UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title
Labor Market Structure and Policy Evaluations

Permalink
https://escholarship.org/uc/item/6xb6926f

Author
Liu, Jianan

Publication Date
2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

LABOR MARKET STRUCTURE AND POLICY EVALUATIONS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Jianan Liu

June 2022

The Dissertation of Jianan Liu
is approved:

Professor Laura Giuliano, Chair

Professor George Bulman

Professor Ajay Shenoy

Peter Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by

Jianan Liu

2022

Table of Contents

List of Figures

List of Tables

Abstract

Labor Market Structure and Policy Evaluations

by

Jianan Liu

My dissertation studies labor market structure and evaluates effects of policies relating to transportation improvements and immigration enforcement in different countries including Germany, the Untied States, and China. Specifically, I examine the nature and extent of labor market frictions, explore the causes and consequences, as well as study the role in policy evaluations.

In the first chapter, I investigate one potential source of labor market frictions from limited labor mobility by examining how integrating labor markets through improved transportation infrastructure affects both wages and the allocation of workers across establishments. I take advantage of the expansion of the High-Speed Rail in Germany, which connected medium-sized districts located on existing rail lines, providing a natural experiment to study the effects of labor market integration. Using administrative panel data on establishments and workers linked to their employers, I estimate difference-in-differences and event-study models that compare newly connected districts to matched controls that were never connected. In theory, policies that improve labor mobility might raise wages both by facilitating more productive matches of workers to firms and by reducing the monopsony power that employers have vis-à-vis workers. I find evidence of increased labor mobility for workers in treated districts, especially those

that are more likely to commute. Worker wages increase significantly and I find that both reduced monopsony power and better match quality are possible mechanisms. I also test for establishment adjustments including entry, exit, size, and wage that are predicted by each mechanism. This study shows evidence of reduced labor market frictions from improved transportation. It also sheds light on the importance of policies that enhance workers' ability to switch employers.

In the second chapter, I explore the effects of LAWA on labor markets based on evidence that LAWA has significantly reduced the population of Hispanic noncitizens in Arizona. Specifically, I focus on the composition of the labor force and industry heterogeneity. I first show that a synthetic control has similar concurrent economic trends with Arizona, and verify that changes in employee composition are due to the replacement of Hispanic noncitizen workers by other subgroups and not by a change in overall employment. In response to LAWA, firms tend to reduce both the new hire rate and the separation rate. Several robustness checks are conducted to test the accuracy of the estimates and several mechanisms are considered that may drive the results.

In the third chapter, we study the structure of labor markets and the effects on wages in China. A growing literature has emphasized the existence of monopsony power stemming from employer concentration within local labor markets, which deviates from the conventional view of labor markets as perfectly competitive. We use firm-level data from the Chinese Annual Survey of Industrial Firms to analyze how employer concentration affects wage behavior. We first verify that local employment concentration measured by the Herfindahl-Hirschman Index of firm employment decreases between

1998 and 2013, in contrast to the increasing trend found in developed countries. Then with OLS and IV models, we show a negative relationship between labor market concentration and wages both the at the market level and at the firm level, indicating the existence of imperfect labor market competition in China.

to myself and my family

Acknowledgments

I wish to express my sincere gratitude to the people who have provided me with great help and support in the completion of my dissertation and the pursuit of my Doctoral Degree.

First and foremost, I am deeply indebtd to my advisors, Professor Laura Giuliano and Professor George Bulman, for their continuous guidance and invaluable patience. Laura has great insights and passion towards research, which inspired me a lot in my research ideas. She provided great support in my data access and spent huge efforts advising my research progress. George supervised me from starting the first research project to writing the final dissertation and gave precious feedbacks along the way. His knowledge, expertise, and enthusiasm encouraged me a lot in every step of my Ph.D. studies. Both of them not only are the mentor to my academic work, but also generously offered kind support in many aspects of my life. I am also very grateful to my other dissertation commitee member, Professor Ajay Shenoy, for his insightful discussions and suggestions. Ajay also offered an opportunity for the research assistant position, from which I learned a lot in research methods and skills.

Furthermore, I also thank my other two oral committee members Professor Justin Marion and Professor Jesse Cunha, as well as all the other workshop participants for their helpful discussions and feedbacks in my research work. Thanks also go to the professors who taught core and field courses in the department. I am extremely grateful to our Ph.D. program directors Professr Jonathan Robinson and Professor Natalia Lazzati for their hard work and great support. Additionally, I hope to express special thanks to our program corodinator Sandra Reebie for the assistance and dedicated involvement during this process. I would also like to acknowledge research staff at German Institute for Employment Research and John Rakuten in the Office of Sponsored Projects of UCSC for help with the data access, and Camille Fernandez in the Center for Labor Economics at UC Berkeley for support with on-site use of the data.

This endeavor would not have been possible without companionship of classmates at UCSC and encouragement of friends. I want to thank my lovely cohorts for help and encouragement at school and joyful gathering moments out of school. I also thank colleagues in other years of the program for the enlightenment and delight they bring. I am also thankful for my friends in and out of the United States for emotional support.

Last but not least, words cannot express my gratitude to my beloved family. I want to thank my parents for their constant support and unconditional love all the time. Thanks my husband, Haoran, for stickying by my side, bearing with me the hardships, and sharing the happiness. I could not have undertaken this journey without you. Thanks my baby, Qiaoyi, for coming to me during the hardest time and brightening my life.

Six years of Ph.D. is a long journey with risks, uncertainties, pains, and also gains. It is also a regretless journey giving me courage and faith that I will carry forward through my career. I look forward to the next chapter of my life.

Chapter 1

The Effects of Labor Market Integration: Evidence from Germany's High-Speed Rail

1.1 Introduction

There is growing interest in the role of labor market frictions that limit workers' ability to switch employers in wage determination (Manning, 2003, 2011). Empirical studies have pointed to labor market frictions as likely explanations for the finding that wages often vary across firms even for the same workers (Card, Heining, and Kline, 2013); and such frictions are increasingly viewed as contributors to the growth in spatial inequality (Hirsch et al., 2020), the persistence of local demand shocks (Autor, Dorn, and Hanson, 2013), and the falling labor share of income[1]. The theoretical literature has identified potential sources of labor market frictions including heterogeneous preferences (Bhaskar et al., 2002), moving costs (Boal and Ransom, 1997) and lack of information (Mortensen and Pissarides, 1999). In theory, the frictions might also stem from limited labor mobility due to high commuting costs (Manning and Petrongolo, 2017). However, the empirical literature has provided limited evidence on the potential for policies to alleviate these frictions. This paper examines the effects of improved transportation infrastructure on wages and and the allocation of workers across establishments using detailed administrative data on establishments and workers linked to their employers and the expansion of Germany's High-Speed Rail (HSR).

With the availability of detailed data on workers and firms, recent studies built on the additive worker and firm effects wage model of Abowd, Kramarz, and Margolis

[1]As mentioned in Dao et al. (2017), the amount of GDP paid out in wages, salaries, and benefits has been declining in developed and, to a lesser extent, emerging economies since the 1980s. In a large sample of 35 advanced economies, it fell on average from around 54 percent in 1980 to 50.5 percent in 2014, a loss of 3.5 percentage points or about 6.5 percent.

