumptions: The log-linearity of wages, additivity, and exogenous mobility. The log-linearity of wages is a widely accepted assumption in labor economics. The additional additivity assumption implies that the effect of the treatment is not heterogeneous by individual characteristics. Exogenous mobility implies that workers do not select into firms based on the idiosyncratic error term, ε_{it}. In Appendix Figure A.1, I carry out the test proposed by Card, Heining, and Kline (2013). The figure shows that workers moving from low-paying establishments to high-paying establishments have a wage increase and that workers moving from high-paying establishments to low-paying establishments have a symmetric wage decrease. If variation in wages across establishments were mainly due to sorting, we would expect wage increases in the latter case.[18]

The DiD also addresses the third challenge—separately identifying the effects of market power from the effects of synergy. In my setting, the merger increased concentration in treated counties, in which firms overlapped, but it did not change concentration in control counties, in which firms did not overlap. Considering that the merger was decided at the national level, increases in concentration due to the merger are arguably exogenous at the county level.[19] The key identifying assumption is that synergies are realized at the national level and are the same for treated and control counties. Examples of synergies at the national level are better bargaining with suppliers, improvements in logistics and optimization of distribution centers, and combining headquarters and administrative departments. Under this assumption, the difference between treated and control groups yields the market power effect, excluding the effects from synergies.

Other papers have used a stronger assumption to address the same challenge. To identify the market power effect, Arnold (2021) assumes that the effect of the mergers on productivity (synergies) is independent of the change in local labor market concentration. For example,

[18] Gerard et al. (2021) use the same matched employer-employee dataset and present additional tests that suggest the assumption holds.

[19] Dafny et al. (2012) use a similar assumption to analyze the effects of the Aetna-Prudential merger in the health insurance market.

mergers with high change in labor market concentration will have the same synergies as mergers with low change in labor market concentration. The difference with my setting is that I use variation within a merger and Arnold compares the effects of different mergers. Similarly, Prager and Schmitt (2021) compare the effects of hospital mergers in which there is an increase in labor market concentration with other hospital mergers in which there is no increase in labor market concentration.

5. Effect of the Merger in Merging Firms

5.1 Effect of the merger on overall wages

Figure 1.3 presents trends in average log wages for all workers in merging firms and yields three takeaways. First, wages increase after the merger for both treated and control groups. We cannot attribute this increase to synergies from the merger, since other events during that period might have also affected wages; minimum wages and average wages for all workers in Brazil were also increasing in that period. Second, the pre-trends are parallel (see Figure 1.4). This gives supportive evidence on the parallel trends assumption and the assertion that establishments in treated and control groups are comparable. Third, wages in the treated and control groups seem to converge after the treatment, covering a large gap. Figure 1.4 shows a statistically significant effect at the 5% significance level.

Panel A of Figure 1.5 presents the difference in trends in the average log wages of pharmacists (see Appendix Figure A.2 for trends). The relative wages of pharmacists start with a slow decline after the merger and reach a strong decrease of 12% after 6 years, with an average effect of -7.3% (Appendix Table A.1). Columns 2 to 4 of Table A.1 use a continuous treatment measure—projected changes in HHI at the county level—instead of the binary treatment. The results show that firms spend less per pharmacist in counties in which concentration increases more. Note that these results do not reflect the causal effect of market power on wages since they do not account for changes labor force composition.

Results are robust to less restrictive samples and other measures of wages. By comparing column 1 to 4 with columns 5 to 8 of Appendix Table A.2, I show that the results for pharmacists do not change if we use average wages instead of wages from December.[20] The sample increases when using average wages because it includes individuals who worked in a merging firm for only part of the year and, for individuals who worked in two different establishments in the same year, both observations. Table A.3 relaxes some sample restrictions. I show that the effect on average wages does not significantly change when I (i) weight for the inverse of the number of pharmacists in each establishment (column 2), (ii) include workers with the top 1% of wages (column 3), (iii) include the year 2012 (column 4) or (iv) include all establishments from merging firms and not just the balanced panel of establishments (column 5).

Panel B of Figure 1.5 shows that the wages of salespeople do not exhibit the same decline as for pharmacists (see Appendix Figure A.2 for trends). The relative wages of salespeople have a small decrease after the merger (-6% in 2015) with a small increase after 2015. The average effect post-merger is not statistically different from zero and is small in magnitude. Once again, the results do not imply that concentration did not reduce the wages of salespeople since they do not account for changes labor force composition.

A recent study by Prager and Schmitt (2021) finds notably similar results. Using a different empirical strategy, the authors study many hospital mergers in the US. They show that after a high-concentration-inducing merger, the average wages of more specialized workers, such as pharmacists and nurses, decrease by 7%, while the average wages of unskilled workers remain stable. These results are consistent with the hypothesis that specialized workers have fewer outside options and are more exposed to the negative effects of concentration. The authors then conclude that mergers may affect wages through market power, but these

[20] Note that these regressions follow Equation 1.1 but are run at the individual level and use the drugstores sample (and not the fully connected sample).

effects only apply in relatively narrow circumstances and do not affect low-skill workers. As with the results in this section, the authors observe the average wages of a pool of workers in each establishment and do not take changes in composition into account. In the next section, I show that changes in labor force composition play a large role in the merger I study.

5.2 Effect of the merger on wages, accounting for changes in composition

Table 1.3 presents the main results of the paper. The table shows estimates of selected parameters from Equation 1.3, which are estimated over the fully connected set. Panel A shows that the results from the previous section on the wages of pharmacists were mostly driven by a change in worker composition within each establishment and not by a change in wages due to market power. The panel presents estimates of δ_3 for pharmacists in merging firms. First, without controls, the wages of pharmacists fall by -5.6% (column 1). The effect of the merger gets stronger when I include age controls (age squared and cubic), decreasing to -7.9% (column 2). The inclusion of establishment fixed effects does not significantly change the estimate (column 3). However, the inclusion of individual fixed effects reduces the magnitude of the effect, with the estimate decreasing to -2.6% (column 4). The latter result is not statistically different than zero at the 5% significance level with standard errors clustered at the county level. The estimate is still not statistically different than zero if I cluster the standard errors at the county-year level or at the establishment level.

I use leads and lags graph to show evidence that treated and control groups are also comparable when I include worker and establishment fixed effects. Figure 1.6 presents leads and lags estimates following the specification from column 4. The omitted category is the difference between treated and control groups in 2011. The figure shows that trends in residualized log wages are parallel before the treatment, which is evidence in favor of the parallel trends assumption. After the merger, there is a slow decline in market wages, but this effect is not statistically significant. Figure A.3 present the residualized trends for both treatment and control groups and shows the same facts. In summary, market wages of pharmacists do not fall once we take changes in composition into account. Notably, this contradicts the hypothesis that more specialized workers have fewer outside options and suffer greatly from events that increase concentration.

Panel B of Table 1.3 shows a decrease in the relative wages of salespeople, which also contrasts with the results from the previous section. First, note that the point estimate is negative in all specifications, but it is only statistically significant in column 4, which includes worker fixed effects. Second, the point estimate increases when I include age controls, which suggests that the average age of salespeople decreases in the treatment group relative to the control group. Third, column 4 shows that the relative wages of salespeople reduce by -3.7%, which is statistically significant at the 10% confidence level. When clustering at the county $\times$ year or establishment level, the effect is statistically significant at the 1% level.[21] The results do not support the assumption that salespeople are less affected by the merger.

One might be concerned that other events might be driving these results. For example, Table 1.2 shows that counties in the treated group are more populous than countries in the control group. The concern is that, for being different, these counties have different trends in the wages of pharmacists and salespeople. I address this concern by including interactions between year indicators and pre-merger county characteristics in column 5 of Table 1.3. County characteristics are indicators for each quartile of the following variables: 2010 population, 2003 HDI, 2010 informality levels, and 2006 county government revenue per capita. The comparison of estimates in columns 4 and 5 of Table 1.3 shows that the inclusion of these variables does not significantly change the magnitude of estimates. This provides additional evidence that results are robust and that control counties are comparable to treated ones.

The increase in minimum wages is not likely to drive these results. Although salespeople

[21]Standard errors are estimated using the Frisch–Waugh–Lovell method.

in the control group receive lower wages than salespeople in the treated group, more than 95% of the salespeople receive wages higher than the minimum wage. While there is evidence that minimum wages have spillover effects to the rest of the wage distribution, these effects should not differ across treatment and control groups.

In section 7., I discuss potential explanations for why the merger has a negative effect on the wages of salespeople and a small effect on the wages of pharmacists.

5.3 Effect of the merger on other labor market outcomes

Table 1.4 presents DiD results for other outcomes of merging firms. Column 1 shows employment effects. Establishments in treated counties have a statistically not different from zero reduction in the employment of pharmacists when compared with establishments in control counties. The magnitude of -0.18 (SE=0.24) represents a decrease of 6% in the employment of pharmacists. Given the large standard errors, I cannot rule out larger declines in employment. The employment of salespeople has an even smaller effect in magnitude which is also not statistically significant. Employment increases by 0.10 (SE=0.57), which represents an increase of only 1% in the number of salespeople. Again, given the large standard errors, I cannot rule out larger declines in employment.

Column 2 of Table 1.4 presents the effects on average age. Changes in labor force composition in terms of age explain the difference between the estimates in columns 1 and 2 of Table 1.3. The average age of pharmacists increases by 2.5%, which is around 9 months, relative to the control group. Although the point estimate is not statistically significant, its magnitude is sufficient to change the estimates in the wage equation of Table 1.3 by 2%. This occurs because older workers usually receive higher wages. Therefore, the estimate of -5.6% is masking a larger decrease in wages, compensated by a change in labor force composition. The inverse occurs for salespeople, where the wage effect of -2.4% partially captures the reduction in the average age of workers.

Column 3 of Table 1.4 presents the changes in labor force composition using worker fixed effects as an outcome. These changes explain the main result of the paper: The difference between estimates in columns 3 and 4 of Table 1.3. There is a relative decrease in worker fixed effects of 3.6% (Panel A). This explains why we see a large negative effect in the wages of pharmacists. What is actually happening is that establishments in the treated group are hiring pharmacists of lower fixed effects after the merger, relative to establishments in the control group. Panel B shows the opposite effect for salespeople (1.6% increase, SE=0.01). Establishments in the treated group are hiring salespeople of higher fixed effects after the merger, relative to the control group.

6. Effect of the Merger in Competing Firms

The effect of the merger on the wages of pharmacists working at competing firms is not statistically significant and varies with each estimation specification. Table 1.3 presents estimates of selected parameters from Equation 1.3 for competing firms. Column 1 shows that the average wages of pharmacists fall by 3.2%. However, when including age controls and worker and establishment fixed effects, we observe a statistically not significant increase in wages of 2.5%. Table 1.5 shows that this difference in estimates arises from a change in composition: Average age decreases by 1.7% and the average of pharmacists' fixed effects decreases by 4.8% (SE=0.012).

Competing firms reduce the relative wages of salespeople in treated counties. Panel B of Table 1.3 shows that the relative wages of salespeople fall by 3.3% after the merger. This effect is statistically significant and only varies by 1% with the removal of controls. Table 1.5 confirms that changes in age and worker fixed effects have low magnitude, such that they are not that relevant for the results. Note that the effect on the wages of salespeople in competing firms has magnitude similar to the effect in merging firms.

Column 1 of Table 1.5 presents the effects of the merger on employment in competing firms. There is a small increase in the number of pharmacists, but the effect is small, 0.187 (SE=0.098). Small effects on the employment of pharmacists might be explained by the production technology and the legal requirements drugstores entail: A pharmacist always has to be working when the facility is open. The effect on the employment of salespeople is larger: an increase of 1.08 workers, which is around 10% of the number of salespeople in competing firms.

The fall in wages and increase in the employment of salespeople in competing firms is consistent with the hypothesis that firms act as strategic substitutes in a monopsony model. As previously discussed, strategic substitutes respond to other firms' hiring strategies by doing the opposite of what their competitors do. In this case, competing firms responded to a decrease in merging firms' employment of salespeople by hiring more salespeople. This possibly prevented the wages of salespeople to fall even more. This new finding helps explain why the estimated effects are modest, despite the fact that concentration increases: The labor market effects of a merger are limited by competitors' responses.

7. Discussion and Magnitudes

In this section, I discuss two questions that emerged from the previous results: Why didn't pharmacists' wages in merging firms fall? And why did the wages of salespeople fall? In Section 5., I showed that the relative wages of pharmacists in merging firms fell by only 2.6% after the merger. On the other hand, the wages of salespeople fell by 3.5%. This evidence is contrary to the ex ante hypothesis that pharmacists would be more affected by the merger because they are more specialized and have fewer outside options than salespeople.

The small effect of the merger on pharmacists' wages is consistent with the higher unionization rate of pharmacists. Pharmacists are both more specialized and more organized than salespeople. Survey data show that 34% of pharmacists are unionized, while only 9% of salespeople working at pharmacies are (PNAD, 2011). One of the unions' functions is to negotiate wage floors and annual wage increases with employers. The results from this paper indicate that a merger between two firms might not weaken the bargaining power of employees' unions. This is consistent with the centralized nature of negotiations and with a model of collective bargaining.

I present additional evidence that supports the efficacy of pharmacists unions by separately analyzing the effect of the merger on incumbent and newly hired workers. Figure 1.7 presents the leads and lags results for pharmacists in these two samples: incumbent workers and newly hired workers. Panel A shows that the average wages of incumbent pharmacists do not change after the merger, compared with a control group. Yet, panel B shows that the wages of pharmacists hired in each year—i.e, entry-level wages for pharmacists—decrease by up to 20%.[22]

A possible explanation for these results is that incumbent workers are shielded by union agreements and new workers are not. First, firms have little room to differentially change incumbent workers' wages across counties. This explains the null results of panel A of Figure 1.7. Second, wage floors are the only instrument that unions have to protect newly hired workers. It turns out that wage floors are not binding for pharmacy chains, which usually pay higher wages than independent drugstores and than the wage floor. The results suggest that the merger changed firm's hiring strategy: merging firms decide to hire pharmacists that accept lower wages, which tend to be of lower worker fixed-effect.

The second surprising result is that the wages of salespeople decrease by 3.5%. Ex ante, we hypothesized that salespeople working in drugstores are fully mobile with respect to other industries, such as supermarkets, general stores, etc. In this scenario, a merger would not increase concentration in the salespeople labor market, and thus it should not affect the wages of salespeople. Accordingly, Panel B in Figure 1.2 shows that the merger does not change the HHI of salespeople when we define a labor market at the county×year×1-digit

[22] These regressions do not include individual characteristics and worker fixed effects, not accounting for changes in composition.

occupation.

However, salespeople working in drugstores might have smaller labor markets than previously thought. In Appendix Table A.4, I show that 74% of salespeople working in a drugstore continue as salespeople in the following year and that 73% still work in a drugstore in the following year. Appendix Table A.5 shows that 24% of the salespeople who were working in a drugstore and switched jobs were still working in a drugstore the following year. The share of salespeople moving to drugstores should be closer to 5% if they had switched randomly to other salespeople jobs. I interpret these results as evidence that salespeople working in drugstores have higher preference to remain in drugstores and do not consider all salespeople jobs equally as their labor market.

I show that pharmacists and salespeople may have preferences over the type of drugstore they wish to work at. In Appendix Figure A.4, I plot a nonparametric regression of the share of workers who move to a pharmacy chain over the share of establishments from pharmacy chains in each county as percentage of total drugstores. Pharmacy chains are defined as firms that have more than five establishments in one year in the entire country. I present results for two groups of workers based on their employer of origin: pharmacy chain or independent drugstore. Panel B of Appendix Figure A.4 shows that salespeople who were working in a pharmacy chain are more likely to switch jobs to another pharmacy chain than salespeople who where originally working in an independent drugstore. Panel A of Appendix Figure A.4 shows a similar pattern for pharmacists. These results provide additional evidence that workers' preferences and/or specific human capital are capable of defining upward-sloping labor supply curves that are specific to each firm—or at least for each type of firm, as in Card et al. (2018).

7.1 Magnitudes

So far, we have seen that the merger increased concentration and decreased wages in some markets. Yet how much did wages decrease after a certain increase in concentration?

Elasticities vary significantly with labor market definitions. To show that, I use the DiD wage estimates for pharmacists and salespeople in merging firms (-2.6% and -3.5%, respectively) and different measures of increases in concentration. Appendix Table A.6 presents measures of concentration using the HHI. In the table, I report the average HHI and average merger-induced change in HHI across counties in the treated group, in which the merging firms overlapped and concentration increased. The merger-induced change in HHI in counties in the control group is always zero by definition.

Wage elasticities with respect to concentration increases are small in the case of pharmacists. The HHI in 2011 for the pharmacists' labor market varies between 457 and 2,316. These measures depend on the labor market definition: all establishments that hired a pharmacist versus only pharmacy chains. Similarly, merger-induced changes in HHI vary from 37 to 651 points. An average increase of 37 points is considered small for DOJ and FTC merger guidelines. However, an increase of 651 points is large enough to receive antitrust attention. The elasticity of wages with respect to labor market concentration ranges from -0.34 to -0.1 for pharmacists.

Panel B of Appendix Table A.6 presents measures of concentration for salespeople. First, using a broad labor market definition, we get that concentration increases by 0.1 points, or 0.2%. This generates the implausibly large elasticity of wages with respect to concentration of -17.5. When restricting the labor market to salespeople in pharmacy chains, the elasticity is -0.13. Hence, the results show that finer labor market definitions give a better description of the salespeople's labor market. The development of a method to define labor markets is left for future research.

8. Conclusion

This paper studies the effects of market power on wages by analysing a merger between two large retail pharmacies in Brazil. I find that the wages of pharmacists do not decrease after the merger and that the wages of salespeople do. The results for pharmacists are consistent with collective bargaining models. The results for salespeople are consistent with modern versions of the monopsony model.

The literature on the labor market effects of mergers is still nascent compared with the industrial organization (IO) literature on the product market effects of mergers. Two features of this paper may sound obvious to IO economists but are still incipient in labor economics. First, the analysis of a single merger in a specific sector provides a narrative that might be useful to other mergers and sectors. In contrast, the aggregate analysis of multiple mergers may underscore important differences across industries and markets. Second, by including individual controls such as age and worker fixed effects, this paper shows the importance of standardizing a unit of labor when following establishments over time.

I present evidence that labor markets can be finer than the year-county intersection. Researchers examining labor markets typically use the intersection between year and regional units such as states, metropolitan areas, commuting zones, or counties to approximate labor markets. In many cases, these choices are guided by data availability. The results in this paper give empirical support to recent work by Manning and Petrongolo (2017); Nimczik (2020); and Schubert et al. (2021), who use data on vacancies and worker flows to show that labor markets are more local than previously thought. In my setting, the merger should not have reduced wages if the labor market had been defined at the county level. However, the merger reduced the wages of salespeople, which suggests that labor markets may be defined by the intersection of months, occupation, industry, county, and firm characteristic, such as firm size.