(1999) (AKM) have examined the role of firm wage premiums and assortative matching on wage dynamics in the United States (Song et al., 2019), Germany (Card, Heining, and Kline, 2013), Portugal (Card et al., 2016), Brazil (Lavetti and Schmutte, 2016), and others. The existence of labor market frictions affect both the monopsony power of firms and the quality of match of workers to firms.

Urban transportation improvements that reduce commuting costs between markets can increase labor mobility, and hence integrate neighboring labor markets. The key hypothesis is that it will reduce labor market frictions and raise wages through two primary mechanisms: more productive matches of workers to firms and reduced monoposony power due to increased labor market competition. Predictions under the first mechanism include wage increases for workers who change jobs, but not for stayers, and increased productivity within firms. Evidence such as convergence of wages across employers and locations, higher wage growth for stayers, and greater turnover of employers are consistent with the second mechanism. However, we should also note that reduced commuting costs could lead to lower wages because of compensating differentials.[2]

To test the above hypotheses and shed light on potential mechanisms, I exploit quasi-random variation in commuting costs due to the expansion of Germany's High-Speed Rail (HSR). The HSR in Germany has several advantageous features for studying labor market integration. First, it is used for passenger transportation only, mitigating

[2]For example, if workers have a strong preference for living in large cities, then firms in less populous districts that are not connected might need to pay a large compensating differential to attract such workers. Getting connected might allow small-city firms to attract workers for lower wages because the workers can commute more easily while still living in big cities.

concerns about confounding effects on multiple (e.g., labor and product) markets. Second, while the original wave of the HSR started in 1991 and connected major districts in the country, the expansion wave started in 1999 and added smaller districts located en route between cities, causing the location of the expansion wave stations to be plausibly exogenous to the labor market outcomes I study. Third, as shown by Heuermann and Schmieder (2019), labor mobility in connected districts increased significantly.

Exploiting this natural experiment, I examine how a district's connection to the HSR affects a range of labor market outcomes measured at the level of the district, establishment and worker. Specifically, I implement a difference-in-differences design comparing districts that are connected in the expansion wave of the HSR to surrounding districts that were never connected. I use an event-study approach to account for treatment effect dynamics and to verify that trends in the treated and control districts are parallel prior to the connection. I also test for treatment effect heterogeneity to check whether impacts are driven by those workers who are most likely to commute between districts and by the establishments most likely to employ such workers.

I use two primary data sources for the analyses: historical HSR schedules, and employment data for establishments and workers from the Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute Employment Research (IAB). First, I collect HSR data that reveals when districts were connected and how the lines were built. I use this data to construct both a binary treatment variable that indicates whether or not the district is connected to the HSR network, and measures of treatment intensity that vary with the relative size and wages of the districts with which the treated district becomes connected. Second, I take advantage of two data sets from the IAB. One is a 50% sample of all establishments throughout Germany which provides information on establishment level number of workers, mean and median wages, location, and industry. It also identifies establishment entry and exit and worker inflows and outflows. The other is a linked employer-employee data set for a representative sample of establishments and all workers employed there for at least one day, as well as the complete employment histories of these workers. This data includes information on worker characteristics such as gender, age, education, daily wage and residence.

The study first reveals that HSR connections lead to an increased number of establishments in treated districts relative to control districts primarily due to new entrants. Further, exiting establishments tend to be smaller in treated districts, which is consistent with increased competition as a mechanism. Using the average wage and skill reliance of each industry, I group establishments into two categories: low-type industries and high-type industries. I find that within treated districts, there is an increasing number of establishments that belong to the high-type industries while there is little change for the low-type industries. That is, treated districts become more specialized in industries with a higher wage and a greater reliance on highly educated and skilled workers.

Establishments in connected districts grow both in size and wage. The average size of establishments within districts increases by 2.51%, in part due to compositional changes. Median wages increase by 0.515% within districts and by 0.393% within establishments, suggesting that the net effect is primarily driven by existing establishments. Heterogeneity analysis at the establishment level shows smaller establishments are more likely to grow in size while larger ones are more likely to grow in median wage. Heterogeneity analysis at the district level suggests that treated districts that are larger or pay higher wages or are connected with districts that pay lower wages are more likely to attract workers.

With worker-level data, I estimate the characteristics of an establishment's workforce, and I distinguish between changes due to new hires and separations. The results indicate that establishments hire from further away after an HSR station opens in their district. This effect is driven largely by male and high-educated workers, groups who are more likely and able to commute. I also estimate wages at the worker level and distinguish between effects for movers and stayers. I find that daily wages within workers increase and that this effect is larger and more significant for high-educated workers. I also find that the wage growth is driven both by increased competition and improved match quality.

The remainder of the paper is organized as follows. Section 2 provides a summary of the related literature. Section 3 introduces the background of the HSR in Germany. Section 4 discusses the data and summary statistics of the variables. Section 5 presents the methodology. Section 6 presents the results for establishment and worker level analyses. Section 7 concludes.

1.2 Related Literature

This paper contributes to the literature on nature and extent of labor market frictions. Researchers estimate that firm-level labor supply elasticities are low, far from infinity as indicated by perfect competition (Webber, 2015; Dube et al., 2020; Bassier et al., 2020). Expanding the conventional explanation of the urban wage premium with agglomeration effects (higher worker productivity in thicker labor markets), Hirsch et al. (2020) argue that a substantial part of the urban premium stems from imperfect competition, as firms in denser labor market face greater competition and have less wage-setting power. Relative to product market concentration and competition, the effects of labor market competition has received little attention. Recent work such as Azar et al. (2020) and Benmelech et al. (2020) consider the effect of labor market concentration on wages. Azar et al. (2020) explore an online job board in the U.S. and measure labor market concentration with the Herfindahl-Hirschman Index (HHI). They define labor markets by a combination of occupation and commuting zones and find higher concentrations are associated with lower posted wages. Benmelech et al. (2020) find similar results using census data from the U.S., and they also notice that the negative relationship between labor market concentration and wages is stronger when unionization rates are low. This paper contributes to this literature by taking advantage of the shock that increase competition through reducing commuting costs to study labor outcomes. It adds empirical evidence in consistent with the theoretical fact that limited labor mobility is a cause for lack of competition.

This paper is also relevant to the growing literature on decomposition of wage structures. Building on AKM's additive worker and firm effects wage model, Card, Heining, and Kline (2013) (CHK) study trends in the dispersion of firm wage premiums and its role in the rise of wage inequality in West Germany. They separately identify the impact of rising heterogeneity in pay across workers and rising heterogeneity in pay received by the same worker across employers to decompose changes in the structure of wages. They find that increasing heterogeneity in establishment wage premiums and rising assortative matching in the assignment of workers to establishments explain a large share of the rise in wage inequality. As for potential sources of firm wage premiums, Hirsch and Mueller (2020) investigate the influence of industrial relations on firm effects from CHK's estimates. They find that wage premiums are larger in firms bound by collective bargaining agreements and with a workers council, and decreasing bargaining coverage contributes to the rise in the premium dispersion. Goldschmidt and Schmieder (2017) find that domestic outsourcing in Germany lead to more dispersed firm wage premiums that, in turn, explain 9% of the rise in wage inequality. Song et al. (2019) investigate wage inequality in the U.S. and confirm the role of firms in the rising earnings gap. However, they find that the increasing between-firm variance is not accounted for by firm wage premiums but by a widening gap between firms in the composition of their workers, which can be split into increasing sorting (high-wage workers become increasingly likely to work in high-wage firms), and increasing segregation (high-wage workers become increasingly likely to work with each other). This paper examines establishment adjustments and worker reallocation in response to labor market integration. It also identifies mechanisms of the wage growth for stayers from both reduced monopsony power and better match quality.

This paper also contributes to the literature on the impacts of transportation improvements. This rich literature has been mainly focused on its effects on aggregate welfare gains and interregional trade (Tsivanidis, 2018; Donaldson, 2018; Allen and Arkolakis, 2019; Banerjee et al., 2020); firm production and organization (Firth, 2017; Charnoz et al., 2018; Bernard et al., 2019; Gumpert et al., 2019) and knowledge diffusion and technology adoption (Dong et al., 2020; Lin et al., 2021). They find significant and heterogeneous improvements in welfare, improvements in management of business organizations and firm performance, and greater idea spillovers among the high-skilled teamwork. Using the same natural experiment as this study, Gumpert et al. (2019) look at the effects of German HSR expansion on establishments that are part of multi-establishment firms. They find that plants grow more quickly when commute time to the headquarters is reduced by the HSR. Using data on worker characteristics, studies examine the effect of transportation on labor mobility and outside options as well as differential gains by skill and gender groups (Heuermann and Schmieder, 2019; Caldwell and Danieli, 2020). Heuermann and Schmieder (2019) examine the causal effect of reductions in commuting time between regions on the commuting decisions of workers and their choices of where to live and where to work. They find the access to HSR increases labor mobility and find evidence that more workers living in bigger cities commute to work in smaller cities. This paper adds on the effects of improved transportation on establishments and workers through integrating labor markets, and emphasizes its role in alleviating labor market frictions.

Lastly, this study is also related to recent studies of the role of local labor markets. For example, Autor et al. (2013) study the effects of import competition from China across US local labor markets, exploiting cross-market variation in import exposure stemming from initial differences in industry specialization. To properly define a labor market, researchers also deal with the localization of economic activities, modifying the size and shape of the local labor market. Manning and Petrongolo (2017) propose a spatial job search model accounting for overlaps and interdependencies of labor markets. Nimczik (2017) introduces a novel method to identify endogenous labor markets which are revealed by job mobility flows rather than pre-determined administrative boundaries. Both studies indicate that the evaluation of local shocks should take into consideration the labor mobility. In application of these ideas, this paper uses the geographical boundaries to define labor markets, and considers commuting costs when determining the size of them.

1.3 Background

Many urban areas have experienced significant development in constructing transportation infrastructure. Countries such as Japan, Spain, France and China have invested substantially in their High-Speed trains, moving passengers across regions at speeds of 200 km/h or more. The High-Speed Rail in Germany, also called the Intercity-Express (ICE), is tightly integrated with pre-existing lines and trains. It was first introduced in 1991 with the first trains operating between Hamburg and Munich and has been expanded in two major waves. The original wave from 1991 to 1998 connected 46 cities such as Berlin and Hamburg. The expansion wave from 1999 to 2010 connected new stations in small- and medium-sized cities along pre-existing routes. The planning process was a joint effect of the Federal Government, the national railway company Deutsche Bahn AG, and states, with the goal of upgrading existing lines by including smaller cities located between larger metropolitan areas.[3] As a result 34 "lucky" cities were connected during this period and the location choice could be regarded as random. Figure 1.1 shows the map of Germany with HSR connections. The green regions are the districts that were connected during the original wave and the red regions were those connected during the expansion wave, and the white regions are those with no connection. The 34 red districts form the treatment group and the control is selected from the white districts. Table 1.A1 lists the 34 districts that are connected in the expansion wave. Table 1.A2 provides the average population and GDP for the three types of districts. The districts that were connected in the original wave have more than twice of the average population of districts that were connected in the expansion wave, and nearly three times the GDP. The districts with no connection are only slightly smaller in population and GDP than the districts in the expansion wave.

[3] Ahlfeldt and Feddersen (2018) provide a detailed description of the twists and turns that the political, legal and administrative process had gone through.

1.4 Data and Summary Statistics

1.4.1 Establishment History Panel

The Establishment History Panel 1975-2019 (BHP 7519) from the Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) is a 50% sample of all establishments throughout Germany. Each firm has at least one employee subject to social security as of 30 June of a given year.[4] The left panel of Table 1.A3 shows the sample size of the BHP data. The core data set comprises establishment information including the anonymized establishment ID, district, economic sector, and details of the employee and wage structure by gender, age, and education. The primary variables used for the analyses are the total employment size, median wage, district, and sector. The employment size is measured by the number of workers under the Employment History (BeH) that integrates notification procedures for health, pension and unemployment insurance, under which employers are required to submit notifications to the responsible social security agencies for employees covered by social security. Civil servants, self-employed and students are not recorded. The top left panel of Figure 1.A1 shows the distribution of the size of all establishments and the top right panel shows the distribution of larger establishments that had at least ten employees in 2004. The variable for the economic sector indicates the economic activity as a 3-digit code in accordance with the Classification of Economic Activities, edition

[4] As for the sampling procedure, a 50% random sample is drawn from all establishment IDs during 1975 to 2019, and once the ID is selected, the information for all the years of its appearance are included in the data.

1993 (WZ93),[5] and contains time-consistent information on the economic activity with the help of re-coding tables.[6] The extension files of the BHP also identify worker inflows and outflows as well as firm entry and exit with the worker flows approach,[7] with which it can identify the start-up of a new establishment or the closure of an establishment separate from an establishment being split off or outsourced, or an existing establishment being given a new ID.

1.4.2 The LIAB Longitudinal Model

The LIAB Longitudinal Model 1993-2010 (LIAB LM 9310) provides linked employer-employee data from the IAB.[8] It contains establishment data from the annual waves of the IAB Establishment Panel (IAB BP) and administrative records for employees who worked there at least one day during the sample period, with each employee linked to the whole employment biographies. The IAB BP is an annually conducted survey of establishments, the sample of which is drawn from the population of all German establishments with at least one employee eligible for social security as of June 30 of the survey year. The sample is stratified according to industry, firm size, and state, and is representative of the population of these establishments. The administrative records for individuals in LIAB come from the Integrated Employment Biographies (IEB) of the IAB, which comprise information on all individuals from two data sources. The first is

[5]The WZ93 is based on the Statistical Classification of Economic Activities in the European Community (NACE) Rev. 1 which has four levels the first two of which are based on the International Standard Industrial Classification (ISIC) Rev. 3.