This paper also investigates how firms strategically interact in the labor market. I show that, following a wage reduction for salespeople, competing firms increased their labor demand, possibly preventing wages from falling even more. That said, how firms compete in other labor markets is an interesting avenue for future research.

Finally, this paper studied the effects of a merger through the mechanism of market power, excluding synergies. Regulatory agencies are interested in the total effects of merger on wages, including synergies. The empirical strategy in this paper does not address these outcomes, since synergies and other shocks to the labor market are not separately identified. As far as I know, the merger in this paper could have had strong synergies that actually increased wages.

9. Figures and Tables

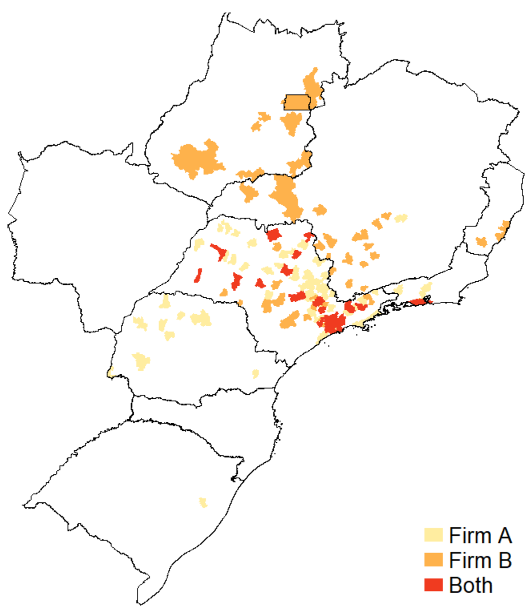
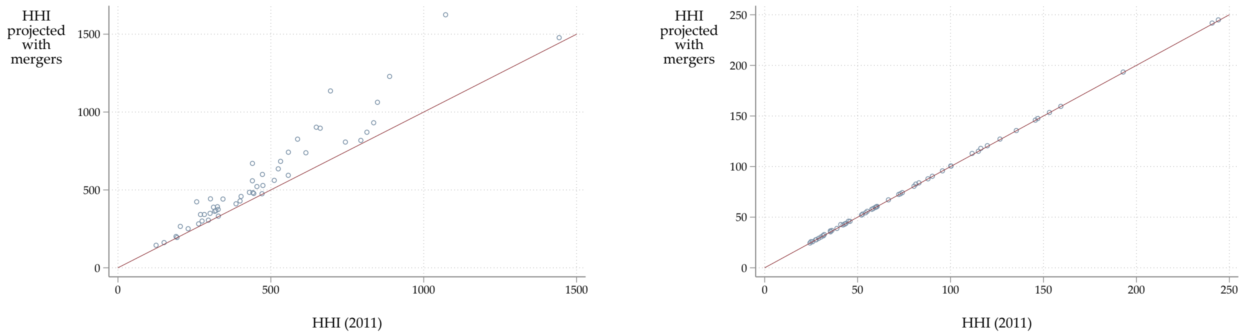


Figure 1.1: Presence of merging firms before the merger (2010)

Note: The figure presents a map of the south of Brazil and highlights counties in which only one of the merging firms had an establishment (69 in light yellow and 54 in orange) and counties in which both firms had establishments prior to the merger (31 in red). The area in white denotes counties in which firms were not present in 2010.



(a) Pharmacists

(b) Salespeople

Figure 1.2: HHI projected with mergers versus 2011 HHI

Note: Each circle in the figure represents a county in which merging firms overlapped. The horizontal axis displays the 2011 Herfindahl–Hirschman Index (HHI) in the labor market of pharmacists (panel A) and salespeople (panel B). The HHI is computed using the shares of all firms within a county that hired workers in these occupations. The vertical axis displays the 2011 HHI projected with the merger. The projection serves to calculate a change in HHI that is due to the merger and not to the entrance and exit of firms or changes in firms' employment shares.

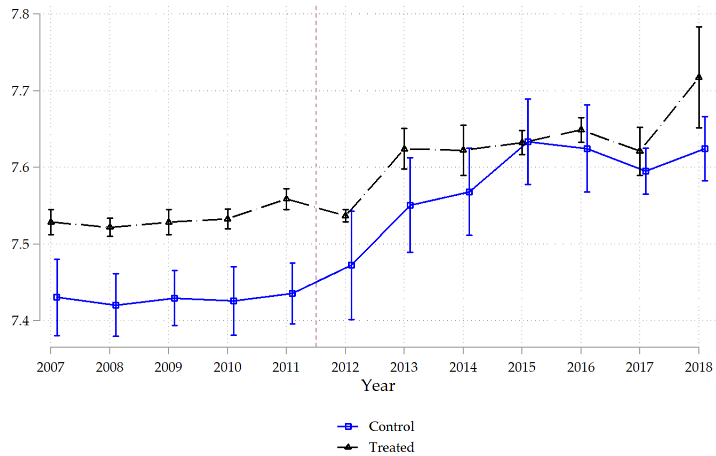


Figure 1.3: Trends - Ln(wage) of all workers in merging firms

Note: The figure presents the average ln(wage) in treated and control groups. Estimates come from a regression that includes all individuals working in merging firms, from 2007 to 2018. Regressions do not include any additional controls. Log wages are measured in December. The sample only includes workers employed on December 31. The sample is composed of a balanced panel of establishments. Standard errors are clustered at the county level.

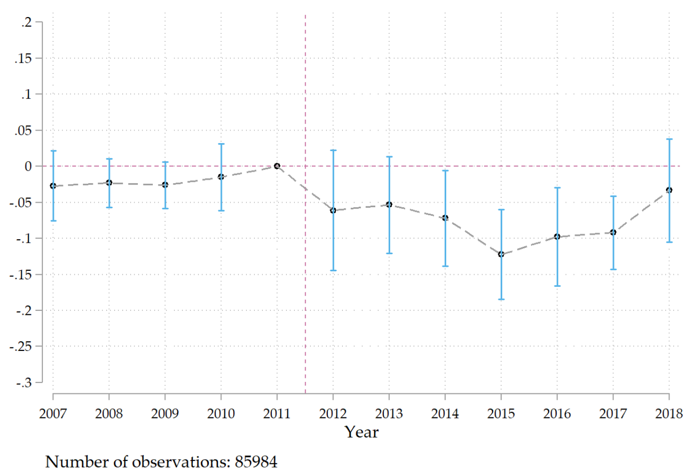
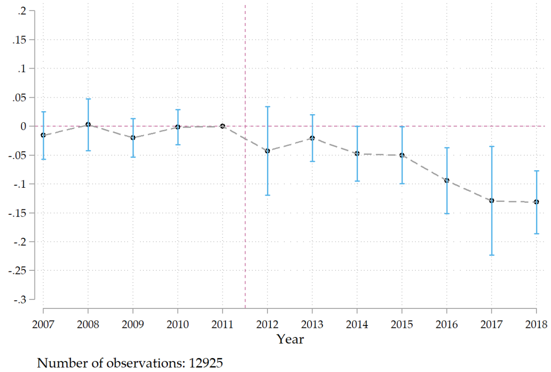
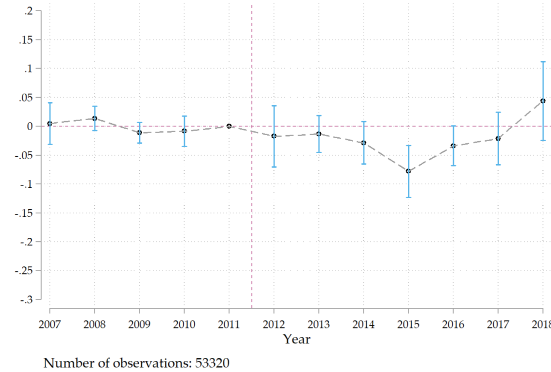


Figure 1.4: Leads and lags - Ln(wage) of all workers in merging firms

Note: The figure presents estimates of the parameters in Equation 1.2. The omitted category is the difference between treated and control groups in 2011. The regression is run at the individual level and does not include any additional controls. Log wages are measured in December. The sample only includes workers employed on December 31. The sample is composed of a balanced panel of establishments. Standard errors are clustered at the county level.



(a) Pharmacists



(b) Salespeople

Figure 1.5: Leads and lags - Ln(wage) of workers in merging firms

Note: The figure presents estimates of the parameters in Equation 1.2. The omitted category is the difference between treated and control groups in 2011. The regression is run at the individual level and does not include any additional controls. Log wages are measured in December. The sample only includes pharmacists (Panel A) or salespeople (Panel B) employed on December 31. The sample is composed of a balanced panel of establishments. Standard errors are clustered at the county level.

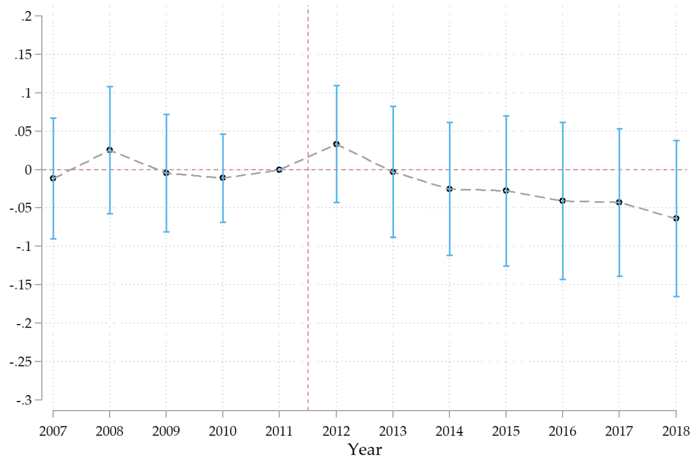
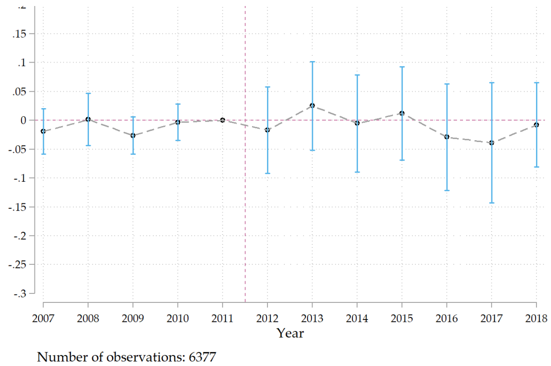
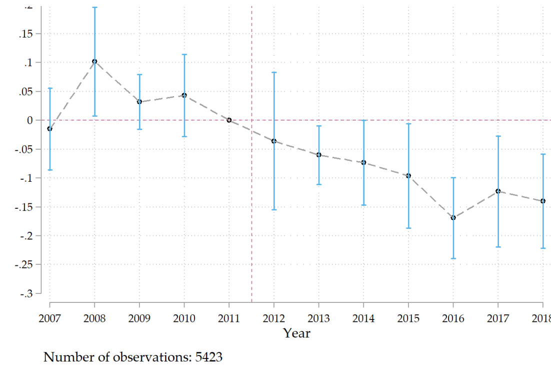


Figure 1.6: Leads and lags - Ln(wage) of pharmacists in merging firms with two-way fixed effects

Note: The figure presents leads and lags estimates. The omitted category is the difference between treated and control groups in 2011. Estimates come from a regression that uses the fully connected set and includes individual characteristics and worker and establishment fixed effects, similar to Equation 1.3. Log wages are measured in December. The sample only includes workers employed on December 31. Standard errors are clustered at the county level.



(a) Incumbent workers



(b) Entry-level wage

Figure 1.7: Leads and lags - Ln(wage) of pharmacists in merging firms

Note: The figure presents estimates of the parameters in Equation 1.2. The omitted category is the difference between treated and control groups in 2011. The regression is run at the individual level and does not include any additional controls. Log wages are measured in December. The sample only includes pharmacists who are incumbent (Panel A) or newly hired (Panel B) and are employed on December 31. Incumbent workers are individuals who were hired before the merger. Entry-level, or newly hired, workers are individuals who were hired in the current year. The sample is composed of a balanced panel of establishments. Standard errors are clustered at the county level.

Table 1.1: Sample size

	Drugstores Sample		Fully Connected
Number of...	Merging firms	Competitors	Sample
Counties	91	91	2,812
Firms	2	9,798	3,328,498
Establishments	392	11,212	4,273,843
Workers	51,380	497,658	29,260,227
Observations	134,216	1,586,235	189,072,473

Note: "Drugstores sample" refers to the balanced panel of establishments in the retail pharmacy sector, where establishments belong to a county in which merging firms were present in 2010. The fully connected sample includes establishments from all sectors that are connected through worker mobility to establishments in the drugstore sample. RAIS 2007—2018.

Table 1.2: Summary Statistics of treated and control groups

	Control		Treated	
	Mean	Obs.	Mean	Obs.
Panel A: County level				
Population (2010) (thousands)	232.6	61	914.8	30
Government revenue (2006) (R$ pc)	2,443.6	61	2,925.4	30
HDI (2000)	0.81	61	0.83	30
% Informality (2010)	28.1	61	25.7	30
Pharmacists HHI (2011)	669	61	478	30
Panel B: Individual Level (2011)				
% Male	32.8	2374	30.5	5660
Age	23.55	2374	24.43	5660
Ln(wage) of all workers	7.44	2374	7.57	5665
Ln(wage) of pharmacists	8.43	286	8.47	738
Ln(wage) of salespeople	7.21	1608	7.33	3519

Note: Population data are from the 2010 census. Government revenue data are from the year 2006 and were accessed through Ipeadata. HDI refers to the human development index of 2000, also acessed through Ipeadata. Informality is accessed from the 2010 Census and defined as the share of workers who do not contribute to the social security system. Pharmacists' HHI refers to the Herfindahl–Hirschman Index for the labor market of pharmacists, where a market is defined by the county and occupation intersection. HHI is computed using 2011 RAIS data. All individual-level characteristics are computed for individuals working in a balanced panel of establishments of merging firms in 2011 using RAIS data.

Table 1.3: Difference-in-differences estimates. Dependent variable is Log(wage)

	(1)	(2)	(3)	(4)	(5)
Panel A: Pharmacists in Merging Firms					
Treat × Post	-0.056	-0.079	-0.069	-0.026	-0.028
	(0.032)	(0.034)	(0.032)	(0.035)	(0.035)
Panel B: Salespeople in Merging Firms					
Treat × Post	-0.024	-0.007	-0.021	-0.037	-0.035
	(0.018)	(0.025)	(0.017)	(0.019)	(0.018)
Panel C: Pharmacists in Competing Firms					
Treat × Post	-0.032	-0.026	-0.016	0.027	0.025
	(0.016)	(0.019)	(0.018)	(0.017)	(0.018)
Panel D: Salespeople in Competing Firms					
Treat × Post	-0.046	-0.042	-0.035	-0.033	-0.032
	(0.019)	(0.022)	(0.025)	(0.014)	(0.015)
Age controls		X	X	X	X
Individual FE				X	X
Establishment FE			X	X	X
Year FE	X	X	X	X	
Year FE × County chars					X

Note: Selected coefficients from the difference-in-differences estimate as in Equation 1.3, where the unit of observation is at the individual level. Each column represents a regression with 189,072,473 observations from 29,260,227 workers and 4,273,843 establishments. Log wages are measured in December and the sample only includes workers employed on December 31. Standard errors are clustered at the county level and presented in parentheses.

Table 1.4: Difference-in-differences estimates. Other outcomes

Outcome:	Employment	Ln(Age)	Worker FE
	(1)	(2)	(3)
Panel A: Pharmacists in Merging firms			
Treat × Post	-0.179	0.025	-0.036
	(0.238)	(0.015)	(0.024)
Observations	4,690	12,449	12,157
Number of workers	-	5,158	4,965
Number of establishments	392	392	392
Panel B: Salespeople in Merging firms			
Treat × Post	0.103	-0.012	0.016
	(0.566)	(0.014)	(0.011)
Observations	4,690	50,869	48,293
Number of workers	-	28,826	26,582
Number of establishments	392	392	392
Year FE	X	X	X

Note: Each cell represents the estimate of the difference-in-differences parameter from Equation 1.1. Regressions in column 1 are at the establishment level and regressions in columns 2 and 3 are at the individual level. Regressions include year and county fixed effects. Standard errors are clustered at the county level and presented in parentheses.

Table 1.5: Difference-in-differences estimates. Other outcomes - Competitors

Outcome:	Employment	Ln(Age)	Worker FE
	(1)	(2)	(3)
Panel C: Pharmacists in Competitors			
Treat × Post	0.187	-0.017	-0.048
	(0.098)	(0.006)	(0.012)
Observations	127,710	82,421	72,450
Number of workers	-	26,443	23,107
Number of establishments	10,839	6,090	5,800
Panel D: Salespeople in Competitors			
Treat × Post	1.078	-0.015	-0.016
	(0.522)	(0.005)	(0.016)
Observations	127,710	500,865	466,628
Number of workers	-	202,291	181,171
Number of establishments	10,839	10,173	9,920
Year FE	X	X	X

Note: Each cell represents the estimate of the difference-in-differences parameter from Equation 1.1. Regressions in column 1 are at the establishment level and regressions in columns 2 and 3 are at the individual level. Regressions include year and county fixed effects. Standard errors are clustered at the county level and presented in parentheses.

Chapter 2

Defining Labor Markets from Worker Flows

1. Introduction

Labor market definitions are important to evaluate and predict the effects of government policies, migration, trade shocks, mergers and acquisitions and other events on workers' outcomes. Researchers have typically relied on ad-hoc approximations for labor markets such as geographic boundaries, occupational codes, or industry codes at differing levels of granularity. However, markets may vary in size and may cross these boundaries. For any given job, it is not clear which definition is most appropriate.

In this paper, we use job transitions across county by industry by occupation cells to draw labor market boundaries in Brazil. We use a method from the network literature—the constant Potts model (CPM)—that aggregates cells (hereafter, nodes) into communities such that most job transitions occur within communities. We refer to communities as data-driven labor markets. We acknowledge that the jobs available to two distinct workers will not have a perfect overlap, such that one should interpret the data-driven labor markets as approximations to labor markets and an improvement from ad-hoc definitions.

We use the Brazilian employer-employee matched dataset (RAIS) between 2007 and 2013 to compile job transitions across nodes. The Brazilian data has many advantages. First, RAIS is a large dataset with detailed geographic, industry, and occupation categories. This allows us to define nodes at the finest level and still have a large number of job transitions across nodes. Second, Brazil is a large country—comparable in size to the US—allowing us to study geographic dispersion. Brazil is also a relatively closed economy, where most workers never worked in another country. It is reasonable to assume that labor markets will not cross international borders, and hence, will be captured with the RAIS data. A set back from RAIS is that it does not include workers in informal jobs.

The method is successful in its purpose of creating communities that are densely linked. By using 40 million job transitions across 246 thousand nodes, the algorithm finds 54,855 data-driven labor markets. We use the Leiden algorithm (Traag et al., 2019) to implement the CPM method. This algorithm corrects a flaw in another widely used algorithm in the networks literature that resulted in communities that were disconnected. In addition, the CPM method allows us to identify small communities. The data driven markets have 75% of job transitions occurring within markets and self-contain four times more job transitions than ad-hoc market definitions.

We outline four stylized facts from the data-driven markets. First, there is substantial heterogeneity in market size and market concentration. The largest 50 markets contain around 55% of job transitions and have more than 100 nodes each. On the other hand, 40 thousand markets have a single node. Most markets have several firms and are not concentrated. We show that 86.7% of workers participate in labor markets that are not concentrated. This should not cast a shade on the fact that more than 10% of workers participate in highly concentrated markets. Data-driven labor markets are usually less concentrated than finer ad-hoc labor market (like county by 6-digit occupation) and more concentrated than broad ad-hoc market (like commuting zone by 3-digit industry).

Second, we show that firms hire workers from several different markets. While 76.4% of firms hire workers from a single market, the 13,308 largest firms—which employ 75% of formal workers in Brazil—hire workers from an average of 10.9 markets and a median of 7 markets. Firms participate in different markets either because they have establishments in many locations or because they hire workers from different occupations. Our results are complementary to the approach by Nimczik (2020). The author uses a similar method to define data-driven labor markets in Austria where nodes are constituted by single firms. As a consequence, in his paper firms cannot participate in different markets. We show that, because markets are spread across occupations, firms may participate in many markets.

Third, we characterize markets by their geographical dispersion. We show that workers participate in markets that are spread across the country (4%), across neighboring states (2%), and across microregions (10%). While 32% of workers participate in markets that include many counties within a microregion, for some workers this definition is too broad. In fact, 27% of workers participate in very local markets that do not cross county borders.

For last, we show that most markets include many occupations and industries. For example, 88% of markets have nodes in at least two different 3-digit occupations. In general, data-driven labor markets pass the common-sense test. For example, the relevant market for airline pilots is a detailed occupation-industry cell at the national level. On the other hand, the relevant market for salespeople consists of many occupations and industries but are restricted to a narrow geographic area.

This paper contributes to the growing literature on labor market definitions. Schmutte (2014) and Nimczik (2020) use different methods from the network literature to identify markets from worker flows (Modularity maximization and Stochastic Block Model). Both studies find that most job transitions are concentrated in very few markets. However, this could be a consequence of the lack detailed industry and occupation data. We use county by 6-digit occupation by 5-digit industry nodes, which allows us to better characterize job transitions. In addition, the CPM method and the algorithm used in this paper have the advantage of being able to identify markets with few nodes. Manning and Petrongolo (2017) estimate a spatial job search model and find that workers have strong preferences for jobs that are in close proximity. We show that while this can be true for most workers, individuals in some occupations and industries have job transition that cross long distances. Schubert et al. (2021) take a different approach by creating a measure of outside-occupation options.

Our data-driven labor markets can be useful to guide policy. In particular, antitrust agencies should be careful when evaluating the potential effects of mergers and acquisitions. A result from this paper is that a merger between two large firms may affect workers in more than a single market and the effects could differ depending on each market's concentration. Prager and Schmitt (2021) and Guanziroli (2022) do a retrospective merger analysis and show that the effects of the merger vary by occupation, consistent with our results.

2. Literature has different Local Labor Market Definitions

In this section, we analyze how recent studies from three different topics define labor markets. For the papers that do not explicitly define labor markets, we use the unit of observation from the main results of the paper.

Appendix Table B.1 presents nine articles that study the effects of increasing labor market concentration on worker's wages and employment. Market definitions are especially important in this literature, as they matter for calculating the level of concentration. In all studies, market definitions seem to be dependent of data availability. Studies that use employer-employee data from the United States tend to define labor markets with the commuting zone X industry cell, but they disagree on the level of aggregation (Benmelech et al., 2022;Arnold, 2021; Berger et al., 2022). Markets are not defined using occupation codes due to the LEHD not having occupation information. That said, Azar et al. (2020) and Azar et al. (2022) gather occupation information from online job postings and argue that commuting zone X occupation is a better labor market definition. Brooks et al. (2021) and Marinescu et al. (2021) use employer employee data from India and France, respectively, and also define labor markets based on data availability (district X industry and Commuting zone X occupation). Prager and Schmitt (2021) and Guanziroli (2022) analyze the effects of mergers within an industry. Using detailed data, they are able to define markets at the location X industry X occupation level.

Appendix Table B.2 shows that the minimum wage literature also has a disagreement regarding market definitions, which is probably due to different data availability. Under the monopsony theory, increases in minimum wage should not affect employment in highly concentrated markets. Hence, market definitions are relevant to these studies. That said, most studies in this literature implicitly use aggregate labor market definitions, like state X industry. As a consequence, the estimated minimum wage effects might be aggregating different heterogenous effects. In addition, there could be contamination between treated and control groups in some studies as markets often cross state boundaries.

Market definitions also vary across papers that study the effects of trade shocks on labor market outcomes. Recent studies show that trade shocks, like trade liberalization, affect some sectors of the economy more than others, a fact that has been explained by the presence of mobility frictions. Market definitions are important to these contexts since the effects of trade shocks are probably heterogenous by labor market. Appendix Table B.3 shows that these definitions also vary across studies.

3. The Brazilian Matched Employer-Employee Dataset

To identify worker flows we use data from RAIS (Relação Annual de Informações Sociais), the Brazilian employer-employee linked administrative dataset.

RAIS is a confidential dataset maintained by the Brazilian Ministry of Labor. We use the data from 2007 to 2014. The dataset includes comprehensive information of firms, establishments, workers, and of the job match. The data is disclosed annually, and firms have to report all their job links within a year. Workers may have more than one entry in a single year, as they switch firms. In this paper, we will use information on occupation, industry sector, and county of employment. To identify worker flows we also use workers' and firms' identification numbers.

The method used in this paper requires the collection of job transition data. The RAIS dataset is well-suited for this purpose because it can provide a large number of transitions between very fine cells, or hereafter, nodes. In this paper, **a node is a county X 5-digit industry X 6-digit occupation cell**. We observe workers switching counties, industry, and occupations when they switch jobs. Thus, we can compute the number of switches for every pair of nodes. The link between two nodes is called an **edge**. Note that an edge may be constituted of one or more **job transitions** and the number of job transitions is referred to as the **weight** of an edge. In the data, job transitions come from: (i) transitions across consecutive years, and (ii) transitions within the year. We compute job transitions for the years between 2007 and 2013.

We impose the following restrictions to the data: First, we exclude workers from the public sector. The reason is that even though workers from the public sector work in very distinct jobs they are sorted into a single industry category. Although governments are an important participant of the labor market, their inclusion would probably agglomerate very distinct nodes, leading to misclassification. Secondly, we exclude temporary workers, internships, and contracts with less than 30 monthly hours of work. These workers are usually not relevant to firms' operation, but given their nature, they might be overrepresented in terms of job transition. Lastly, we exclude workers with more than two observations within a year, as these might reflect reporting mistakes and their inclusion could create false transitions between disconnected nodes.

Column 1 of Table 2.1 presents sample sizes of the resulting raw data. Between 2007 and 2013, we identified workers in 5,360,609 nodes and their transitions between nodes led to 45,270,686 edges. However, most nodes are too small, and most edges contain a single transition. Hence, to reduce noise and prevent the use of misclassified transitions, we impose additional restrictions. We exclude nodes—and all their edges—with weight lower than five. I.e., we exclude nodes in which less than five workers switched jobs between 2007 and 2013 to another firm in the same node. We also exclude edges with only one transition. Column 2 of Table 2.1 presents sample sizes of the main data used in the paper. We are left with 246,638 nodes and 6,819,564 edges. There is an average of 178 transitions per node, but many transitions occur within nodes (35% of them.). I.e., workers tend to switch jobs to firms in the same county, same occupation, and same industry.

In the next section, I discuss a method that aggregates nodes into markets. Appendix Table B.4 shows that, after data cleaning and restrictions, we are left with 4,246 counties, 2,026 6-digit occupations and 646 5-digit industry sectors. Their combination leads to 246,638 nodes. Studies using data with less precise categories or missing industry, occupational or regional data will have fewer nodes. Depending on the structure of job transitions, these could lead to few markets being detected or to wrongful detection.

4. A Method to Define Labor Market Approximations

This section describes the method to identify communities, which are approximations to labor markets. The Labor market is a complex network of workers and firms. While an individual entering the market can in theory apply for any available job, workers and firms tend to cluster at different levels of occupation, industry sector, and location. The goal of this paper is to identify such clusters, or communities.

To identify communities, we use a method called constant Potts model (CPM). The CPM method is a reiterative process that assigns nodes to communities until there are strong connections between the nodes within a community and weak connections between nodes across communities.

The notation is the following: the connected graph G includes n nodes and m edges. Nodes take the atomistic form and may also be referred to as vertex. Edges are the structure within a graph that attach two nodes. The adjacency matrix $A_(n \times n)$ maps the edges between nodes, where $A_{ij} = 1$ if there is an edge between nodes i and j. Edges have weight w_{ij}. The community of node i is denoted by σ_i and the function $\delta(\sigma_i, \sigma_j)$ takes value one if $\sigma_i = \sigma_j$, i.e., nodes i and j belong to the same community, and zero otherwise.

The constant Potts model maximizes the expression below by choosing a value σ_i for all i.

$$H = \sum_{ij}(A_{ij}w_{ij} - \gamma)\delta(\sigma_i, \sigma_j) \tag{2.1}$$

Where γ is a constant, also called the resolution parameter. Traag et al (2011) show that γ balances the trade-off between maximizing the number of internal edges within a community and keeping communities relatively small. The constant can also be understood as a penalty for the inclusion of a new edge in the community. As a result of this maximization, communities should have strong links within them and weak links across them.

An alternative method previously used in the literature is the modularity maximization (Schmutte, 2014;Nimczik, 2020). Similar to CPM, the modularity maximization approach yields communities with strongly connected nodes. However, Fortunato and Barthélemy (2007) show that modularity maximization has a problem named the resolution limit problem that prevents the detection of smaller communities. The CPM method is resolution-limit-free (Traag et al., 2011).

4.1 Algorithm and implementation

In this paper, nodes are defined by the intersection between 6-digit occupations, 5-digit industries and counties. This is the finest cell in the data. Two nodes are connected by an edge if at least two worker switched jobs between these nodes. The edge weight is the total number of job transitions between the nodes. Hence, the CPM method aggregates occupations, industries, and counties with dense worker transition into a community, or a data-driven labor market.

The CPM method is implemented through Python using the Leiden algorithm (Traag et al., 2019). The Leiden algorithm corrects a flaw in the widely used Louvain algorithm which generates badly connected communities. We set the resolution parameter in the CPM method to $\gamma = 0.1$.

5. Characterizing Labor Markets

The Leiden algorithm identifies 54,855 communities, or approximations to labor markets. Column 1 of Table 2.2 presents descriptive statistics of the data-driven markets. The table shows that 75% of job transitions occur within data-driven markets, with the rest occurring across markets. This is an indicator of the algorithm's success.

The algorithm aggregates nodes into markets to maximize the density of a market and minimize cross market density. As a comparison, Column 2 of Table 2.2 presents descriptive statistics of a commonly used ad-hoc labor market definition—microregion X 6-digit occupation cells— that gives a similar number of markets to the data-driven method. The table shows that 45.8% of job transitions occurred within the microregion X 6-digit occupation cells. To highlight the contrast, I exclude transitions within the same node: 61.8% of job transitions occurred within data-driven markets, while only 16.8% occurred within microregion X 6-digit occupation cells.

Next, I discuss four stylized facts from the analysis of data-driven labor markets.

(i) **There is substantial heterogeneity in market size and market concentration**

Table 2.2 reveals considerable heterogeneity in market size. Markets have on average 4.5 nodes and 800 job transitions. But nodes are not equally distributed across markets. While 40,171 markets have a single node, nine markets have more than 1000 nodes. A significant number of markets is in between, with 14,675 markets having between 2 and 1000 nodes.

This translates into two facts: (a) A great share of workers and transitions are part of few large labor markets, and (b) A considerable share of workers participates in small and concentrated labor markets. The top 50 markets with more transitions have 55% of the transitions in the data and a proportional fraction of workers. That said, 950 markets have 33% of transitions and 53,000 markets have 11.6% of transitions. These numbers contrast with Schmutte (2014) and Nimczik (2020) that find that most transitions and workers belong to very few markets. This discrepancy could be either to differences in context or in the choice of algorithm and data availability.

Panel A of Table 2.3 shows that some markets are highly concentrated, and some are not. Labor market concentration is measured with the Herfindahl–Hirschman Index (HHI). The HHI for a single market is defined as the sum of the square of firm's employment share in that market. The scale ranges between 0 and 10,000, with 0 representing perfectly competitive markets and 10,000 representing a monopoly. The average HHI across all markets is of 6,450.7 points, which is substantially high. To put into context, the U.S Department of Justice (DOJ) considers that product markets with HHI between 1,000 and 1,800 are moderately concentrated and markets with HHI above 1,800 are highly concentrated.

However, when considering the size of each market, we learn that most workers participate into markets that are not concentrated. The previous numbers are led by markets of a single node and that have very few workers. The average HHI weighted by employment in 2014 is of 614.6. Appendix Table B.5 shows that 86.7% of workers participate in labor markets with HHI lower than 1000 points, which are not concentrated according to the DOJ's guidelines.

The use of ad-hoc proxies for labor markets, such as the microregion X 6-digit occupation definition leads to the misleading conclusion that markets are highly concentrated. Table 3 shows that the average HHI weighted by employment is of 1,341 points. In addition, the metrics suggests that more than 25% of workers participate in moderate to high concentrated markets. However, ad-hoc proxies fail to aggregate workers and firms by how interconnected they are. For example, this measure allocates airplane pilots into different regional markets, each with very few firms. This would lead to a highly concentrated market even though there are many firms at the national level. The next stylized facts show that markets often cross regional boundaries.

(ii) **Firms hire workers from many markets**

Most firms hire workers from a single market, but the largest firms hire workers from many labor markets. Panel B of Table 2.3 shows the distribution of firms by the number of markets in which they hire workers. In 2014, we identified 1.4 million firms, and more than 75% of them hire workers from a single market. However, the 13,308 largest firms (by labor force size) hire workers from an average of 10.9 markets. The largest 738 firms in the country hire workers from an average of 40.9 markets.

Labor market definitions are important when considering the effects of mergers and acquisitions. A merger between two firms increases concentration in the labor market since there is one less firm in the market. Measurement of increases in concentration depend on the market definition. A merger in a market with 1,000 equally sized firms will not significantly increase concentration, but a merger in a market with 10 firms will. The increase in concentration (measured by the HHI) is one of the main metrics used by antitrust agencies around the world when analyzing product markets. Our approach shows that firms hire workers from many markets. As a consequence, they may increase concentration in some of these markets, but not necessarily in all markets.

Prager and Schmitt (2021) and Guanziroli (2022) show the importance of labor market definitions when analyzing the effect of a merger in the hospital and retail pharmacy sectors, respectively. In both their analysis, firms are assumed to hire workers from different markets, as defined by their occupations. The authors show that this margin is important, with the effects of mergers being different across occupations. Our approach provides a guide to future researchers and antitrust agencies on how to determine the markets in which firms participate.

(iii) **Some markets are geographically dispersed**

One of the main problems of using ad-hoc labor market proxies such as occupation X microregion cells is that some types of workers move across regions. In our approach, we find that some workers participate in markets that are spread across the country, across neighboring states, across microregions and within a state. On the other hand, some markets are smaller than a commuting zone, or a microregion. Next, we discuss each case.

a. Nationally

Figures 2.1.A and 2.1.B present two examples in which the data-driven markets are spread across the country. Figure 2.1.A shows market number 91, which is composed mostly of workers in the civil engineering occupation. In the figure, each circle represents a county that belongs to the market. The size of the circle represents the sum of transitions of all nodes that belong to market 91 in that county. Market 91 is composed of 233 nodes that include 56 counties, 21 occupations and 29 industries. Civil engineers cover 59% of nodes and the other occupations are closely related even though some are categorized in different 1-digit occupation. The market is defined over 53,832 transitions.

Nationally dispersed labor markets are not necessarily associated to workers with a college degree. Figure 2.1.B shows market number 80, which is composed mostly of workers in the welder 6-digit occupation. Market 80 is composed of 252 nodes that include 82 counties, 13 occupations and 51 industries. The welder occupation is in 84.1% of nodes. The market is defined over 64,959 transitions.

Table 2.4 shows that many workers participate in nationally dispersed markets.[1] There are 432 markets classified as nationally dispersed and they contain 4.8% of job transitions. Interestingly, nationally dispersion is not a characteristic of all large markets. From the 1000 largest markets in terms of transitions, only 39 are nationally dispersed.

b. Across Neighboring States

Some markets cross state borders but are not nationally dispersed. Figure 2.2 plots the map of markets 325 and 977. Market 325 has soybean farming as the modal sector, containing 68% of job transitions. The modal occupation is tractor driver, containing only 15% of transitions. Most of the other occupations are in the context of operating agricultural machines or agriculture work. Panel A of Figure 2.2 shows that most nodes of market 325 are contiguous and concentrated in the states of Maranhão and Piauí.

Panel B of Figure 2.2 shows market 897, which is contained in the states of São Paulo and Rio de Janeiro, with most transitions occurring in the capitals of these states. The market includes 20 occupations and 9 industries, with elevator maintenance having only 20.5% of transitions. Other occupations are within the scope of electromechanics.

Table 2.4 shows that there are 401 markets that cross neighboring states, containing only 2.1% of job transitions.

[1] I define nationally dispersed labor markets the markets that have nodes in more than one region and in more than two states and satisfy at least one of the following conditions: (1) more than 15% of transitions in at least two different regions; (2) more than 10% of transitions in at least three regions; (3) more than 5% of transitions in at least four regions; (4) more than 2.5% of transitions in all regions; (5) more than 5% of transitions in at least four states; (6) more than 2.5% of transitions in at least six states; and (7) more than 1.5% of transitions in at least eight states.

c. Within States, across microregions

Some markets cross commuting zones, or microregions, but do not cross state boundaries. Figure 2.3 plots the nodes of markets 49 and 133. The modal occupation in both is truck driver (modes have 66% and 77% of transitions, respectively), but market 49 is restricted to the state of Rio Grande do Sul and market 133 is restricted to the state of Mato Grosso. Each market includes nodes from around 50 industries.

A substantial number of workers participate in state dispersed markets. Table 2.4 shows that 1,194 markets are state dispersed, containing 10.6% of job transitions.

d. Within microregions

Most workers participate in markets that are concentrated within a microregion and include many counties. Table 2.4 shows that there are 2,830 markets—containing 32.2% of job transitions—in which more than 90% of transitions occurred within a single microregion (and are not part of the next category). From the top 1000 markets, 266 are contained within a microregion.

e. Within States, across microregions

Most markets pertain to a single county. Table 2.4 shows that there are 46,059 markets in which more than 90% of job transitions occurred in a single county. These markets correspond to 27% of job transitions, and while most markets are small and have a single node, 369 markets are within the top 1000 markets.