[6]A detailed description of generating time-consistent codes for the classification of economic activities WZ93 can be found in Eberle et al. (2011).

[7]Detailed information about this procedure can be found in Hethey-Maier and Schmieder (2013).

[8]Klosterhuber et al. (2013) describes the data.

the Employment History (BeH) that has the same source as the BHP. The second is the Benefit Recipient History (LeH) that covers periods during which individuals receive earnings replacement benefits such as unemployment benefits. Hence individuals under this source is deleted for my analyses. The sample size is shown in the right panel of Table 1.A3. Variables include worker characteristics such as gender, age, education, district of residence and work, and daily wage. The bottom left panel in Figure 1.A1 shows the distribution of the weighted size of establishments and the right panel shows the distribution of the size without weighting. Using the information on education, I define a worker to be low-educated if he has a lower secondary, intermediate secondary or upper secondary school certificate without vocational qualifications and a worker is defined as high-educated if he has a upper secondary school certificate with a vocational qualification or a degree from a university. A dummy variable for commuter is created based on whether the worker works in the same district as he lives. Table 1.A4 provides a summary of establishment and worker characteristics in the LIAB data.

1.4.3 The HSR Data

The HSR data comes from the historical schedules of High-Speed trains.[9] It contains information on when the districts were connected by the HSR as well as how the lines were built. Using web scraping, I collect information on the HSR schedules from 1991 to 2010, from which I first obtain the year of connection for the stations in the expansion wave, and then based on the lines that the HSR operates, I identify

[9]https://www.fernbahn.de/datenbank/suche/

district pairs with direct connections between them. Figure 1.3 shows the timeline of the connections.

1.4.4 Eurostat

The geographical unit used for the analyses is the district at the NUTS-III level in Germany. In 2010, the country consisted of 412 districts in 16 states, a quarter of which are urban districts.[10] 32 cities are, for historical reasons, split up into a core city and the surrounding area and I merge these into single districts. From the Eurostat website, I collect information on the district level area, population and per capita GDP, as well as the distances between all district pairs. The commuting distance for commuters is defined as the distance between their district of residence and district of work and for non-commuters it is defined with the radius of their district of work, given the area and assuming the shape is a circle.

1.5 Methodology

This paper attempts to identify the effects of labor market integrating due to reducing commuting costs. There are two primary challenges of identifying the effects of transportation networks on labor markets in this context. First, the choice of train stations is usually not independent of local economic development, resulting in endogeneity issues. A richer district might be more likely to be connected to the

[10]The LIAB data uses the 2010 boundary and the BHP data uses the 2016 boundary that contains 401 districts.

network as it may have higher demand for transportation services. Alternatively, a less developed district might be more likely to get a station in order to promote development. To deal with such selection issues, the existing literature often relies on instrumental variables such as planned or historical lines (Banerjee et al. (2020)) and least cost path spanning tree networks (Faber (2014)). Second, transportation infrastructure affects both the labor market and the product market and ideally one would find an instrument that only affects the mobility of labor while leaving the mobility of goods unchanged. To address these challenges, I exploit the expansion of High-Speed Rail in Germany. The HSR was first introduced from 1991 to 1998 and connected 46 major cities. An expansion wave from 1999 to 2010 connected new stations in smaller cities along pre-existing routes. So 34 “lucky” cities were connected during this period with the location choice determined by convenience rather than economic promise. Further, the HSR is used only for the transportation of people and not goods. Heuermann and Schmieder (2019) show that labor mobility increases significantly in districts that were connected during the expansion wave of the HSR. One important pattern is that young and well-educated workers with a preference for urban life commute to work in the periphery. Thus, this context provides an opportunity to study the effects of the integration of labor markets in districts that become connected with large urban labor markets.

1.5.1 Empirical Design: Establishment Level Data

With the staggered adoption of the HSR stations during the expansion wave in Germany as a natural experiment, the empirical strategy is a generalized difference-in-differences design with variation in treatment timing[11]:

$$Y_{fdt} = \alpha_f + \gamma_t + \beta_1 HSR_{dt} + \epsilon_{fdt} \tag{1.1}$$

HSR_{dt} is a dummy variable indicating whether a district d is connected by the HSR in year t. Y_{fdt} indicates the outcome variables for establishment f. Each specification includes either district fixed effects or establishment fixed effects and year fixed effects. Standard errors are clustered at the district level. I also conduct analyses at the district level (e.g. establishment entry and exit) with a slightly different version of Equation (1) where the outcome are changes at the district level, and district fixed effects are included. In Figure 1.2, the treated districts are the red regions that are connected during the expansion wave of the HSR, and the control districts are the blue regions that are selected based on geographical proximity. Specifically, blue districts are those that near the red districts (within 100km), excluding those that are too close to the green districts (within 20km).[12] The assumption here is that treated districts are similar to their neighbors in the levels and trends of the outcome variables prior to the expansion of the HSR network. With the above criteria, 152 districts form the control group, with their weights based on the number of districts in the control pool matched to each of the 34 treated districts.[13]

To examine the validity of the design and show the dynamic effects of treat-

[11] The timing of the connections is shown in Figure 1.3.

[12] The districts that are too close to the green districts are excluded since establishments in those districts can easily get access to the workforce in the nearby metropolitan areas.

[13] For example, if 10 districts are within 100km to Berlin, then each of them will be assigned a weight that is equal to 0.1.

ment, I first implement an event studies approach using pre- and post- expansion years:

$$Y_{fdt} = \alpha_f + \gamma_t + \sum_{j=-2}^{4} \beta_j HSR_{dt}^{j} + \epsilon_{fdt} \tag{1.2}$$

HSR_{dt}^{j} is a dummy variable and equals to 1 if it is j years after district d was connected with the HSR. I first aggregate the BHP data to the district level and focus on the total number of establishments as well as the number of establishments that enter and exit, where the outcome variables in Equation (2) are the log number of total establishments in the district and the log of establishments that enter and exit. I also consider industry specialization within districts with the event studies. For the establishment level analyses, I restrict my sample to establishments that have at least 10 employees before the treatment. The outcomes of interest are establishment size, measured by the log number of workers, and establishment wages measured by the log of median wage. I also examine employee inflows and outflows for each establishment, measured by the log number of workers that enter or exit. With the same sample (establishments that have more than 10 employees before the treatment) and the same period (two years before the treatment to four years after the treatment), I estimate the average treatment effects for establishment size and wage with Equation (1).

Next, I conduct analyses to shed light on whether the effects are driven by certain types of establishments. First, I examine the characteristics of establishments that enter and exit:

$$Y_{fdt} = \alpha_d + \gamma_t + \beta_1 HSR_{dt} + \beta_2 S_{fdt} + \beta_3 HSR_{dt} * S_{fdt} + \epsilon_{fdt} \tag{1.3}$$

S_{fdt} is a dummy variable indicating the entry or exit status of the establishment. Y_{fdt}

represents establishment size and wage. For example, to answer whether small or low-wage establishments are more likely to exit (when increasing competition among establishments), S_{fdt} indicates whether establishment f exits district d (i.e. exits the labor market) in year t. β_3 reveals the effects of the HSR on exit across smaller and larger establishments and those with higher and lower wages. This approach is also used to evaluate the characteristics of establishments that enter.

Second, I examine heterogeneity of effects on size and wage by establishment characteristics at baseline:

$$Y_{fdt} = \alpha_f + \gamma_t + \beta_1 HSR_{dt} + \beta_2 HSR_{dt} * Est_f + \epsilon_{fdt} \tag{1.4}$$

Est_f represents establishment size and wage before the treatment. Y_{fdt} represents establishment size and wage in year t. Heterogeneity by establishment size is theoretically ambiguous. On the one hand, larger establishments have greater latitude to set wages, but, on the other hand, they may have greater ability to increase wages when facing competition. The interaction term reveals whether larger or smaller establishments are more likely to grow in size and to increase wages, as well as whether those that pay higher or lower wages at baseline are more likely to grow in size and wage.

In order to shed light on the mechanisms of the above effects, and to differentiate the effects by treatment intensity, I consider the characteristics of the treated district. For example, there may be larger or smaller effects in more populous districts that are connected during the expansion wave. Similarly, effects may differ across districts with higher or lower wages in the baseline period. This can be examined with a similar approach to Equation (4):

$$Y_{fdt} = \alpha_d + \gamma_t + \beta_1 HSR_{dt} + \beta_2 HSR_{dt} * Dist_d + \epsilon_{fdt} \tag{1.5}$$

$Dist_d$ here represents district size measured by population and district wage measured by per capita GDP.

The above analyses use binary treatment measures. To take into account the relative size and wage of the treated districts with the connected ones, I exploit the idea of “market access”, similar to Donaldson and Hornbeck (2016). Based on a model of cross-city labor sourcing, this approach allows a city’s “market access” to be affected by the city’s HSR connections to other cities. Empirically, the measure of a city’s “market access” is approximately the average of other cities’ GDP inversely weighted by the bilateral costs of passenger travel. In my context, I first assume that labor comes from not only within the district but from nearby districts from where workers can easily commute (by non-HSR and HSR transportation). After getting connected to the HSR network, the establishments can hire labor from further districts. As shown in Figure 1.4, the “accessible workers” for establishments in district A before HSR connections is the blue circle in the left panel, which includes districts B and C that are close to A. After being connected to district B, as shown in the right panel of Figure 1.4, the “accessible workers” for A now expands to include the green circles that are further away. So, the HSR reduces the travel time between A and B_i(i=1,2) and brings B_j(j=3,4) into commuting range.[14] According to Heuermann and Schmieder (2019), the expansion of HSR in Germany led to a shift toward rail commuting of medium distances between

[14] I only consider direct HSR connections in my analyses.

150 and 400 km, so the choice of the distance for non-HSR transportation is 150km and 400km for the HSR transportation. I calculate the following two measures of the HSR-induced change in "market access":

(1) Percentage change in the size of the "accessible labor market" (PCN), which is measured by the total population inversely weighted by the travel cost in the labor market and can be expressed as:

$$Population_d + \sum_{m=1}^{M} \frac{Population_m}{Distance_{dm}/240} + \sum_{n=1}^{N} \frac{Population_n}{Distance_{dn}/80},$$

where d represents the district where the establishment is located, and m represents the connected districts, and n represents the unconnected districts. The weight is measured by the travel time assuming the average speed of HSR transportation is 240km/h and non-HSR transportation is 80km/h.

(2) Percentage change in the average wage of the "accessible labor market" (PCW), which is measured by the average per capita GDP in the labor market and can be expressed as:

$$\frac{\sum_{d,m,n} PGDP * Population}{\sum_{d,m,n} Population}.$$

The two expressions use the baseline measure of population and per capita GDP so that the change in market access is only driven by a district's connection to the HSR.[15] The following equation is applied to capture the heterogeneous effects considering treatment

[15] I calculate the changes in the two measures of market access for all years and assign the largest ones as the measure for intensity.

intensity:

$$Y_{fdt} = \alpha_d + \gamma_t + \beta_1 HSR_{dt} + \beta_2 HSR_{dt} * Intensity_d + \epsilon_{fdt} \tag{1.6}$$

$Intensity_d$ represents the percentage change in the size (PCN) or wage (PCW) of the accessible labor market. Figure 1.5 shows the distributions of these two measures for the treated districts.

1.5.2 Empirical Design: Linked Employer-Employee Data

The above district and establishment level designs focus on changes in establishment composition and characteristics. The LIAB data provides information on a representative sample of establishments from 2000-2008 and all workers employed for at least one day during the time period. I first examine what kind of workers, in terms of gender, age, education, and commuting behavior, are entering into and leaving the establishments using the individual analogue of Equation (3):

$$Y_{ifdt} = \alpha_f + \gamma_t + \beta_1 HSR_{dt} + \beta_2 S_{ifdt} + \beta_3 HSR_{dt} * S_{ifdt} + \epsilon_{ifdt} \tag{1.7}$$

S_{ifdt} indicates whether worker i is an entrant or leaver. Y_{ifdt} represents worker characteristics including a dummy for male, a dummy for high education level, log age, and a dummy for commuter as well as commuting distance. Hence, β_3 estimates whether entrants/leavers are more/less likely have certain characteristics. And, from the establishment perspective, I can look into the change in worker composition, especially by gender and education using Equation (1).

Lastly, I consider the mechanisms for changes in establishment median wages.

Changes could be driven by increasing competition among establishments who will need to pay more to attract new workers and retain workers, or could stem from improvements in match quality due to higher labor mobility and more outside options after getting connected. To test the first hypothesis, I look at the change in the wage for the stayers:

$$Y_{ifdt} = \alpha_{if} + \gamma_t + \beta_1 HSR_{dt} + \epsilon_{ifdt} \tag{1.8}$$

Y_{ifdt} is the daily wage for workers and α_{if} are worker by establishment fixed effects, with which I can focus on the stayers. I also consider worker by district fixed effects as well as worker fixed effects and the details are described in Section 6.7.

To examine the second hypothesis, I look into labor market sorting by estimating whether workers are more likely to work in high-premium or large establishments at baseline. As a measure for establishment premium, I first regress worker wages on worker characteristics including gender, age, education as well as the interactions and include establishment by year fixed effects. The resulting fixed effects are used to measure the establishment effects controlling for worker effects. The following equation is used to test worker sorting:

$$Y_f = \alpha_i + \gamma_t + \beta_1 HSR_{dt} + \epsilon_{ifdt} \tag{1.9}$$

Y_f represents the baseline establishment effects and number of workers. The worker fixed effects α_i is used to capture whether workers are sorted towards high-premium or large establishments.