A substantial number of markets does not fit in any category. Most of these markets are concentrated within a county but have a reasonable share of transitions in other counties, microregions or states.

(iii) **Most markets include many occupations and industries**

Data-driven markets are incredibly sparse across industries and occupations. Table 2.5 shows that most markets have nodes in two or more different 2-digit industries and 1-digit occupations (69% and 78% of markets, respectively). While ad-hoc labor market definition yields similar numbers for industry dispersion, by construction, these markets are not spread across occupations.

6. Discussion

The labor market is a global network of firms and individuals in which firms offer wages to individuals in order to compensate them for their work. All workers and firms are somehow connected, such that a shock in the soybean production in Argentina will eventually affect the wages of workers in the retail sector in China. That said, workers tend to cluster by some observable characteristics, like occupation, industry, and geographic location. Any event within the cluster should affect workers and firms in a faster and more direct way.

In this paper, we presented an empirical method that attempts to identify such clusters by using job transitions across fine occupation-industry-location cells. We refer to clusters as data-driven labor markets, or labor market approximations. The method used in this paper—the constant Potts model—is successful in identifying data-driven labor markets where most job transitions occur within labor markets, and not across labor markets.

There are four important takeaways from the analysis of data-driven labor markets: (i) There is substantial heterogeneity in market size and market concentration; (ii) Firms hire workers from many different markets; (iii) Some markets are geographically dispersed; and (iv) Most markets include many occupations and industries. These takeaways highlight the benefits of using the labor markets identified through our method instead of ad-hoc labor market definitions.

In this paper we did not verify the robustness and validity of data-driven labor markets.

First, the CPM method requires the choice of a constant, which will help determine the number of markets. It is possible that changes in this constant will change the composition and number of markets. Future work should verify the robustness of data-driven labor markets to changes in the Potts constant. Second, it is necessary to develop an approach that determines if data-driven markets predict the consequences of events in the real world. For example, do movements in wages correlate across workers within markets? Does the effect of a plant closure on workers wages dissipate as predicted by the data-driven labor markets? And do increases in minimum wages reduce employment of workers in less concentrated markets? Future work should check the validity of data-driven labor markets.

7. Figures and Tables

(a) Market #91

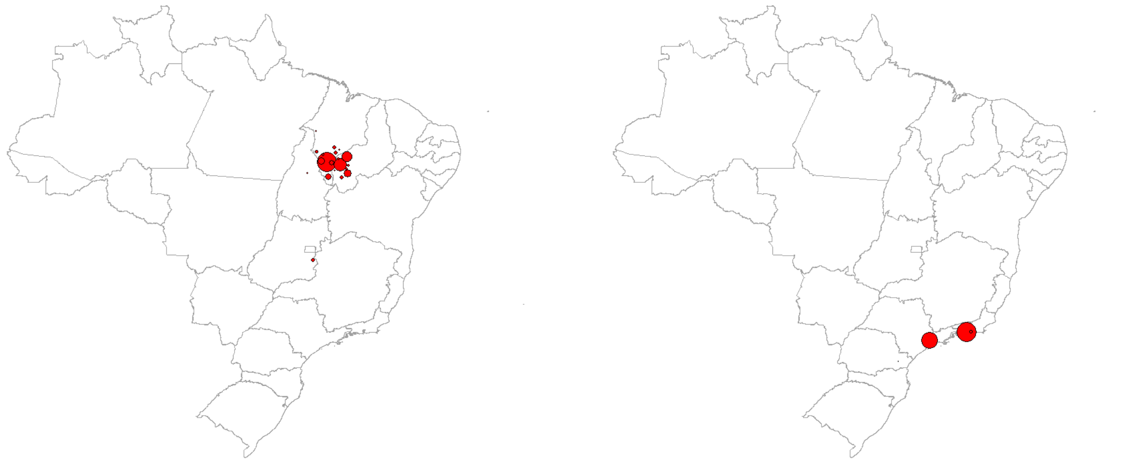
Modal Occupation: Civil Engineer

(b) Market #80

Modal occupation: Welder

Figure 2.1: Nationally dispersed markets

Note: The figures describe two distinct data-driven labor markets. Each circle represents a county from a node in the market, and the size of the circle is proportional to the number of workers from that county that participate in the market.



(a) Market #325

Modal sector: Soybean farming

(b) Market #897

Modal occupation: Elevator Maintenance

Figure 2.2: Markets in neighboring states

Note: See Figure 2.1.

(a) Market #49

Modal sector: Truck driver

(b) Market #133

Modal occupation: Truck driver

Figure 2.3: State dispersed markets

Note: See Figure 2.1.

Table 2.1: Job transitions and sample size

	Raw data	Main data
Nodes (county X 5-digit industry X 6-digit occupation)	5,360,609	246,638
Edges	45,270,686	6,819,564
Transitions (2007-2013)	82,321,061	43,874,702
Transitions between the same node (% of total transitions)	9.9%	34.8%
Average edges per node	8.4	27.7
Average transitions per node	15.4	177.9
Average transitions per node (except own)	13.8	115.9

Note: The table presents sample sizes of the raw dataset and the main dataset used in the paper. A node is a county X 5-digit industry X 6-digit occupation cell. Edges are non-empty links (or transitions) between nodes. Edges between two cells may have more than one transition. RAIS 2007-2013.

Table 2.2: Descriptive Statistics of Markets

	Data-driven Markets	Microregion X 6-digit occ.
Number of markets	54,855	71,361
Number of nodes	246,638	
Number of transitions	43,874,702	
Transitions within markets / Total transitions	75.1%	45.8%
Transitions within markets / Total transitions (excludes transitions within the same node)	61.8%	16.8%
Nodes per market	4.5	3.5
Transitions per market	799.8	614.8
Markets with...		
1 node	40,171	39,254
2 to 5 nodes	6,912	23,672
6 to 9 nodes	3,800	4,021
10 to 99 nodes	3,666	4,303
100 to 999 nodes	297	111
1000 or more nodes	9	0
Maximum number of nodes in a market	2,794	478
Share of job transitions in the...		
Top 50 markets	55.4%	17.0%
Top 100 markets	64.6%	23.7%
Top 500 markets	83.4%	44.5%
Top 1000 markets	88.4%	55.2%

Note: Nodes are defined as the county X 6-digit occupation X 5-digit industry cell. The first column presents descriptive statistics of the data-driven labor markets that are produced through the Leiden algorithm. The second column presents descriptive statistics of a commonly used ad-hoc labor market definition where a market is defined by the intersection between a microregion and a 6-digit occupation. RAIS 2007-2013.

Table 2.3: Labor market concentration and number of markets by firm

	Min	p25	p50	Mean	p75	Max	N
Panel A: HHI							
Data-driven markets							
Market distribution	9.7	2800.0	7222.2	6450.7	10000	10000	54,855
Worker distribution	9.7	17.8	51.2	614.6	266.6	10000	16,828,223
Microregion X 6-digit occupation							
Market distribution	3.2	1533.3	4583.3	5194.3	10000	10,000	71,361
Worker distribution	3.2	59.5	249.6	1341.3	1136.6	10000	16,828,223
Panel B: Number of markets per firm							
All firms	1	1	1	1.5	1	757	1,440,839
Top employers (75%)	1	3	7	10.8	13	757	13,308
Top employers (50%)	1	9	22	40.9	47	757	738

Note: Panel A presents the distribution of markets over the Herfindahl Hirschman Index (HHI). The HHI was computed using employment shares of each firm in the market. Panel B presents the distribution of firms by number of markets in which they hire workers from. The line "Top employers (75% of workers)" refer to the firms, ranked by employment size, that employed 75% of the workers in Brazil in 2014. Data: RAIS 2014.

Table 2.4: Types of markets by geography

Geography Group	All		Top 1000 markets	
	Count	Transitions (%)	Count	Transitions (%)
a. Nationally dispersed	432	4.8%	39	4.3%
b. Neighboring states	401	2.1%	17	1.9%
c. State dispersed	1,194	10.6%	149	9.4%
d. Single Microregion	2,830	32.2%	266	29.8%
e. Single county	46,059	27.0%	369	22.2%
f. Other categories	3,622	23.4%	160	32.5%

Note: See footnote 1 for definition of nationally dispersed markets. Top 1000 markets are ranked by number of transitions.

Table 2.5: Market composition

Share of markets composed of at least two different...	Data-driven Markets	Microregion X 6-digit occ.
States	26%	-
Mesoregions	43%	-
Microregions	53%	-
Counties	69%	69%
2-digit industries	69%	70%
3-digit industries	76%	81%
1-digit occupations	78%	-
2-digit occupations	83%	-
3-digit occupations	88%	-
Number of markets	14,684	32,107

Note: The table only includes markets that have more than one node.

Chapter 3

Has the College Premium really Flattened? The Entry of New Schools

1. Introduction

The statistics commonly known as the college wage premium has received great attention in the past fifty years. Its charm comes from highlighting movements in wage inequality (Freeman, 1976;Katz and Murphy, 1992;Card and Lemieux, 2001;Goldin and Katz, 2008;Acemoglu and Autor, 2011) and for serving as reference for public and private investments in higher education (Rodriguez et al., 2016; Bank, 2019). While the college premium has increased in developed countries, it has flattened or even decreased in many Latin American countries during the last 20 years (Fernandez and Messina, 2018). At the same time, there was a great expansion in the number of college graduates and college institutions in the region. Although it is usual to interpret the effects of such expansion on the college wage premium through the lenses of market equilibrium, the characteristics of workers with a college degree has considerably changed over time and possibly affected this measure.

In this paper, we study whether a reduction in the average quality of college institutions is responsible for the flattening/decrease in the college premium in Brazil. First, we show that students from older institutions perform better in the end-of-degree standardized exams, which suggests that these are better schools. Second, we show that the college premium has actually increased when we hold constant a set of universities. Combining these two facts, we infer that a composition change is driving the flattening of the college premium. There are more workers with a college degree, but with lower quality degrees, which reflects into lower average wages.

To perform this analysis, we use a novel data of one million students that graduated between 1990 and 2020 in 42 Brazilian universities. The data was constructed from FOIL requests (Freedom of Information Law) and is representative of the best universities in the country. We use students and workers full names to match this data with the Brazilian employer-employee matched dataset (RAIS) from 2007 to 2018. We use a machine learning approach to decide which is the best match for student-workers with multiple matches. As a result, we have annual data on workers of different ages and different cohorts that graduated in a constant set of universities.

We first show that the quantity and composition of college graduates changed over time. The number of people graduating from any college institution increased by three times between 2000 and 2018, from 400 thousand to 1.2 million people. Yet, most of the growth came from new institutions and majors. The share of people graduating from the top universities has significantly decreased during the period. In addition, we show that newer universities are often lower ranked, and their students perform worse in standardized exams. We interpret this as preliminary evidence that the average skill level of college graduates decreased over time.

Our main results show that the college premium restricted to workers from a constant set of universities has actually increased. For this analysis we use the RAIS dataset matched to our sample of college graduates as well as all the other workers with a college or high school degree. After residualizing log wages using set of fixed effects, we find that the college wage premium increased by 23% between 2007 and 2018—an average increase of 2% per year. When we do not condition on the same set of universities, we find that the college premium decreased by 12%.

A story in which all universities—including lower ranked institutions—have increasing college premium trends that aggregate into a decreasing overall college premium trend is consistent with our results. This conclusion depends on three facts: (i) All universities have an increasing college premium trend; (ii) At any given year, students that graduated from higher ranked universities earn higher wages; (iii) The share of students graduating from lower ranked universities increased over time. We have already mentioned that (iii) holds in our data. In addition, we show that (ii) holds by presenting a correlation between earnings and university ranking. We find that students from universities ranked in the 1st place receive 20% higher wages than students from universities ranked in 100th place. Additional data is required to prove that (i) holds and this should be the object of future research.

This paper contributes to few strands of the literature. We are the first to our knowledge to show that changes in the quality composition of higher education institutions are responsible for the trends in the college premium in Brazil. Although this hypothesis has been brought up by several studies, the literature had not reached a consensus probably due to data limitations.[1] Rodriguez et al. (2016) and Camacho et al. (2017) show that returns to college are heterogeneous in Chile and Colombia, respectively, with recently created programs having lower returns.[2] However, due to data limitations, they do not show the effect on the college premium trends. We contribute to the discussion by constructing a long panel with matched information on workers' wages and universities for different cohorts of workers.

Our results are also relevant to the growing literature that looks at the causes of the decline in earnings inequality in Latin America. Barros et al. (2010) argue that half of

[1] This hypothesis is sometimes referred to as "degraded tertiary hypothesis". See Messina and Silva (2017) for a review on the topic.

[2] Camacho et al. (2017) argue that the effects come from self-selection.

the decline in inequality was caused by an acceleration of educational progress. Ferreira et al. (2017), Alvarez et al. (2018) and Fernandez and Messina (2018) disagree with the latter statement and argue the decrease in earnings inequality is due to a compression of returns to firm and worker characteristics, such as experience and education We argue that decompositions of certain measures of inequality mistakenly attribute decreasing returns to education to a group that had increasing returns, possibly biasing the results.[3]

The results in this paper have strong implications regarding the public decision to invest in higher education. Previous studies found that the labor demand for college workers is relatively inelastic and that firms can easily substitute skilled for non-skilled workers (Katz and Murphy, 1992; Acemoglu, 2002; Ciccone and Peri, 2005;Haanwinckel, 2020). This implies that public investments that increase the supply of workers have the unintended effect of reducing relative wages for all workers with college. Our results show that returns to high quality education have continued to increase, consistent with skill-biased technological change and/or a more elastic labor demand. As a consequence, we learned that investments in higher education had increasing returns in the past, a trend that may continue in the future.

2. Data Sources and Descriptive Statistics

We gathered information on college graduates from a selected sample of universities and match them to the Brazilian employer-employee dataset in order to compute the college premium.[4]

A. Employer-Employee Matched Dataset (RAIS)

RAIS (Relação Anual de Informações Sociais) is an employer-employee panel data that

[3]Common measures of inequality are the ratio of log earnings of the 90th and 50th or 10th percentile in the wage distribution.

[4]College premiums are usually computed using household surveys, which are representative at the national level. However, such surveys do not include information on worker's schools.

covers the universe of formal employees and firms in the private and public sectors. The data is administered by the Ministry of Labor and has restricted access. RAIS has information on individuals (CPF, full name, age, gender, race, schooling), on firms (CNPJ and sector), on establishments (county, zip code and name) and on the employer-employee match (wages, occupation, tenure, dates of hiring and firing/separation). We use the data from 2007 to 2018.

B. College Graduates Sample

We gathered data on college graduates from public universities in Brazil through FOIL requests (Freedom of Information Law). The data consists of full name, university, major, year of admission, and year of graduation. Appendix Table C.1 presents the sample of universities that agreed with providing this information. We have information on 1.2 million students that graduated from 42 federal/state universities between 1990 and 2020.[5]

To estimate the returns to college adjusted by university composition, we match the college graduates' sample with the RAIS dataset. We match these datasets using student-workers' full names and, when there are multiple matches, we select the best match using a machine learning algorithm. After imposing some sample restrictions, we are left with 545,478 students matched to a worker with the same name. We explain the matching and data restrictions in Appendix 2..

C. Higher Education Census and Ranks

We use the Brazilian census of higher education, the ENADE national examination (Exame Nacional de Desempenho dos Estudantes), and two rankings of higher institutions that are published online (RUF, from Folha de Sao Paulo newspaper; and Web ranking) to complement the analysis. We use these data to describe the changes in composition of college graduates over time in terms of school's age and ranking in section 3.

[5] The variables and the number of cohorts available vary across universities.

Table 3.1 presents the age distribution of universities included in the college graduates' sample and all other universities. The data comes from the RUF ranking and is limited to 194 universities. The table shows that most universities in our sample were founded more than 50 years ago (59.5%). Table 3.2 presents the RUF score and ranking, and the web ranking for these two samples. Universities in our sample have better scores and as a consequence are better ranked, according to the RUF ranking. The Web ranking includes more universities (1,285) and shows a greater discrepancy between the sample universities and the out of sample universities. The median ranking in the sample is 37, and 663 out of sample. In summary, our sample includes many of the best and oldest universities in the country.

D. Age-adjusted College Premium

Figure 3.1 presents the evolution of the college premium over time using the nationally representative household survey (PNAD, in Panel A) and the employer-employee matched dataset (RAIS, in Panel B). In this analysis, college premiums are defined as the weighted ratio of earnings for workers with a college degree and workers with a high school degree. Similar to Fernandez and Messina (2018), premiums are constructed using a fixed-weight average of every age subgroup, for workers of ages between 21 and 65 years old. The weights are equal to the mean employment share of each subgroup across all years. We present the weighted average by aggregating all groups.

Using the PNAD survey data, Panel A of Figure 3.1 shows that the college premium increased between 1997 and 2004 and then decreased between 2004 and 2015. Fernandez and Messina (2018), describe a similar picture. Using the sample of formal workers, Panel B of Figure 3.1 shows that the college premium had a period of strong increase between 1994 and 2003, a weaker increase between 2004 and 2012 and decreased between 2012 and 2019. These trends have been interpreted as the equilibrium response of positive supply shocks of skilled workers. In fact, Figure 3.2 shows that the number of formal workers with a college degree has substantially increased between 1995 and 2019.

In the rest of the paper, we compare the college premium trends presented in panel B of Figure 3.1 with measures that account for changes in university composition.

3. Preliminary Evidence of Changes in Graduates' Skill Composition

We argue that the growth in the supply of college workers reduced the average skill of college graduates.

First, most of the growth in the supply of college graduates is due to the introduction of new majors and institutions. Figure 3.3 uses annual data from the higher education census and presents the number of students that graduated from a bachelor's program in each year. In 2000, around 400 thousand people graduated from a bachelor's program. The number of graduates was three times greater by 2018. Much of this growth comes from new majors and new institutions. The figure shows that the number of graduates from majors that were created before the year 2000 is steady over time and possibly decreasing. Growth in number of graduates in the 2000's (2010's) comes from majors founded in the 2000's (2010's) and not by increases in class size from older majors.[6]

Secondly, new institutions are lower ranked, and their students perform worse on national exams. We regress the ENADE score and RUF ranking on dummy categories of institutions' age, as defined by the year of foundation of the university's first major. Table 3.3 presents the results, which should be interpreted as deviations from institutions founded before 1940. The table shows that students in universities founded recently have worse scores in the ENADE's

[6]Appendix Figure C.1 shows similar trends categorized by the age of the oldest major in the higher education institution. The figure shows that half the growth in the number of college graduates comes from institutions that were founded between 1990 and 2000 and after the year 2000.

exam in both specific knowledge (related to the major) and in general knowledge. Column 4 shows that RUD ranking is increasing on university's age.