1.6 Results

1.6.1 Establishment Entry and Exit

After aggregating the BHP data to the district level, I calculate the total number of establishments, the number of establishments that enter into the district, and the number of establishments that exit the district every year. Few establishments change district, hence entry and exit refers to new establishments and establishments that close. First, I conduct event studies with Equation (2). Figures 1.6 and 1.7 verify parallel trends between the treated and control districts before the treatment. Figure 1.6 indicates a 1% insignificant increase in the number of establishments in treated districts. Figure 1.7 looks into the effects on entry in the top panel and exit in the bottom panel. The graphs indicate that the increase in the number of establishments is driven by an increase in new establishments rather than a reduction in exits. The average number of establishments that enter into treated district in the baseline year is 244 and experiences a 5% increase after the district is connected to the HSR. There is no significant change in exits except for a short-lived decrease one year after connection.

Next, I examine the characteristics of establishments that open and close in treated districts. Specifically, with the triple difference approach in Equation (3), I compare the size and wage of establishments between those that enter/exit and the existing ones. The results are shown in Table 1, where columns 1 to 3 use the establishment size measured by the log number of workers as the outcome and columns 4 to 6 use the establishment wage measured by the log median wage. Columns 1,2,4 and 5 look at all establishments. As indicated by columns 1 and 4, new establishments are, on average, 38.4% smaller in size and pay 14.1% lower wages, and they are 6.7% larger in treated districts relative to control districts. Similar patterns have been found in the closing establishments in columns 2 and 5. Columns 3 and 6 follow establishments that exist in the baseline and focus on the exit behavior of the larger ones (more than 10 employees at baseline) only. It suggests that smaller establishments are more likely to exit in treated districts and the size of the closing establishments are 24.5% smaller. The analysis reveals no significant differences in the wages of entering and exiting establishments in the treated districts relative to the controls. These district level results tell us that the HSR brings more new establishments into treated districts and the smaller ones are less likely to survive, indicating increasing competition among establishments.

1.6.2 Industry Specialization

As people have different willingness and ability to commute, the HSR might affect different groups of workers differently, or affect establishments in different industries differently, especially based on their wages and reliance on skills. Lin (2017) studies China's HSR and suggests that industries with a higher reliance on nonroutine cognitive skills benefit more from HSR-induced market access to other cities. With the first two digits of WZ93, I group establishments based on ISIC codes. I then calculate the average median wage and fraction of high-educated workers[16] by industry and group them into 2 categories: low-type industries and high-type industries. Table 1.A5 lists

[16] The high-educated workers are those who have a degree from a university or a university of applied sciences.

the 13 industries and their characteristics. I look into industry specialization within districts by calculating the total number of establishments that belong to each type of industry within the district. With Equation (2), I run the event studies to see the changes in the log number of establishments across the two types. The results are shown in Figure 1.8. In treated districts, there is an increasing number of establishment in the high-type industries while the number of establishment in the low-type industries remains unchanged.

1.6.3 Establishment Size and Wage

With the establishment level BHP data, I estimate changes in establishment size, measured by the total number of employees, and establishment wages, measured by the median wage of all full time workers. First, I conduct event studies with Equation (2). Figures 1.9 and 1.10 verify parallel trends between treated and control districts before the treatment. The top panels include district fixed effects and the bottom panels use establishment fixed effects. In Figure 1.9, we can observe an immediate and significant increase in the size of the treated establishments of 3% within the district and 1% within the establishment. The 3% increase in average size of the establishment within the district is driven by both the growth of existing establishments as well as the change in the composition of establishments, which is consistent with the result in Table 1.1 that smaller establishments are more likely to exit. As for the median wage, Figure 1.10 indicates that the establishments increase their wages for the average worker by around 0.75% four years after connection, and the effect is nearly identical when looking within establishments, suggesting that the overall increase is driven by existing establishments. With Equation (1) I then estimate the average treatment effects for the event window. The results are shown in the first two columns of Tables 1.2 and 1.3. In Table 1.2 we can see a 2.51% increase in establishment size in treated districts, while the within establishment growth is 0.67% and insignificant. Columns 1 and 2 of Table 1.3 show that the average treatment effects for wages are 0.515% with district fixed effects and 0.393% with establishment fixed effects.

Guided by the above findings, I further explore the heterogeneous effects across establishments. Specifically, with Equation (4), I estimate whether large or high-wage establishments grow more in size and wage. In Tables 1.2 and 1.3, *dlsize* represents the demeaned log size of the establishment at baseline and *dlwage* represents the demeaned log median wage of the establishment at baseline. Column 3 of Tables 1.2 and 1.3 indicate that a 1% increase in the baseline establishment size is associated with a 0.0233% decrease in treatment effects in terms of establishment size and a 0.0101% increase in the treatment effects in wage. However, nothing has been found in column 4 of Tables 1.2 and 1.3, saying there is no significant difference in the treatment effects among establishments with different baseline wages. So smaller establishments on the one hand are more likely to grow their business but on the other hand are also more likely to exit the labor market, while the larger ones are more likely to increase their wages as they have more latitude to set wages hence have more ability to increase wages.

1.6.4 Treatment Intensity

The above analyses consider the binary effect of being connected, without considering characteristics of districts. With Equation (5), I first explore the heterogeneous effects by size and wage of the treated districts and the results are shown in columns 3 and 5 of Tables 1.4 and 1.5. I use two measures for $Dist_d$: dln and dlw that take the demeaned values of log size and wage of the district at baseline. Size is measured by the district population and wage is measured by per capita GDP of the district. From Table 1.4, we can see that if the population of the treated district is 1% larger then the treatment effect on the size of the establishment increases by 0.0232%. Likewise a 1% increase in the per capita GDP is associated with a 0.119% increase in the treatment effect on establishment size. From Table 1.5 we can see there are no significant heterogeneous effects in wages. Hence large and high-wage treated districts are more likely to attract new workers.

Next with Equation (6), I use two measures of treatment intensity to consider characteristics of districts that the treated districts get connected with. pcn and pcw are percentage changes in population and per capita GDP of the "accessible workers" in the labor markets after the expansion driven by HSR connections. The results are shown in columns 4 and 6 of Tables 1.4 and 1.5, which indicate that a 1% increase in per capita GDP in the local labor market after getting connected is associated a 0.16% drop in the treatment effect in establishment size and a 0.0639% drop in the treatment effect in establishment wage. This means that if a treated district is connected with districts with higher per capita GDP, it becomes less attractive for new workers. The average establishment wages fall which could be driven by changes in wage-setting power of the establishment and the composition of workers.

1.6.5 Inflows and Outflows

Before using LIAB data to examine worker characteristics, I estimate event studies of worker inflows and outflows with the BHP data. The inflows are defined as the number of workers that are in the establishment in the current year but not the previous year and the outflows are defined as the number of workers that are in the establishment in the current year but not the next year. Previous results indicate no evidence of within establishment growth but increases in the size within districts driven by the exit of small establishments. The results for Equation (2) with log inflows and outflows as the outcome and district fixed effects are shown in Figure 1.11. We see increased turnover of workers in treated districts.

1.6.6 Worker Composition

Due to the fact that the HSR mainly affects workers who are more likely and more able to commute, we may expect workers with certain characteristics to have larger effects. As mentioned by Caldwell and Danieli (2020) regarding the expansion wave of the HSR in Germany, high educated workers (that are more likely to use the HSR) benefit from getting access to more distant jobs, and female workers (who tend to work closer to home) benefit from the increase in the supply of local jobs. With the LIAB data, I first test the relationship between several worker characteristics and commuting behaviors. The results are shown in Table 1.A6 and we can see that young, male, high-educated, and high-wage workers are more likely to be commuters and to commute longer distances. I then use worker characteristics to disentangle the effects of within establishment growth. Since the data contains information on all workers in representative establishments as well as their working biographies, I can observe their status as stayers, entrants or leavers. With the triple difference approach in Equation (7), I first compare the characteristics of workers between the entrants/leavers and the stayers, focusing on age, gender, education, and commuting behavior.

The results are shown in Tables 1.6 and 1.7, which indicate that within establishments, the entrants are more likely to be young, high-educated commuters and are more likely to commute longer distance in the treated districts. The result in gender is not significant for entrants. No significant effects are evident for leavers, except that they tend to be younger in the treated establishments. Furthermore, based on the characteristics, I calculate the number and fraction of workers by gender, education, commuting behavior as well as the average commuting distance at the establishment level. I also calculate the number and fraction of male commuters and high-educated commuters as they are most directly affected by the HSR. After aggregating these measures to the establishment level, I conduct analyses with Equation (1) and the results are shown in Tables 1.8 and 1.9. The top panels include district fixed effects and the bottom panels include establishment fixed effects. Table 1.8 shows the results by gender and education, and we can see that within districts, there are increasing numbers of workers in different subgroups except for the low-educated. The fraction of high-educated workers significantly increases by 0.02 while the fraction of low-educated workers decreases by 0.0183. Less has been found within establishments. Table 1.9 shows the results for commuters, indicating significant increases in the number and fraction of high-educated commuters within districts as well as significant increases in the number and fraction of commuters, especially male commuters, within establishments. The average commuting distance increases by around 5% in both specifications. By digging into worker composition, Table 1.6-1.9 suggest that establishments are hiring more workers from other districts, especially more young, male, high-educated workers.

1.6.7 Wage of Stayers

Establishment level analyses indicate an increase in the median wage within existing establishments. However, it could be driven by both the growth of individual wages as well as the change in the composition of workers. With details on worker wages, I can test to what extent the wage growth is driven by an increase in the wage of stayers, which could be addressed by adding worker fixed effects in the worker level regression as in Equation (8). The results are shown in Table 1.10. Columns 1 to 4 show the results for all workers and columns 5 to 8 focus on high-educated workers. Four specifications are considered. The first one examines the initial workers in the sample by using worker fixed effects. The second and third ones test to what extent the effect is driven by workers changing jobs to other districts and establishments by adding district and establishment fixed effects. The last one is to test the competition mechanism by controlling for match-level fixed effects. The treatment is defined with their baseline district of work and remains the same even if they change their job to another district. For all workers, there is small and insignificant increase in their wages as shown in column 1. High-educated workers experience a larger and significant wage growth as shown in column 4, which is consistent with the fact that they are more likely to be affected by the HSR. As shown in columns 6 to 8, the wage growth for high-educated workers reduces from 1.14% to 1.02% when adding establishment fixed effects, indicating part of the growth in driven by workers changing their jobs to establishments with higher wages, which indicates increased competition. It reduces to 0.807% in the last column, indicating that better match quality plays a role, and reduced monopsony power leads to significant wage growth even for stayers.

1.6.8 Worker Sorting

Lastly, the increase in labor mobility and turnover of workers has raised the probability of changes in labor market sorting. To understand whether workers are more likely to work in establishments that have higher wage premiums at baseline, I first estimate the regression-adjusted wages of establishments controlling for observed worker characteristics. With the saved establishment by year fixed effects as the outcome in Equation (1), panel A of Table 1.11 reveals that there are significant increases in the adjusted wage of establishments both within district and establishment. Columns 1 and 2 use all establishment in the data while columns 3 and 4 only use the representative sample.[17] Panels B and C of Table 1.11 show the results regarding worker sorting with Equation (9). Panel B uses the baseline measure of the regression-adjusted wage as the outcome and panel C uses the log number of workers in the baseline as the outcome. The treatment is similarly defined as in Table 1.10. We see little evidence of workers sorting towards high premium establishments for all workers, however, workers tend to work in smaller establishments. This is also consistent with the establishment level analyses that smaller establishments are more likely to grow in size, adding some evidence that they are hiring more new workers either from other districts or who just enter into the labor market.

1.7 Conclusion

This paper examines the responses of establishments and workers as a result of labor market integration due to transportation improvements. It provides another angel to understand the benefits of infrastructure network through reducing monopsony power and improving match quality as well as sheds light one potential source of labor market frictions. With historical train schedules and the staggered adoption of stations in the expansion wave of HSR in Germany as a natural experiment, this paper first finds increased establishment entry into treated districts, and that treated districts become more specialized in industries with higher wages and higher reliance on skills. Smaller establishments are less likely to survive but are also more likely to grow in size. Larger

[17] A lot of establishments that are not in the sample appear in the data since the data includes the workers in the sample as well as their working biographies, so the information for the establishments that ever hire those workers are also available.

establishments, however, have greater ability to increase wages. Treated districts that are larger or pay higher wages or are connected with districts that pay lower wages are more likely to attract workers. With the detailed information on workers, I find that establishments are hiring more young, male, high-educated workers from other districts. The fractions of male and high-educated commuters within establishments have been found to increase. Controlling for observed worker characteristics, there is still increase in wages of establishments. The wage of stayers increases significantly, especially for the high-educated workers, and both reduced monopsony power and better match quality seem to play a role. The results show evidence of reduced labor market frictions from improved labor mobility. It also sheds light on the importance of policies that enhance workers' ability to switch employers.

References

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

G. M. Ahlfeldt and A. Feddersen. From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2):355–390, 2018.

T. Allen and C. Arkolakis. The welfare effects of transportation infrastructure improvements. Technical report, National Bureau of Economic Research, 2019.

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

J. Azar, I. Marinescu, M. Steinbaum, and B. Taska. Concentration in us labor markets: Evidence from online vacancy data. *Labour Economics*, 66:101886, 2020.

A. Banerjee, E. Duflo, and N. Qian. On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, 145:102442, 2020.

I. Bassier, A. Dube, and S. Naidu. Monopsony in movers: The elasticity of labor supply to firm wage policies. Technical report, National Bureau of Economic Research, 2020.

E. Benmelech, N. K. Bergman, and H. Kim. Strong employers and weak employees: How does employer concentration affect wages? *Journal of Human Resources*, pages 0119–10007R1, 2020.

A. B. Bernard, A. Moxnes, and Y. U. Saito. Production networks, geography, and firm performance. *Journal of Political Economy*, 127(2):639–688, 2019.

V. Bhaskar, A. Manning, and T. To. Oligopsony and monopsonistic competition in labor markets. *Journal of Economic Perspectives*, 16(2):155–174, 2002.