In summary, older universities are better ranked but they used to represent a higher share of graduates. For example, Figure 4 shows that the share of graduates from Top 25 universities according to ENADE ranking has substantially decreased.

4. The Adjusted College Premium Trends

In the previous section, we learned that a smaller share of students graduated from the older and higher ranked universities in the past years. In this section, we study the effects of such change in composition on the college premium. We do that by decomposing the college premium into two samples: (a) the college graduates' sample from FOIL requests matched to RAIS, and (b) all workers in the RAIS dataset with a college degree, excluding workers from the previous sample. Note that "sample a" holds the university composition constant over time, but "sample b" does not.

To compute the college premium, we regress log wages of individual i, of schooling s, age a, on year t on a set of dummies as shown by Equation 3.1:

$$ln(wage)_{i(s,a)t} = \delta_{t,s} + \alpha_{a,s} + \psi_c + \eta_{u,m} + \varepsilon_{it} \tag{3.1}$$

Where $\delta_{t,s}$ are year by schooling fixed effects, $\alpha_{a,s}$ are age by schooling fixed effects; ψ_c are cohort of graduation fixed effect; and $\eta_{u,m}$ are university by major fixed effects.[7] The previous two fixed effects also equal to zero for high school workers and for workers whom we do not have university data. We omit dummies for the first year in the data due to collinearity. The sample includes all workers in the RAIS data between ages 21 and 65 that

[7]We are able to separately identify age, cohort, and year effects because we observe workers of different ages in the same cohort of graduation and over time. I.e., age, cohort, and year do not form a collinear relation.

were employed on December 31st, worked 40 hours a week, had positive wages in December and had either complete high school or college education. The resulting sample has 263 million observations.

As shown in Section 2, the unconditional college premium has a decreasing pattern. Figure 3.5 presents the estimates of $\delta_{t,s}$ from Equation 3.1, which represent the changes in log wages of each sample relative to the wages of high school workers, taking the year of 2007 as the benchmark. The dashed line represents the college premium for the sample of all workers in the RAIS dataset with a college degree, excluding workers from “sample a”. The figure shows that the college premium remained constant between 2007 and 2011 and decreased by 14.5% between 2011 and 2018—a 1.8% annual decrease.

However, when fixing the sample to students that graduated from the same group of universities, we note that the college premium has actually increased. The black line in Figure 3.5 presents the college premium for the college graduate’s sample from FOIL requests (Sample Universities). The figure shows that the college premium increased by 19% between 2007 and 2011—a 4% annual increase. The college premium increases at a slower pace between 2011 and 2018, by 5% in total or 0.6% annually.

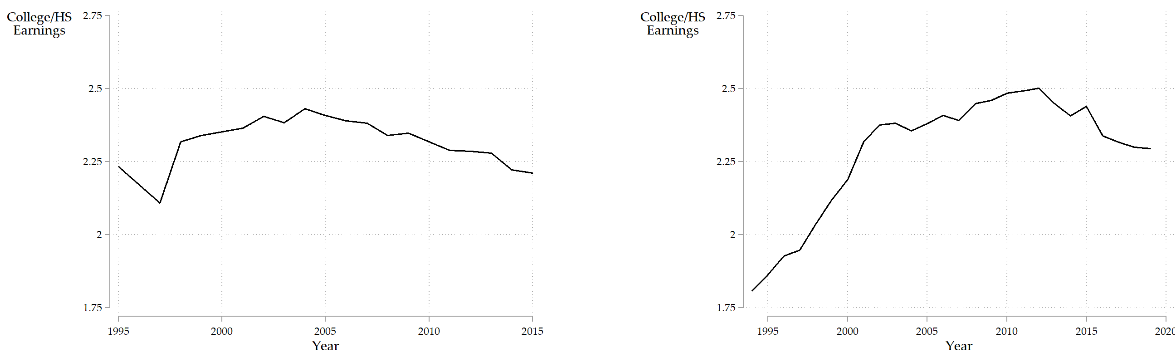
In theory, all universities could have an increasing college premium that aggregates into a decreasing overall college premium. Table 3.4 presents the correlation between the university fixed effect from Equation 3.1 with university rankings (where #1 is the best). The table shows that a move in ten positions in the university ranking is associated with a wage difference of 2%. For example, students from universities ranked in the 1st place receive 20% higher wages than students from universities ranked in 100th place. The point is that all these universities could have increasing trends in the college premium but at different levels. However, the number of students graduating from lower ranked universities (and lower fixed effect) has increased. As a result, the overall college premium gives stronger weight to workers from lower ranked universities by the end of the period, showing a decreasing trend.

5. Conclusion

We presented evidence that the college premium in Brazil has increased, opposing previous results in the literature. The difference in results comes from the construction of a new dataset that identifies worker's university. We find that the college wage premium increased by 23% between 2007 and 2018 when holding constant the set of universities for which we have data and decreases by 12% in the overall sample. In addition, we showed that the supply of workers with college degree has significantly increased, but much of this increase came from newer, lower ranked and lower wage-premium universities.

The results are relevant for the estimation of labor demand elasticities and to calculate the importance of skill biased technological change. Future research should try to include measures of skill in such models in order to account for changes in skill composition of workers in the same schooling group. The results also inform individuals and policymakers regarding the decision to invest in higher education. That said, future research should focus on verifying if these trends are similar for all universities in Brazil and in other countries.

6. Figures and Tables



(a) PNAD **(b)** RAIS

Figure 3.1: College/High School wage premium

Note: The graph plots the ratio between wages of workers with college degree and workers with high school degree adjusted for age composition. The sample is restricted to workers between 21 and 65 years old. Panel A uses the Brazilian household survey (PNAD). Panel B uses the Brazilian matched employer-employee data (RAIS).

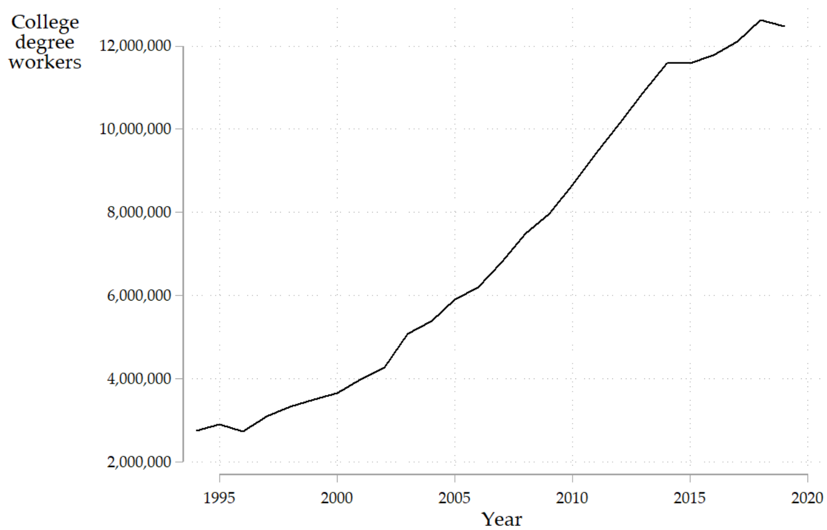


Figure 3.2: Formal workers with a college degree

Note: The figure plots the trends of the number of formal job relations with positive wage in which workers had a college degree. Data: RAIS 1994-2019.

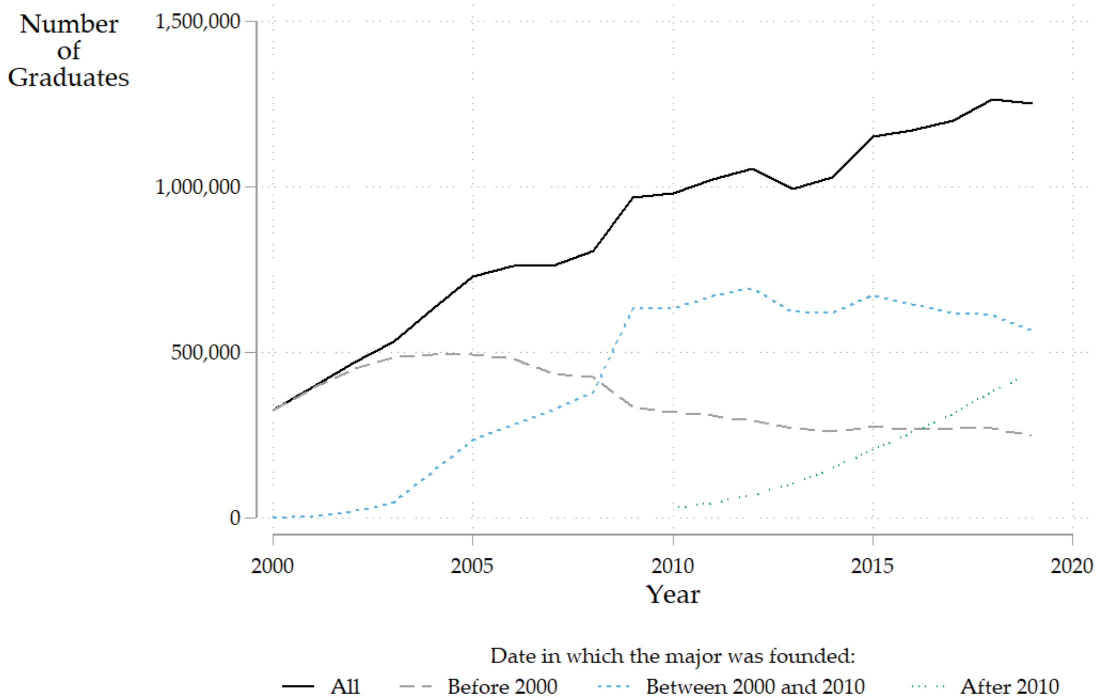


Figure 3.3: Number of graduates by major's date of foundation

Note: The figure presents the number of students that graduated in each year. Categories are defined by the year in which the major was founded. Source: Higher Education Census (2000-2018)

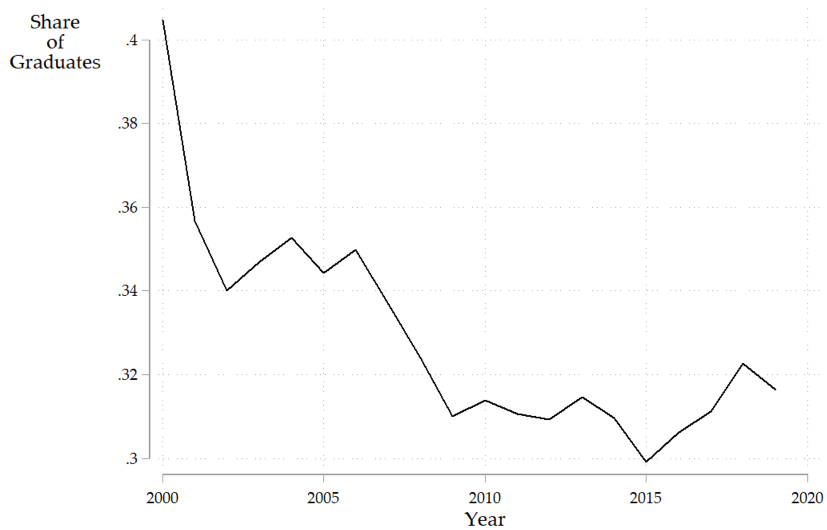


Figure 3.4: Share of graduates from Top 25 Universities according to ENADE ranking

Note: The number of graduates from each university comes from the higher education census.

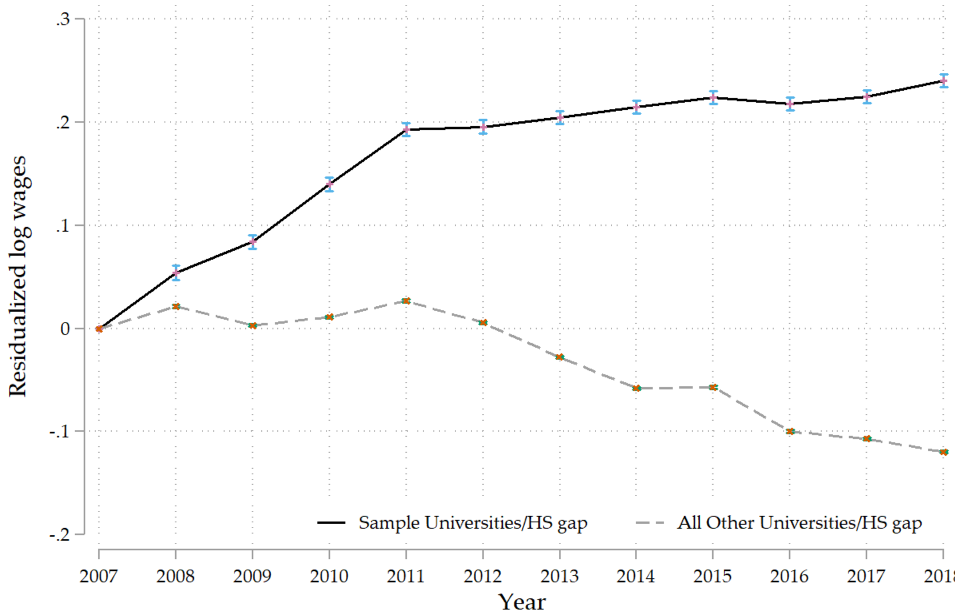


Figure 3.5: Trends in residualized log wages (college premium)

Note: The figure plots the evolution of the college premium, relative to 2007 values, for the college graduate's sample from FOIL requests (Sample Universities) and for the sample of all workers in the RAIS dataset with a college degree, excluding workers in the Sample Universities (All Other Universities). Both curves are relative to the same trends in the high school residualized wages. Estimates come from the estimation of Equation 1 over a sample of 263,181,905 observations. The figure plots 95% confidence intervals for each point estimate.

Table 3.1: University's age

University's Age	Sample	Out of Sample
$<$ 30 years	10.80%	31.20%
30 to 50 years	29.70%	33.10%
$>$ 50 years	59.50%	35.70%
N	37	157

Note: The first column refers to the universities from the college graduates' sample. The second column refers to all other universities in the RUF ranking.

Table 3.2: Institutions ranking and score

	Mean	S.D.	Median	Min	Max	N
RUF score						
Sample	69.6	21.7	42.1	4.8	98	37
Out of sample	43	21	74	4.2	97	157
RUF ranking						
Sample	48.3	48.7	33	3	197	37
Out of sample	110.2	52.6	111	1	196	157
Web ranking						
Sample	52.7	50.4	37	3	219	41
Out of sample	662.5	361	663.5	1	1285	1244

Note: RUF range between 0 and 100. The RUF ranking includes 194 universities and the Web ranking includes 1,285 universities.

Table 3.3: Association between university's age and quality

Dependent variable:	ENADE score			RUF	
	All	Specific knowledge	General knowledge	Ranking	Score
	(1)	(2)	(3)	(4)	(5)
Year in which the institution's first major was founded:					
1940 < year < 1960	-2.948***	-3.211**	-2.160*	46.12***	-19.85***
	(1.075)	(1.261)	(1.175)	(10.21)	(3.953)
1960 < year < 1980	-5.138***	-5.294***	-4.671***	72.66***	-31.13***
	(0.857)	(1.006)	(0.937)	(8.735)	(3.336)
1980 < year < 2000	-6.299***	-6.248***	-6.452***	97.31***	-37.90***
	(0.867)	(1.017)	(0.948)	(13.86)	(5.815)
2000 < year < 2020	-5.863***	-5.654***	-6.490***	112.2***	-43.10***
	(0.818)	(0.959)	(0.894)	(13.54)	(6.301)
Constant	49.32***	46.04***	59.13***	41.76***	74.76***
	(0.796)	(0.933)	(0.870)	(7.174)	(2.700)
Observations	1,469	1,469	1,469	197	171
R-squared	0.046	0.030	0.058	0.364	0.402

Note: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Table 3.4: Correlation between University ranking and wages

Dependent Variable:	University fixed effect	
	(1)	(2)
Ranking RUF	-0.0016**	
	(0.0008)	
Ranking Web		-0.0021**
		(0.0008)
Observations	33	36

Note: The dependent variable is the fixed effects from the estimation of Equation 1, with the exception that we include university fixed effects and do not include major by university fixed effects. Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Appendix A

Appendix Materials for Chapter 1

1. Data Appendix

In this section I describe data restrictions to the drugstore and the fully connected samples.

- Workers with multiple jobs during a year will appear several times in the data. I restrict the sample to individuals with a formal employment link on December 31.
- It is possible that establishments such as distribution centers, headquarters, and warehouses are wrongly classified as drugstores. To prevent that, I drop from the sample establishments that are sorted in one of following criteria:
    1. establishments with more than 100 full-time employees in any year
    2. establishments with more than 50 full-time employees and no pharmacist in any year, and
    3. establishments with more than 10 full-time employees and more than 60% of employees in warehouse occupations in any year.
- In the fully connected sample, I restrict the data to workers between 18 and 60 years old, who have 40 or more contractual weekly hours of work, and who work in the

states in which the merging firms had an establishment in 2010. If workers have more than two observations in a year, I keep the two observations with the earliest dates of admission.

- In 2012, the merging firms had to change the establishments' identification numbers. I re-code these establishments' IDs so that they have the same IDs throughout all of the time series. I also re-code workers' tenure. I assume that two establishments with different IDs are actually the same establishment if (i) they are in the same county, (ii) they share 50% or more of their labor force, and (iii) one of the establishments exits the data when the other enters. In addition to changing IDs, firms had to report their workers' information to RAIS twice in 2012. Although I can identify this in the data, I exclude the year 2012 from the main specification to prevent reporting mistakes. That said, including 2012 does not significantly change the results.

In Appendix Table A.7, I present additional statistics on the drugstore sample and on less restrictive samples.

2. The retail pharmacy sector in Brazil

In this section, I present more detailed information on the retail pharmacy sector and the evolution of pharmacists' occupation.

What does a pharmacy look like in Brazil? Retail pharmacies, or drugstores, are facilities that sell medications, cosmetics, and pharmaceutical products. Pharmacies may also administer vaccines and compound medication, and can sell some food products for special purposes (Sebrae-SP, 2015). Pharmacies in Brazil are typically smaller than pharmacies in the US, both in terms of space and number of employees. In Brazil, an average drugstore has 8.5 employees, where 1 or 2 of these employees are a pharmacist and 4 to 5 are salespeople (including cashiers). Other employees can be working in managerial tasks, cleaning, product delivery, or organizing inventory.

Even though pharmacies are private companies, they are highly regulated and form part of the Brazilian health system. The list of products a pharmacy is allowed to sell is regulated by state laws and by Anvisa (Agencia Nacional de Vigilancia Sanitaria), the national health regulatory agency. In Brazil, there are also maximum "factory price" and "consumer price" restrictions on a list of essential medicaments. These restrictions are defined at the national level by Anvisa.[1] A pharmacy must also have an employee registered with the state's pharmacy council (Conselho Regional de Farmácia) available at all times when the store is open to the public.[2] In 2014, a new law specified that this professional should be a pharmacist, but this requirement had been binding since at least 2009 via decree.[3] Last, a requirement to work as a pharmacist in Brazil is having a university degree in pharmacy, which usually takes between four and five years.[4]

Informality is not a big issue in the retail pharmacy sector. As in most developing countries, Brazil has high levels of informality. However, probably due to the regulations and enforcement by Anvisa, the retail pharmacy sector has much lower levels of informality compared with other sectors. In 2012, 45% of workers in Brazil were in the informal sector. However, only 17.7% of people working in the sector that includes drugstores, fragrance shops, and shops for other medical and orthodontic products were working informally.[5] Since fragrance shops and shops for other medical and orthodontic products are not as regulated as pharmacies, I expect informality rates at pharmacies to be even lower. I use RAIS data in this paper, which only has information on workers formally employed. Thus, low informality

[1] CMED Resolution nº2, March 5, 2004.

[2] Law number 5,991 from 1973.

[3] Law number 13,021, from 2014 and Resolução – RDC nº 44/09 da ANVISA.

[4] As a comparison, it usually takes 8 years of college study to earn a pharmacist degree in the U.S.

[5] Own calculations based on the Brazilian household survey (PNAD, Pesquisa Nacional por Amostra de Domicilios).

rates are one of the attractive features of studying the retail pharmaceutical sector in Brazil.

Table A.8 shows that the retail pharmacy sector in Brazil was growing during the period of analysis. In 2018, there were around 30,000 firms and 47,000 establishments in the retail pharmacy sector in Brazil—42% more than in 2007. The sector employed around 400,000 workers, with 20% being pharmacists. The total revenue in the sector, of 154 billion reais, corresponded to 2.2% of Brazil's GDP. Between 2007 and 2018, the number of stores increased by 3.3% per year and total revenue in the sector increased by 8.9% per year. The substantial growth in the sector has been attributed to Brazil's economic growth and to higher demand for pharmaceutical products coming from a population that is aging.

The growth in the sector promoted changes in the composition of pharmacists and other workers. Table A.8 shows that the number of individuals working in pharmacies has almost doubled. This is true for both pharmacists and salespeople. Figure A.5 shows that the number of pharmacists per 10,000 residents increased from 3.3 to 6.6 between 2007 and 2018. While this increase did not significantly change the share of female pharmacists (69% to 73%, Figure A.6), it did increase racial diversity in the occupation. Figure A.7 shows that the share of whites decreased from 81% to 67% and the share of black and brown individuals increased from 18% to 32%. Age composition also changed, with an increase of 5 years in the modal age of pharmacists (Figure A.8). Last, the share of workers with a high school degree increased from 78% to 91% (Table A.8).

Despite the increase in employment, the wages of pharmacists also increased in the period. Appendix Table A.8 shows that pharmacists earn on average 4.2 times the minimum wage and salespeople employed in pharmacies earn 1.65 times the minimum wage. Although these ratios are constant or even decreasing over time, both the numerator—real wages—and the denominator—real minimum wage—have increased in the period. Between 2007 and 2018, real minimum wages had a strong increase of 2.7% per year. Figure A.9 shows that the real wages of pharmacists had a steady increase during the same period. Furthermore, wage increases seem to be similar throughout the wage distribution, with the 25th percentile, median wages, average wages, and the 75th percentile all growing by the same rate during the period. The fact that the employment and wages of pharmacists increased suggests that the demand for pharmaceutical services increased as well.

In the 2000s, some pharmacy chains started a consolidation and growth process. This was driven in part by Brazil's economic growth and the development of its financial market. A few groups concluded their IPOs in the mid-2000s. In 2007, firms with more than 100 establishments employed only 14% of those working in pharmacies. At the time, financial analysts expected the sector to go through a consolidation process similar to the U.S. retail pharmaceutical sector. By 2016, firms with more than 100 establishments employed 33% of those working in pharmacies. The share of revenue from larger chains also increased in the same period, reaching around 53% of the sector's revenue in 2016. However, the growth of pharmacy chains in terms of market share has been limited in the past few years and did not reach analysts' expectations. This is attributed to the steady market share of independent pharmacies, supported by the sales of generic medication through a governmental program (RD, 2018).

3. Appendix Figures and Tables

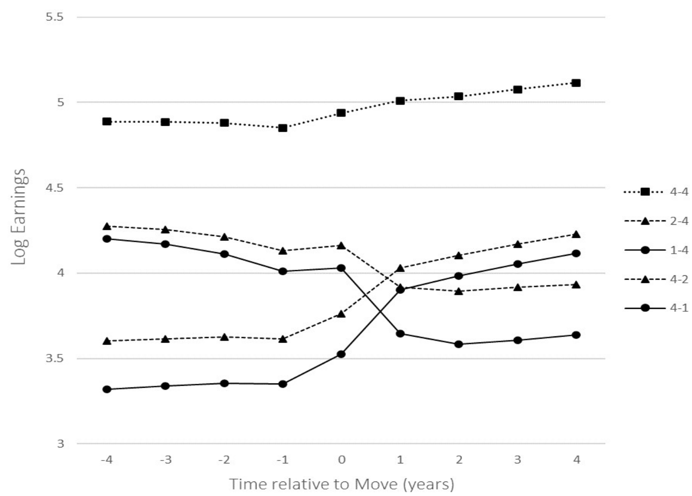
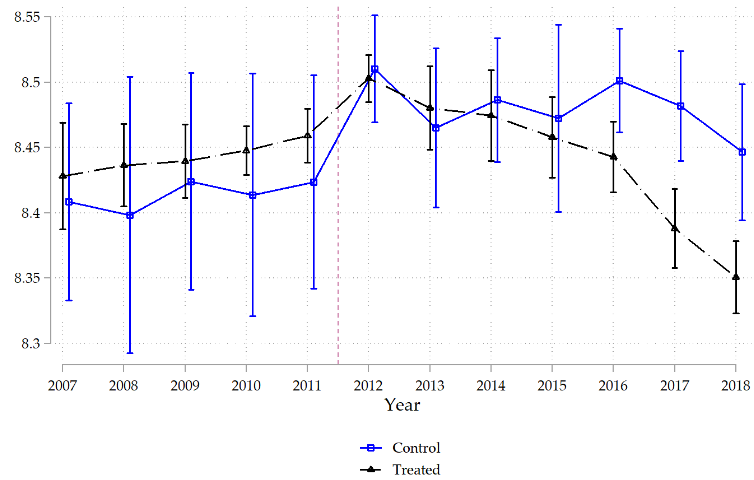
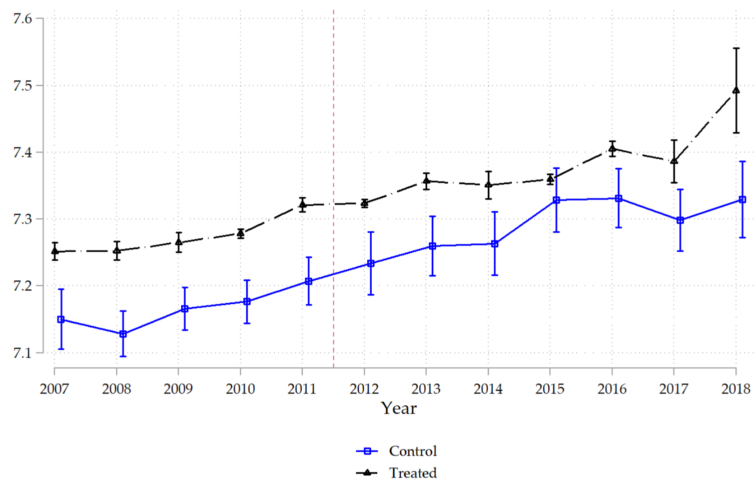


Figure A.1: Event Study of Changes in Earnings when Workers Move Between Firms

Note: In this figure, I classify firms into four equally sized groups based on the mean earnings of non-movers in the firm (with 1 and 4 being the group with the lowest and highest mean earnings, respectively). I then compute mean log earnings for the workers who move between these groups of firms in the years before and after the move. Note that the employer differs between event times -1 and 1, but we do not know exactly when the change in employer occurred. Thus, to avoid concerns over workers exiting and entering employment during these years, one might prefer to compare earnings in event years -2 and 2.



(a) Pharmacists



(b) Salespeople

Figure A.2: Trends - Ln(wage) of workers in merging firms

Note: The figure presents the average ln(wage) in treated and control groups. Estimates come from a regression that includes pharmacists (panel A) or salespeople (panel B) working in merging firms, from 2007 to 2018. Regressions do not include any additional controls. Log wages are measured in December. The sample only includes workers employed on December 31. The sample is composed of a balanced panel of establishments. Standard errors are clustered at the county level.

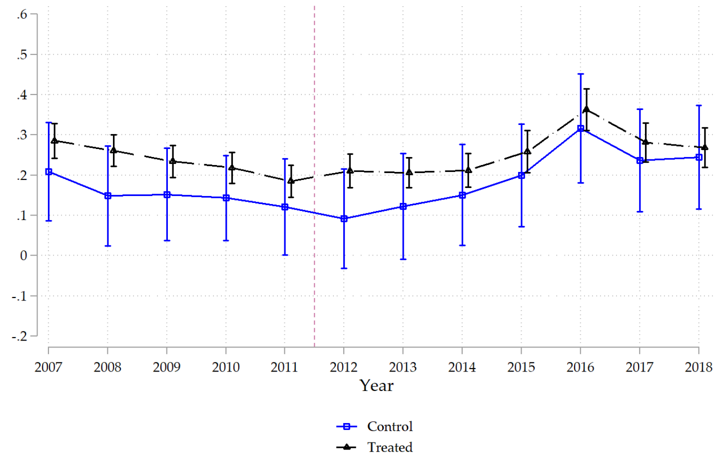
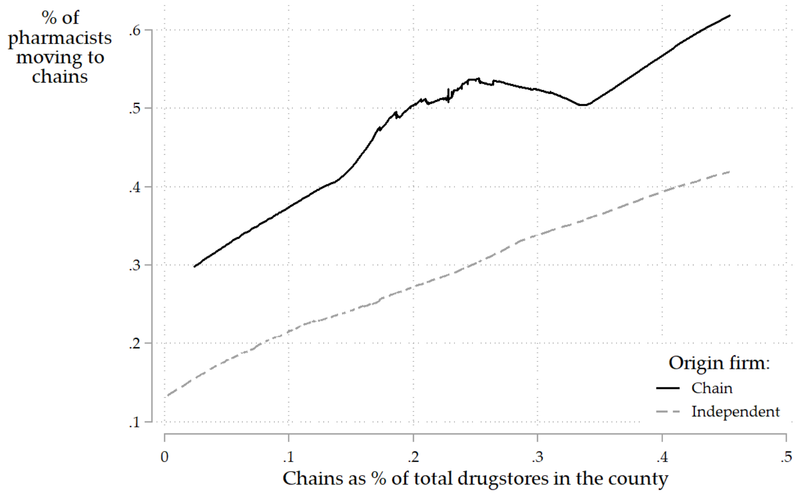
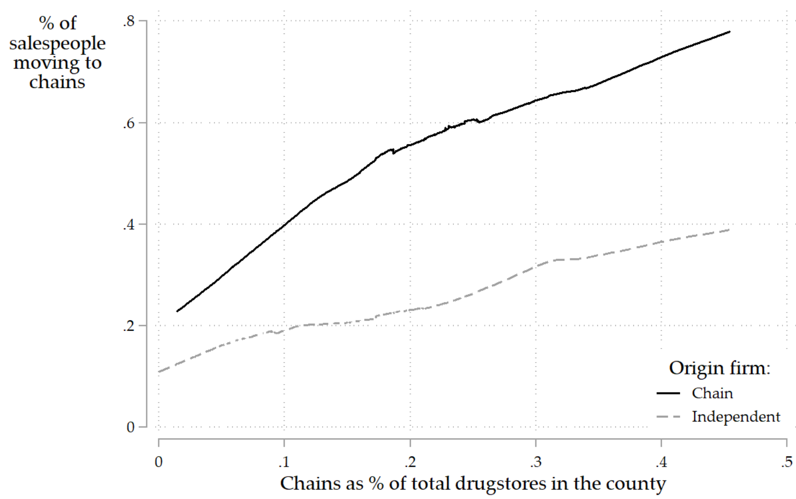


Figure A.3: Residual ln(wage) of pharmacists in merging firms with two-way fixed effects

Note: The figure presents the residual ln(wage) in treated and control groups for pharmacists in merging firms. Estimates come from a regression that uses the fully connected set and includes individual characteristics and worker and establishment fixed effects. Unlike Equation 1.3, treatment and control indicators are interacted with all year indicators and not just the post indicator. The figure reports the estimate and standard errors of the parameters associated with these interactions. Log wages are measured in December. The sample only includes workers employed on December 31. Standard errors are clustered at the county level.



(a) Pharmacists



(b) Salespeople

Figure A.4: Job transition probabilities conditional on moving to another drugstore

Note: The figures plot nonparametric regressions at the county level. The sample includes individuals working in drugstores in 2009 who switched jobs between 2009 and 2010 and stayed at a drugstore. The dependent variable is the percentage of pharmacists (Panel A) or salespeople (Panel B) who were working in a pharmacy chain in 2010. This is regressed on the ratio between the number of establishments from pharmacy chains and the total number of pharmacies in the county. The figure separates individuals who were working in 2010 in a pharmacy chain (continuous line) and individuals who were working in an independent pharmacy (dashed line). Source: RAIS 2009-2010.

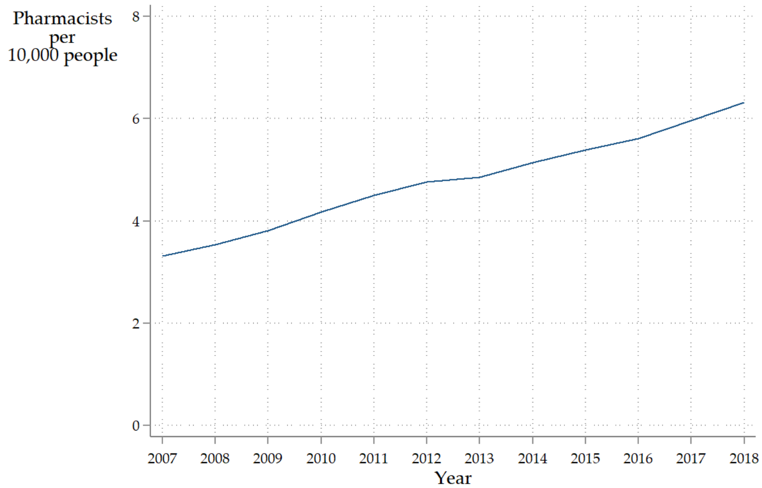


Figure A.5: Pharmacists per capita - Brazil

Note: The figure plots the evolution in the number of pharmacists per capita. Pharmacist data comes from RAIS 2007-2018. Yearly population estimates come from the Census and Ipeadata.

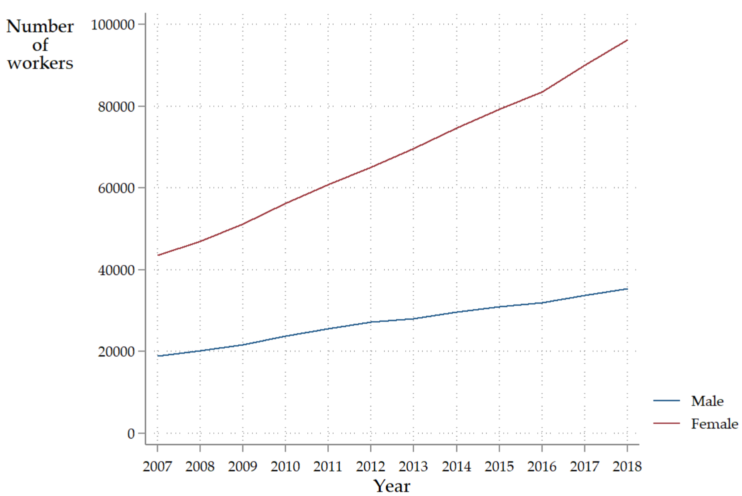


Figure A.6: Trends in pharmacists' gender composition - Brazil

Note: The figure plots the evolution in the gender composition of pharmacists from 2007 to 2018. The share of female pharmacists raises from 69% in 2007 to 73% in 2018. Source: RAIS 2007-2018.

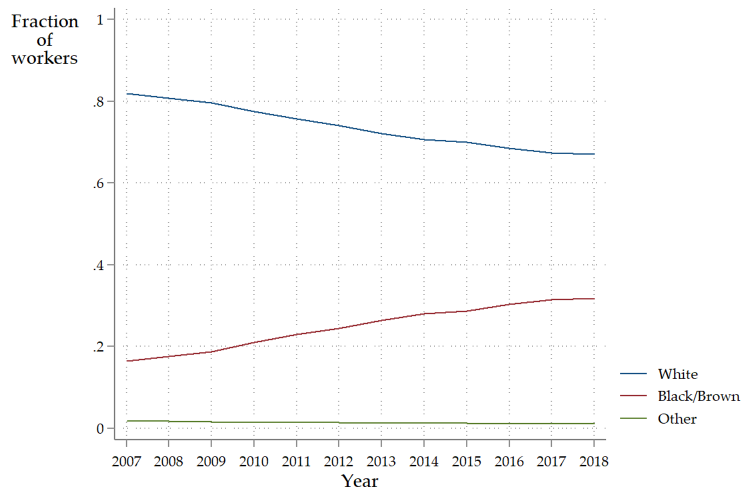


Figure A.7: Trends in pharmacists' race composition - Brazil

Note: The figure plots the evolution in the race composition of pharmacists from 2007 to 2018. The profession gets more representative of the Brazilian population over time, with the share of whites falling from 81% to 67% and the share of black and brown increasing from 18% to 32%. Source: RAIS 2007-2018.

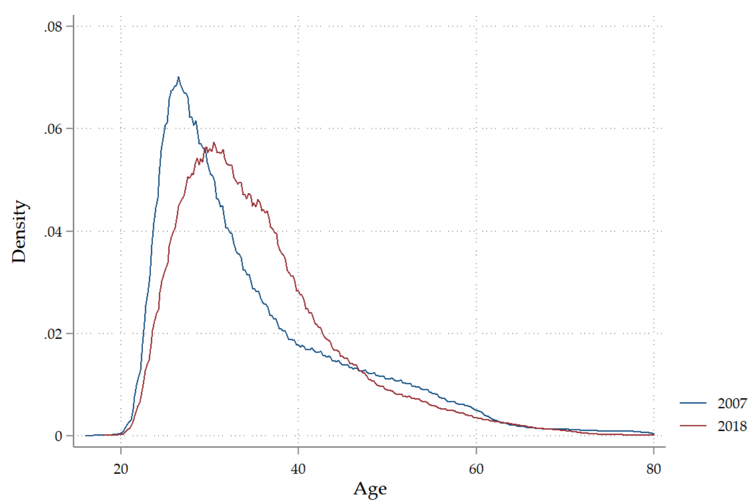


Figure A.8: Age distribution of pharmacists - Brazil

Note: The figure plots the age distribution of pharmacists in 2007 and 2018. The age distribution shifts to the right from 2007 to 2018. Source: RAIS 2007-2018.

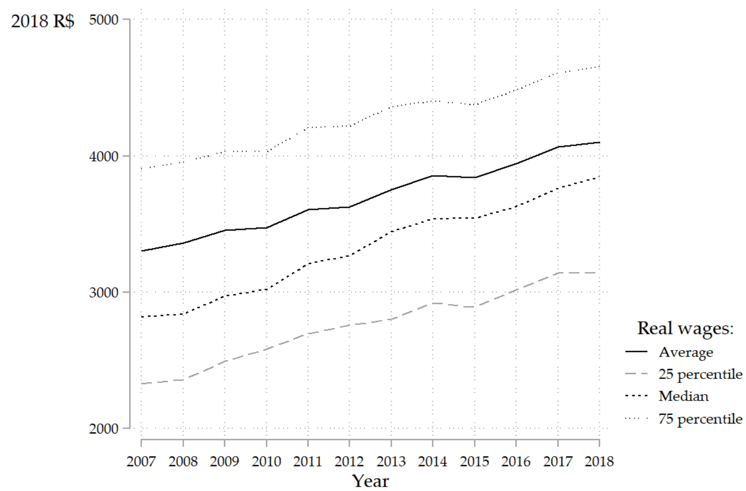


Figure A.9: Trends in pharmacists' wage distribution - Brazil

Note: The figure plots the evolution in the wage distribution of pharmacists from 2007 to 2018. Real wages are reported in 2018 reais. Source: RAIS 2007-2018.

Table A.1: Difference-in-differences estimates. Dep. variable is ln(wage)

Sample:	Pharmacists in merging firms			
	(1)	(2)	(3)	(4)
Post × Treat	-0.0731			-0.0628
	(0.0296)			(0.0416)
Post × ΔHHI		-3.878	-3.818	-0.870
		(1.876)	(1.908)	(2.345)
Treat	0.0310		-0.0672	-0.0266
	(0.0458)		(0.0434)	(0.0446)
ΔHHI		3.935	7.070	5.177
		(2.215)	(2.002)	(1.180)
Constant	8.401	8.391	8.412	8.399
	(0.0421)	(0.0305)	(0.0382)	(0.0422)
Observations	12,452	12,452	12,452	12,452
R-squared	0.023	0.022	0.032	0.033

Note: All specifications include year fixed effects. ΔHHI refers to the projected change in the 2011 Herfindahl-Hirschman Index (HHI) due to the merger. A labor market is defined by the interesection between county and occupation. The HHI is divided by 10,000, such that ΔHHI varies between 0 and 1. Average projected change in the HHI is of 0.0111 (or 111 point in the standard HHI scale). Standard errors are clustered at the county level.

Table A.2: Robustness. Difference-in-differences estimates. Dep.variable is Log(wage)

	December Wage				Average Wage			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Pharmacists in Merging Firms								
Treat	0.029	0.027	0.031		0.026	0.025	0.030	
	(0.045)	(0.044)	(0.044)		(0.056)	(0.054)	(0.054)	
Treat × Post	-0.073	-0.077	-0.076	-0.064	-0.076	-0.082	-0.082	-0.066
Cluster: County	(0.029)	(0.029)	(0.028)	(0.029)	(0.033)	(0.033)	(0.032)	(0.030)
Cluster: County X Year	(0.025)	(0.025)	(0.025)	(0.013)	(0.030)	(0.030)	(0.030)	(0.014)
Cluster: Establishment	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)	(0.012)
Robust	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)
Age controls		X	X	X	X	X	X	X
Other individual characteristics			X	X			X	X
Individual FE								
Establishment FE								
County FE				X				X
Year FE	X	X	X	X	X	X	X	X
Observations	12,328	12,325	12,324	12,324	18,776	18,772	18,771	18,771
Number of workers	5,127	5,127	5,127	5,127	6,612	6,612	6,612	6,612
Number of establishments	392	392	392	392	392	392	392	392

Note: The table presents estimate of the difference-in-differences parameter from Equations 1.1 at the individual level. The dependent variable is either the log of December wage or log of average wages within a year. Columns 1 to 4 only include workers employed on December 31. Columns 5 to 8 include more observations, since workers can have more than one employer in a year.

Table A.3: Robustness. Difference-in-differences estimates. Dep. variable is Log(wage)

Sample restrictions	All restrictions (1)	All, with weights (2)	includes outliers (3)	includes 2012 (4)	includes all establishments (5)
Panel A: Pharmacists in Merging Firms					
Treat × Post	-0.066	-0.061	-0.062	-0.056	-0.096
	(0.030)	(0.028)	(0.031)	(0.026)	(0.030)
Age controls	X	X	X	X	X
Other individual characteristics	X	X	X	X	X
Individual FE					
Establishment FE					
County FE	X	X	X	X	X
Year FE	X	X	X	X	X
Observations	18,771	18,771	18,960	20,499	43,969
Number of workers	6,612	6,612	6,637	6,746	13,309
Number of establishments	392	392	392	392	1,410

Note: The table presents estimates of the difference-in-differences parameter from Equation 1.1 at the individual level. Column 1 is similar to column 8 of Table A.2. The dependent variable is log of average wages. Columns 1 and 2 include all sample restrictions. Column 2 uses inverse employment weights, such that establishments have the same weight. Column 3 includes outliers: the top 1% of the observations in terms of wages. Column 4 includes observations from 2012. Column 5 includes all establishments from merging firms, not limiting it to the balanced panel of establishments. Standard errors are clustered at the county level and presented in parentheses.

Table A.4: Job transition for all workers

	Status in 2010				
Occupation in 2009	Same occupation as in 2009	Same county as in 2009	Same microregion as in 2009	Still works in a pharmacy	N
Pharmacist	83%	73%	81%	73%	50,594
Salespeople	74%	82%	86%	73%	301,893
Manager	77%	83%	87%	82%	30,359

Note: The table shows the job transition probabilities of all workers employed in the retail pharmacy sector in 2009. Columns 1 to 4 show the share of workers who were working in the same occupation, county, micro-region and pharmacy in 2010, respectively. Column 5 shows the sample size for each occupation. Source: RAIS 2009 and 2010.

Table A.5: Job transition for movers

	Status in 2010				
Occupation in 2009	Same occupation as in 2009	Same county as in 2009	Same microregion as in 2009	Still works in a pharmacy	N
Pharmacist	65%	47%	63%	41%	30,098
Salespeople	36%	54%	63%	24%	122,336
Manager	25%	49%	58%	31%	9,008

Note: The table shows the job transition probabilities of movers—that is, workers employed in the retail pharmacy sector in 2009 who were working in a different establishment in 2010. Columns 1 to 4 show the share of workers who were working in the same occupation, county, micro-region and pharmacy in 2010, respectively. Column 5 shows the sample size for each occupation. Source: RAIS 2009 and 2010.

Table A.6: Merger-induced changes in concentration

Labor market definition:	HHI in 2011 (1)	Projected HHI (2)	ΔHHI (3)	%ΔHHI (4)
Panel A: Pharmacists				
County X Occupation X All workers X Year	457	494	37	8.1%
County X Occupation X All workers X Year X Pharmacies	529	623	94	17.7%
County X Occupation X All workers X Year X Pharmacy chain	2,316	2,967	651	28.1%
Panel B: Salespeople				
County X Occupation X All workers X Year	77	77	0	0.2%
County X Occupation X All workers X Year X Pharmacies	433	488	55	12.8%
County X Occupation X All workers X Year X Pharmacy chain	2,269	2,870	601	26.5%
County X Occupation X Newly hired workers X Year	91	91	0	0.3%
County X Occupation X Newly hired workers X Year X Pharmacies	528	621	93	17.6%
County X Occupation X Newly hired workers X Year X Pharmacy chain	2,364	2,927	563	23.8%
County X Occupation X Newly hired workers X Semester X Pharmacy chain	2,987	3,556	569	19.0%
Panel C: Salespeople, 6-digit occupation code				
County X Occupation X All workers X Year	632	755	123	19.5%
County X Occupation X All workers X Year X Pharmacies	935	1,206	271	29.0%
County X Occupation X All workers X Year X Pharmacy chain	3,149	4,543	1,394	44.3%

Note: The table shows the average Herfindahl–Hirschman Index (HHI), projected HHI, and projected changes in the HHI induced by the merger. Averages are taken over the 31 treated counties—i.e., counties in which merging firms overlapped in 2011. Projected changes in the HHI in control counties are equal to zero, and thus are not included. Each line of the table presents the HHI calculations using a different labor market definition. For example, the first line defines a labor market for pharmacists as all employed pharmacists in 2011 within a county. In the second line, I restrict the market to workers in pharmacies. In the third line. I further restrict the sample to pharmacists in pharmacy chains. Pharmacy chains are defined as firms that have more than 5 establishments in 2011, nationwide. I also present calculations in which I define the labor market only for workers who were hired in 2011, i.e. newly hired workers. Lastly, I define the labor market for workers hired in the first semester of 2011 (line 11). I use workers who were employed on December 31 of 2011 in all calculations.

Table A.7: Sample size

Number of...	All		Relevant Counties		Balanced		
	Merging	Competitor	Merging	Competitor	Merging	Competitor	
	(1)	(2)	(3)	(4)	(6)	(7)	(8)
Counties	319	3,698	154	154	91	153	91
Firms	2	91,379	2	63,439	2	12,766	9,798
Establishments	1,821	114,761	1,404	77,187	392	14,751	11,212
Workers	92,267	1,228,711	91,505	1,225,704	51,380	632,906	497,658
Observations	269,877	4,361,448	261,827	4,029,495	134,216	2,015,316	1,586,235
Workers:							
Pharmacists	8,024	74,358	7,722	62,326	3,190	27,177	20,919
Salespeople	64,597	741,511	63,918	712,101	33,579	323,486	256,218
% of Observations:							
Pharmacists	16.3	11.0	15.9	9.4	13.2	8.5	8.3
Salespeople	61.8	53.7	62.8	54.0	66.0	53.3	53.4

Note: The table presents sample sizes. The first two columns do not restrict the sample. They include all observations in merging firms and their competitors from 2007 to 2018. Columns 3 and 4 restrict the sample to counties in which merging firms had an establishment in 2010. Column 6 and 7 restrict the sample to establishments that had at least one employee in every year between 2007 and 2018. Column 8 restricts the sample of competitors to the same 91 counties from the sample of merging firms. Source: RAIS 2007-2018.

Table A.8: Descriptive statistics on the retail pharmacy sector

	2007	2018	Annual growth (%)
Number of...			
Firms	26,187	30,094	1.3
Establishments	33,216	47,400	3.3
Employees	233,400	403,228	5.1
Pharmacists	39,905	78,977	6.4
Salespeople	121,897	216,055	5.3
Total Revenue (2018 R$)(billions)	59.9	153.7	8.9
Minimum wage (2021 R$)	807	1086	2.7
Avg. wage/min. wage:			
All workers	2.3	2.3	0.0
Pharmacists	4.6	4.2	-0.8
Salespeople	1.7	1.65	-0.3
% Completed High school	78	91	1.4
% Female	58	64	0.9

Note: Revenue data come from the Brazilian annual retail trade survey (PAC, Pesquisa Anual de Comércio). Minimum wage data are extracted from Ipeadata and deflated using the consumer price index. The rest of the statistics are based on the RAIS dataset and include the full count of establishments in the retail pharmacy sector that had at least one pharmacist.

Appendix B

Appendix Materials for Chapter 2

1. Appendix Tables

Table B.1: Recent Literature on Labor Market Concentration

Paper	Market Definition	Country
Azar et al. (2020)	Commuting zone X 6-digit occupation	US
Benmelech et al. (2022)	County X 4-digit industry Commuting zone X 4-digit industry	US
Prager and Schmitt (2021)	Commuting zone X Hospitals X Occupation Commuting zone X All health care X Occupation	US
Arnold (2021)	Commuting zone X 4-digit industry	US
Azar et al. (2022)	Commuting zone X 6-digit occupation	US
Berger et al. (2022)	Commuting zone X 3-digit industry	US
Brooks et al. (2021)	District X 4-digit industry State X 4-digit industry	India
Marinescu et al. (2021)	Commuting zone X 4-digit occupation X quarter	France
Guanziroli (2022)	County X Retail Pharmacies X Occupation	Brazil

Note: The table presents a selected sample of recent studies in the labor market concentration literature. Colum 2 describes the market definition explicitly or implicitly used in each study.

Table B.2: Recent Literature on the Effects of Minimum Wages

Paper	Market Definition	Country
Saltiel and Urzua (2022)	Microregion	Brazil
Dube et al. (2010)	County X Restaurants	US
Cengiz et al. (2019)	State X Demographic groups State X 1-digit industry	US
Dustmann et al. (2021)	District X Demographic groups	Germany
Azar et al. (2019)	County X 6-digit occupations	US

Note: The table presents a selected sample of recent studies in the labor market concentration literature. Colum 2 describes the market definition explicitly or implicitly used in each study.

Table B.3: Recent Literature on the Effects of Trade Shocks

Paper	Market Definition	Country
Felix (2022)	Microregion X 6-digit occupation	Brazil
Adão (2016)	Microregion X Schooling	Brazil
Kovak (2013)	Microregion	Brazil
Dix-Carneiro and Kovak (2017)	Microregion	Brazil
Autor et al. (2013)	Commuting zones X SIC codes	US
Hakobyan and McLaren (2016)	County X 2-digit industry	US
Caliendo et al. (2019)	States X 12 manufacturing sectors	US

Note: The table presents a selected sample of recent studies in the labor market concentration literature. Colum 2 describes the market definition explicitly or implicitly used in each study.

Table B.4: Sample size

	Raw data	Main Data
Region:		
States	27	27
Microregions	137	137
Mesoregions	558	553
Counties	5551	4246
Occupation:		
1-digit	10	10
2-digits	49	45
3-digits	193	185
4-digits	614	577
6-digits	2588	2026
Industry:		
2-digits	88	87
3-digits	285	274
5-digits	673	646

Note: The table presents the number of regions, occupations and industries in the raw dataset and in the main dataset used in the paper. RAIS 2007-2013.

Table B.5: Labor market concentration

Data-driven markets	# markets	# workers	% workers
HHI <=1000	6,033	14,598,459	86.70%
1000>HHI>=1800	3,466	527,595	3.10%
HHI>1800	45,356	1,702,170	10.10%

Note: The table categorizes markets in terms of concentration levels, as measured by the Herfindahl–Hirschman Index (HHI). The number of workers comes from RAIS 2014.

Appendix C

Appendix Materials for Chapter 3

1. Matching the College Graduates Sample with RAIS

In this section, we describe the procedure to match the college graduates' sample with the employer-employee dataset (RAIS) and clean the data. We first match the college graduate sample with the RAIS using student/worker's full names. This procedure leads to multiple matches. Secondly, we use a machine learning algorithm to select the best match. Third, we impose some sample restrictions.

The raw college graduates sample includes information of 1,217,440 students that graduated from 42 universities. After removing special characters, we are able to match 74% of these students to one or more workers with the exact same full name in the RAIS dataset. As a result, 2,093,069 workers are matched to 906,420 students. Out of these students, 78% are matched to a single worker and 22% are matched to at most 20 workers with the exact same name. We drop students with very common names that are matched to more than 20 workers.

Among students with multiple matches, we select the best match using a machine learning procedure. One university provided us with students' identification number such that, for this university, we can match students and workers by both name and identification number. We proceed by matching students by name, and, for a training sample, we estimate a model that uses student and worker's characteristics to predict whether the match is correct—as defined by the match using the identification number.[1]

We estimate two different models—a logit regression, and a random forest model—using the training sample. Using the model's estimates, we calculate a score for each match. We define the correct match based on 3 rules: (i) The match has the highest score of all matches; (ii) The score of the match is sufficiently large (greater than 5%); and (iii) The score of the top match is sufficiently large relative to the second-best match (the ratio of scores is greater than 1.1).

Appendix Table C.3 compares the results of each model using two metrics: the positive predictive value (PPV or accuracy) and the true positive rate (TPR or efficiency). While the random forest model has better PPV and TPR rates in the training sample (96.6% and 97.2%, respectively), these numbers do not produce better results in the test sample. Therefore, we decide to use the logit approach due to the consistency of training sample and test sample metrics (PPV of 87.1% and 87.2%). Appendix Table C.4 presents the sample sizes used for the in sample PPV and TPR calculations.

The machine learning algorithm finds a match for 806,893 students/workers. We further restrict the sample to students that graduated between 2000 and 2017. The final dataset includes 17,455,296 employment observations from 545,478 students/workers, that graduated from 42 universities.

[1]We use the following characteristics that are specific to the student-worker match: five dummies for age at college admission (<17, >20, >25, >35, >45 years old) where age is determined by the worker variable; difference between worker's earliest year on the data with year of college admission; a dummy for whether the individual had a full time job during college; a dummy for whether the schooling variable at RAIS reports an educational achievement inferior to college after the year of graduation; the number of worker observations at RAIS; dummies for the maximum schooling from all worker's observations.

2. Appendix Figures and Tables

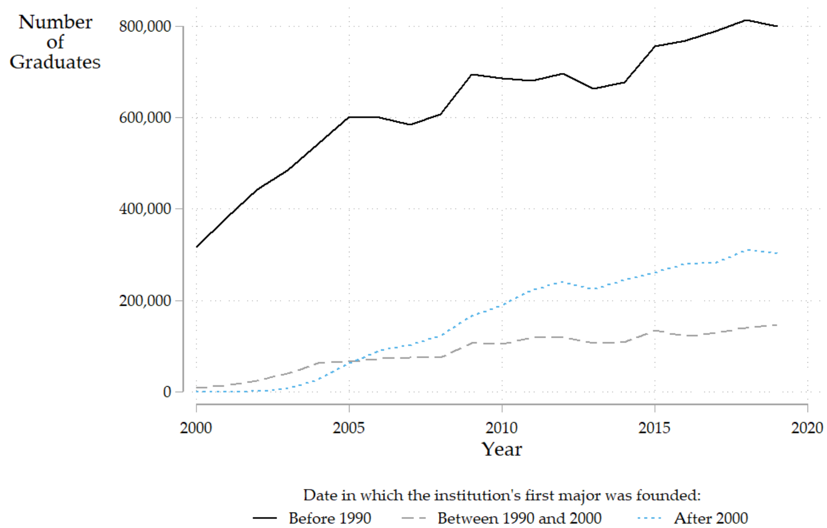


Figure C.1: Number of graduates by institution's date of foundation

Note: The figure presents the number of students that graduated in each year. Categories are defined by the year in which the institution's first major was founded. Source: Higher Education Census (2000-2018)

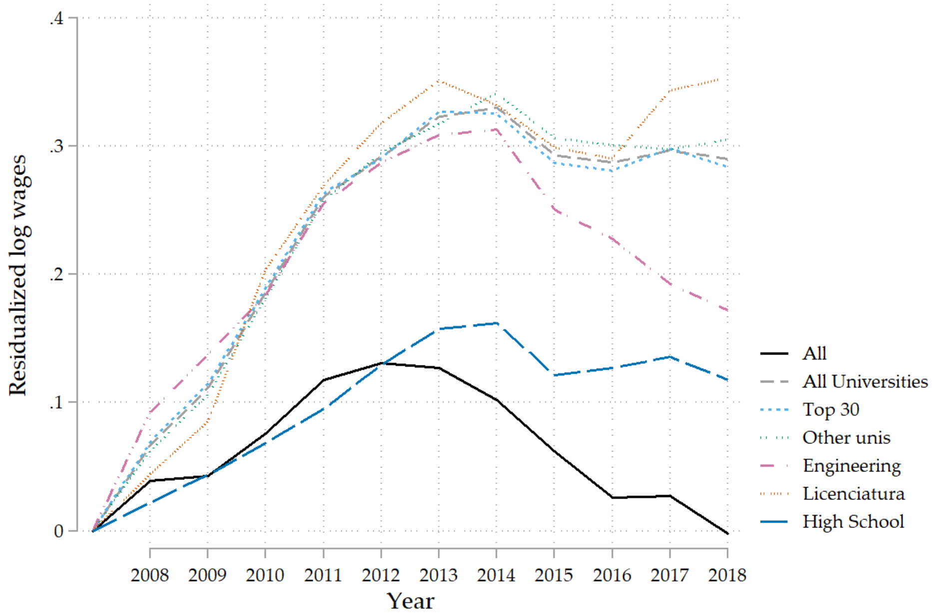


Figure C.2: Trends in residualized log wages

Note: The figure plots the evolution of earnings, relative to 2007 values, for different samples. Estimates come from the estimation of Equation 1 over a sample of 263,181,905 observations.

Table C.1: Schools with access to the data

Universidade/ Instituto	Acronym
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca	CEFET-RJ
Fundação Universidade do Amazonas	UFAM
Fundação Universidade Federal de Mato Grosso	UFMT
Fundação Universidade Federal de Ouro Preto	UFOP
Fundação Universidade Federal de Pelotas	UFPel
Fundação Universidade Federal de Rondônia	UNIR
Fundação Universidade Federal de Roraima	UFRR
Fundação Universidade Federal de São João Del Rei	FUNRei
Fundação Universidade Federal de Sergipe	UFS
Fundação Universidade Federal do ABC	UFABC
Fundação Universidade Federal do Acre	UFAC
Fundação Universidade Federal do Maranhão	UFMA
Fundação Universidade Federal do Tocantins	UFT
Fundação Universidade Federal do Vale do São Francisco	UNIVASF
Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro	IFRJ
Instituto Federal de Educação, Ciência e Tecnologia Fluminense	IFF

Table C.1: Schools with access to the data (cont.)

Universidade/ Instituto	Acronym
Instituto Militar de Engenharia	IME
Universidade de Brasilia	UNB
Universidade do Estado do Rio de Janeiro	UERJ
Universidade Estadual de Maringa	UEM
Universidade Estadual Paulista Júlio de Mesquita Filho	UNESP
Universidade Federal de Alagoas	UFAL
Universidade Federal de Alfenas	UNIFAL
Universidade Federal de Campina Grande	UFCG
Universidade Federal de Goiás	UFG
Universidade Federal de Itajubá	UNIFEI
Universidade Federal de Juiz de Fora	UFJF
Universidade Federal de Lavras	UFLA
Universidade Federal de Minas Gerais	UFMG
Universidade Federal de Santa Catarina	UFSC
Universidade Federal de Santa Maria	UFSM
Universidade Federal de Uberlândia	UFU
Universidade Federal de Vicosa	UFV
Universidade Federal do Ceará	UFC
Universidade Federal do Espírito Santo	UFES
Universidade Federal do Estado do Rio de Janeiro	UNIRIO
Universidade Federal do Pará	UFPA
Universidade Federal do Rio de Janeiro	UFRJ
Universidade Federal do Rio Grande do Norte	UFRN
Universidade Federal do Sul e Sudeste do Pará	UNIFESSPA
Universidade Federal Rural do Rio de Janeiro	UFRRJ
Universidade Tecnológica Federal do Paraná	UTFPR
Total	42

Table C.2: Sample size by university

University	Students	Matched to RAIS	% Matched
CEFET-RJ	30,446	23,029	75.60%
FUNRei	2,971	2,370	79.80%
IFF	16,387	10,682	65.20%
IFRJ	1,896	1,336	70.50%
IME	2,558	2,042	79.80%
UEM	53,160	40,605	76.40%
UERJ	72,562	56,330	77.60%
UFA	439	331	75.40%
UFAL	37,418	28,050	75.00%
UFABC	2,271	1,864	82.10%
UFAM	2,544	1,975	77.60%
UFC	65,155	51,206	78.60%
UFCG	32,647	23,678	72.50%
UFES	49,408	39,269	79.50%
UFG	46,335	36,233	78.20%
UFJF	31,459	23,471	74.60%
UFLA	13,231	9,480	71.60%
UFMA	1,851	1,511	81.60%
UFMG	99,656	75,839	76.10%
UFMT	45,547	35,156	77.20%
UFOP	27,024	21,423	79.30%
UFPA	47,025	35,051	74.50%
UFPel	24,189	16,107	66.60%
UFRJ	105,380	78,696	74.70%

Table C.2: Sample size by university (cont.)

University	Students	Matched to RAIS	% Matched
UFRN	81,761	50,166	61.40%
UFRR	325	259	79.70%
UFRRJ	13,395	10,003	74.70%
UFS	37,705	29,139	77.30%
UFSC	53,713	41,741	77.70%
UFSM	7,272	5,457	75.00%
UFT	1,537	1,101	71.60%
UFU	25,170	13,893	55.20%
UFV	26,383	18,231	69.10%
UNB	62,106	45,813	73.80%
UNESP	3,077	2,639	85.80%
UNIFAL	10,573	6,766	64.00%
UNIFEI	6,553	5,070	77.40%
UNIFESSPA	4,701	3,398	72.30%
UNIR	18,813	15,192	80.80%
UNIRIO	19,464	14,887	76.50%
UNIVASF	4,405	3,138	71.20%
UTFPR	28,928	23,793	82.20%
Total	1,217,440	906,420	74.50%

Table C.3: Comparing Matching Algorithms

Algorithm	Hyper Parameters		Algorithm Quality			
			Training sample (50%)		Test sample (50%)	
	b1	b2	PPV	TPR	PPV	TPR
Logit	0.1	1.25	87.10%	92.00%	87.20%	92.20%
	0.05	1	86.70%	92.50%	87.10%	93.00%
Random Forest	0.1	1.25	96.60%	97.20%	88.80%	90.30%
	0.05	1	96.00%	98.70%	86.70%	92.70%

Note: Hyper parameters b1 and b2 are the threshold for whether the match's score is sufficiently large and the threshold for whether the ratio between the best and the second-best scores is sufficiently large, respectively. Positive Predictive Value (PPV or Accuracy): number of true positives over total number of positives. True Positive Rate (TPR or Efficiency): number of true positives over total number of correct cases.

Table C.4: Confusion Matrix—Out of Sample Predictions

Algorithm Prediction	True Status		Total
	False	Correct	
Not Matched	13,613	691	14,304
Matched	1,362	8,900	10,262
Total	14,975	9,591	24,566

Note: The table presents the sample sizes from the logit model with b1=0.05 and b2=1.1. Positive Predictive Value (PPV or Accuracy): number of true positives over total number of positives = 8900/(8900+1362) = 86.7%. True Positive Rate (TPR or Efficiency): number of true positives over total number of correct cases = 8900/(8900+691)= 92.8%.

Bibliography

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Acemoglu, D. (2002, March). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature 40*(1), 7–72.

Acemoglu, D. and D. Autor (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4 of *Handbook of Labor Economics*, Chapter 12, pp. 1043–1171. Elsevier.

Adão, R. (2016). Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil. Unpublished.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018, January). Firms and the Decline in Earnings Inequality in Brazil. *American Economic Journal: Macroeconomics 10*(1), 149–89.

Arnold, D. (2021). Mergers and acquisitions, local labor market concentration, and worker outcomes. Unpublished.

Autor, D. H., D. Dorn, and G. H. Hanson (2013, October). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review 103*(6), 2121–68.

Azar, J., E. Huet-Vaughn, I. Marinescu, B. Taska, and T. von Wachter (2019). Minimum wage employment effects and labor market concentration. Working Paper 26101, National Bureau of Economic Research.

Azar, J., I. Marinescu, and M. Steinbaum (2022). Labor Market Concentration. *Journal of Human Resources 57*(S), 167–199.

Azar, J., I. Marinescu, M. Steinbaum, and B. Taska (2020). Concentration in us labor markets: Evidence from online vacancy data. *Labour Economics 66*(C).

Bank, W. (2019). *World Development Report 2019: The changing nature of work*. The World Bank Group.

Barros, R., M. de Carvalho, S. Franco, and R. Mendonça (2010). Markets, the state, and the dynamics of inequality in Brazil. In L. F. López-Calva and N. Lustig (Eds.), *Declining inequality in Latin America: A decade of progress?*, pp. 74–134. Brookings Institution Press.

Benmelech, E., N. K. Bergman, and H. Kim (2022). Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages? *Journal of Human Resources 57*(S), 200–250.

Berger, D., K. Herkenhoff, and S. Mongey (2022, April). Labor market power. *American Economic Review 112*(4), 1147–93.

Berry, S., M. Gaynor, and F. Scott Morton (2019, August). Do Increasing Markups Matter? Lessons from Empirical Industrial Organization. *Journal of Economic Perspectives 33*(3), 44–68.

Brooks, W. J., J. P. Kaboski, I. O. Kondo, Y. A. Li, and W. Qian (2021, September). Infrastructure Investment and Labor Monopsony Power. *IMF Economic Review 69*(3), 470–504.

Brown, C. and J. L. Medoff (1988). *The Impact of Firm Acquisitions on Labor*, pp. 9–32. University of Chicago Press.

Caliendo, L., M. Dvorkin, and F. Parro (2019, May). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica 87*(3), 741–835.

Camacho, A., J. Messina, and J. P. Uribe (2017, February). The Expansion of Higher Education in Colombia: Bad Students or Bad Programs? Documentos CEDE 015352, Universidad de los Andes - CEDE.

Card, D. (2022). Who set your wage? Working Paper 29683, National Bureau of Economic Research.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

Card, D., J. Heining, and P. Kline (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics 128*(3), 967–1015.

Card, D. and A. B. Krueger (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review 84*(4), 772–793.

Card, D. and T. Lemieux (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019, 05). The Effect of Minimum Wages on Low-Wage Jobs*. *The Quarterly Journal of Economics 134*(3), 1405–1454.

Ciccone, A. and G. Peri (2005). Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950-1990. *The Review of Economics and Statistics 87*(4), 652–663.

Dafny, L., M. Duggan, and S. Ramanarayanan (2012). Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry. *American Economic Review 102*(2), 1161–1185.

DePasquale, C. (2018). Hospital consolidation and the nurse labor market. Unpublished.

Dix-Carneiro, R. and B. K. Kovak (2017, October). Trade liberalization and regional dynamics. *American Economic Review 107*(10), 2908–46.

Dube, A., T. W. Lester, and M. Reich (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics 92*(4), 945–964.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, and P. vom Berge (2021, 08). Reallocation Effects of the Minimum Wage*. *The Quarterly Journal of Economics 137*(1), 267–328.

Elsby, M., B. Hobijn, and A. Şahin (2013). The decline of the U.S. labor share. *Brookings Papers on Economic Activity*, 1–52.

Falch, T. (2010). The elasticity of labor supply at the establishment level. *Journal of Labor Economics 28*(2), 237–266.

Felix, M. (2022). Trade, labor market concentration, and wages. Unpublished.

Fernandez, M. S. and J. Messina (2018). Skill premium, labor supply, and changes in the structure of wages in Latin America. *Journal of Development Economics 135*(C), 555–573.

Ferreira, F., S. Firpo, and J. Messina (2017). Ageing Poorly? Accounting for the Decline in Earnings Inequality in Brazil, 1995-2012. IZA Discussion Papers 10656, Institute of Labor Economics (IZA).

Fortunato, S. and M. Barthélemy (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences 104*(1), 36–41.

Freeman, R. B. (1976). *The Overeducated American.* San Diego: Academic Press.

Gehrke, B., E. Maug, S. Obernberger, and C. Schneider (2021). Post-merger restructuring of the labor force. IZA Discussion Paper No. 14409.

Gerard, F., L. Lagos, E. Severnini, and D. Card (2021). Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil. *American Economic Review 111*(10), 3418–57.

Giroldo, R. and B. Hollenbeck (2021). Concentration, Retail Markups, and Countervailing Power: Evidence from Retail Lotteries. MPRA Paper 109039, University Library of Munich, Germany.

Goldin, C. and L. Katz (2008). *The Race between Education and Technology.* Harvard University Press, Cambridge.

Guanziroli, T. (2022). Does labor market concentration decrease wages? evidence from a retail pharmacy merger. Unpublished.

Haanwinckel, D. (2020). Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution. Unpublished.

Hakobyan, S. and J. McLaren (2016, 10). Looking for Local Labor Market Effects of NAFTA. *The Review of Economics and Statistics 98*(4), 728–741.

He, A. and D. L. Maire (2020). Mergers and Managers: Manager-specific Wage Premiums and Rent Extraction in M&As. Unpublished.

Jarosch, G., J. S. Nimczik, and I. Sorkin (2019). Granular search, market structure, and wages. Working Paper 26239, National Bureau of Economic Research.

Kambourov, G. and I. Manovskii (2009). Occupational specificity of human capital. *International Economic Review 50*(1), 63–115.

Karabarbounis, L. and B. Neiman (2013). The Global Decline of the Labor Share. *Quarterly Journal of Economics 129*(1), 61–103.

Katz, L. and K. M. Murphy (1992). Changes in Relative Wages, 1963–1987: Supply and Demand Factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kovak, B. K. (2013, August). Regional effects of trade reform: What is the correct measure of liberalization? *American Economic Review 103*(5), 1960–76.

Krueger, A. B. and O. Ashenfelter (2018). Theory and evidence on employer collusion in the franchise sector. Working Paper 24831, National Bureau of Economic Research.

Lagaras, S. (2020). M&As, Employee Costs, and Labor Reallocation. Unpublished.

Lamadon, T., M. Mogstad, and B. Setzler (2021). Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market. *American Economic Review*, forthcoming.

Li, X. (2012). Workers, Unions, and Takeovers. *Journal of Labor Research 33*(4), 443–460.

Manning, A. (2011). Imperfect Competition in the Labor Market. *Handbook of Labor Economics 4B*, 973–1041.

Manning, A. and B. Petrongolo (2017). How Local Are Labor Markets? Evidence from a Spatial Job Search Model. *American Economic Review 107*(10), 2877–2907.

Marinescu, I., I. Ouss, and L.-D. Pape (2021). Wages, hires, and labor market concentration. *Journal of Economic Behavior & Organization 184*(C), 506–605.

Matsudaira, J. D. (2014). Monopsony in the Low-Wage Labor Market? Evidence from Minimum Nurse Staffing Regulations. *Review of Economics and Statistics 96*(1), 92–102.

McGuckin, R. H. and S. Nguyen (2001). The impact of ownership changes: A view from labor markets. *International Journal of Industrial Organization 19*(5), 739–762.

Menezes-Filho, N. A., H. Zylberstajn, J. P. Chahad, and E. Pazello (2011, September). Unions and the Economic Performance of Brazilian Establishments. IDB Publications (Working Papers) 3280, Inter-American Development Bank.

Messina, J. and J. Silva (2017). *Wage inequality in Latin America: Understanding the past to prepare for the future*. The World Bank Group.

Morlacco, M. (2020). Market Power in Input Markets: Theory and Evidence from French Manufacturing. Unpublished.

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics 13*(4), 653–77.

Nevo, A. and M. D. Whinston (2010, June). Taking the dogma out of econometrics: Structural modeling and credible inference. *Journal of Economic Perspectives 24*(2), 69–82.

Nimczik, J. S. (2020). Job Mobility Networks and Data-driven Labor Markets. Unpublished.

Parent, D. (2000). Industry-Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics. *Journal of Labor Economics 18*(2), 306–323.

Piketty, T. and G. Zucman (2014). Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010 . *Quarterly Journal of Economics 129*(3), 1255–1310.

Poletaev, M. and C. Robinson (2008). Human capital specificity: Evidence from the dictionary of occupational titles and displaced worker surveys, 1984–2000. *Journal of Labor Economics 26*(3), 387–420.

Prager, E. and M. Schmitt (2021). Employer consolidation and wages: Evidence from hospitals. *American Economic Review 111*(2), 397–427.

Qiu, Y. and A. J. Sojourner (2019). Labor-market concentration and labor compensation. IZA Discussion Papers 12089.

Rinz, K. (2020). Labor market concentration, earnings, and inequality. *Journal of Human Resources*, 0219–10025R1.

Rodriguez, J., S. Urzua, and L. Reyes (2016). Heterogeneous Economic Returns to Post-Secondary Degrees: Evidence from Chile. *Journal of Human Resources 51*(2), 416–460.

Rubens, M. (2021). Market structure, oligopsony power, and productivity. Unpublished.

Saltiel, F. A. and S. Urzua (2022). Does an Increasing Minimum Wage Reduce Formal Sector Employment? Evidence from Brazil. *Economic Development and Cultural Change*, forthcoming.

Schmutte, I. (2014). Free to move? a network analytic approach for learning the limits to job mobility. *Labour Economics 29*(C), 49–61.

Schubert, G., A. Stansbury, and B. Taska (2021). Employer concentration and outside options. Unpublished.

Sebrae-SP (2015). Como abrir e fazer a gestão de farmácias e drogarias. Technical report.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2018). Firming Up Inequality. *Quarterly Journal of Economics 134*(1), 1–50.

Staiger, D. O., J. Spetz, and C. Phibbs (2010). Is There Monopsony in the Labor Market? Evidence from a Natural Experiment. *Journal of Labor Economics 28*(2), 211–236.

Starr, E. P., J. Prescott, and N. D. Bishara (2021). Noncompete Agreements in the US Labor Force. *Journal of Law and Economics 64*(1), 53–84.

Sullivan, P. (2010). Empirical evidence on occupation and industry specific human capital. *Labour Economics 17*(3), 567–580.

Todd, K. and J. Heining (2020). The Labor Market Impacts of Employer Consolidation: Evidence from Germany. IAB Discussion Paper 202020.

Traag, V., L. Waltman, and N. V. Eck (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep 9*(5233).

Traag, V. A., P. Van Dooren, and Y. Nesterov (2011, Jul). Narrow scope for resolution-limit-free community detection. *Phys. Rev. E 84*, 016114.