W. M. Boal and M. R. Ransom. Monopsony in the labor market. *Journal of economic literature*, 35(1):86–112, 1997.

S. Caldwell and O. Danieli. *Outside options in the labor market.* Pinhas Sapir Center for Development, Tel Aviv University, 2020.

D. Card, J. Heining, and P. Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015, 2013.

D. Card, A. R. Cardoso, and P. Kline. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2):633–686, 2016.

P. Charnoz, C. Lelarge, and C. Trevien. Communication costs and the internal organisation of multi-plant businesses: Evidence from the impact of the french high-speed rail. *The Economic Journal*, 128(610):949–994, 2018.

M. C. Dao, M. Das, Z. Koczan, and W. Lian. Understanding the downward trend in labor income shares. *World Economic Outlook April 2017: Gaining Momentum*, pages 121–71, 2017.

D. Donaldson. Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934, 2018.

D. Donaldson and R. Hornbeck. Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858, 2016.

X. Dong, S. Zheng, and M. E. Kahn. The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, 115:103212, 2020.

A. Dube, J. Jacobs, S. Naidu, and S. Suri. Monopsony in online labor markets. *American Economic Review: Insights*, 2(1):33–46, 2020.

J. Eberle, P. Jacobebbinghaus, J. Ludsteck, and J. Witter. Generation of time-consistent industry codes in the face of classification changes. *Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport*, 5:2011, 2011.

B. Faber. Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070, 2014.

J. Firth. I've been waiting on the railroad: The effects of congestion on firm production. Technical report, MIT Working Paper, 2017.

D. Goldschmidt and J. F. Schmieder. The rise of domestic outsourcing and the evolution of the german wage structure. *The Quarterly Journal of Economics*, 132(3):1165–1217, 2017.

A. Gumpert, H. Steimer, and M. Antoni. Firm organization with multiple establishments. 2019.

T. Hethey-Maier and J. F. Schmieder. Does the use of worker flows improve the analysis of establishment turnover? evidence from german administrative data. Technical report, National Bureau of Economic Research, 2013.

D. F. Heuermann and J. F. Schmieder. The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in germany. *Journal of economic geography*, 19(2):335–372, 2019.

B. Hirsch and S. Mueller. Firm wage premia, industrial relations, and rent sharing in germany. *ILR Review*, 73(5):1119–1146, 2020.

B. Hirsch, E. J. Jahn, A. Manning, and M. Oberfichtner. The urban wage premium in imperfect labor markets. *Journal of Human Resources*, pages 0119–9960R1, 2020.

W. Klosterhuber, J. Heining, S. Seth, et al. Linked-employer-employee-daten des iab: Liab-längsschnittmodell 1993-2010 (liab lm 9310). *FDZ-Datenreport*, 8:2013, 2013.

K. Lavetti and I. M. Schmutte. Estimating compensating wage differentials with endogenous job mobility. 2016.

Y. Lin. Travel costs and urban specialization patterns: Evidence from china's high speed railway system. *Journal of Urban Economics*, 98:98–123, 2017.

Y. Lin, Y. Qin, and Z. Xie. Does foreign technology transfer spur domestic innovation? evidence from the high-speed rail sector in china. *Journal of Comparative Economics*, 49(1):212–229, 2021.

A. Manning. *Monopsony in motion: Imperfect competition in labor markets.* Princeton University Press, 2003.

A. Manning. Imperfect competition in the labor market. In *Handbook of labor economics*, volume 4, pages 973–1041. Elsevier, 2011.

A. Manning and B. Petrongolo. How local are labor markets? evidence from a spatial job search model. *American Economic Review*, 107(10):2877–2907, 2017.

D. T. Mortensen and C. A. Pissarides. New developments in models of search in the labor market. *Handbook of labor economics*, 3:2567–2627, 1999.

J. S. Nimczik. Job mobility networks and endogenous labor markets. 2017.

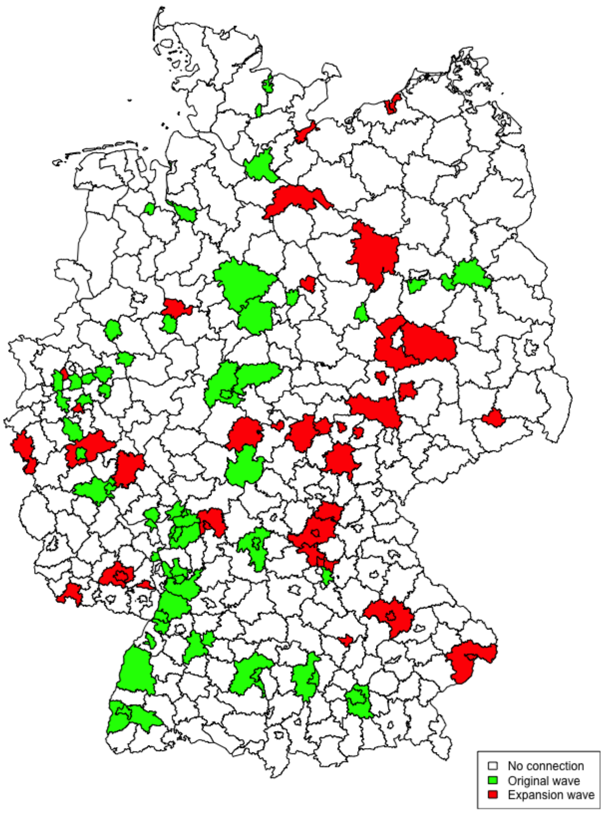
J. Song, D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter. Firming up inequality. *The Quarterly journal of economics*, 134(1):1–50, 2019.

N. Tsivanidis. The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio. *Unpublished manuscript*, 2018.

D. A. Webber. Firm market power and the earnings distribution. *Labour Economics*, 35:123–134, 2015.

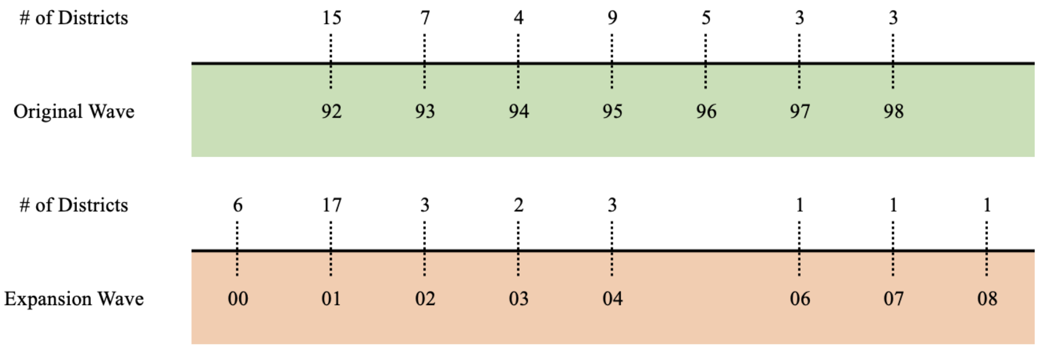
Figures and Tables

Figure 1.1: German Districts with HSR Stations



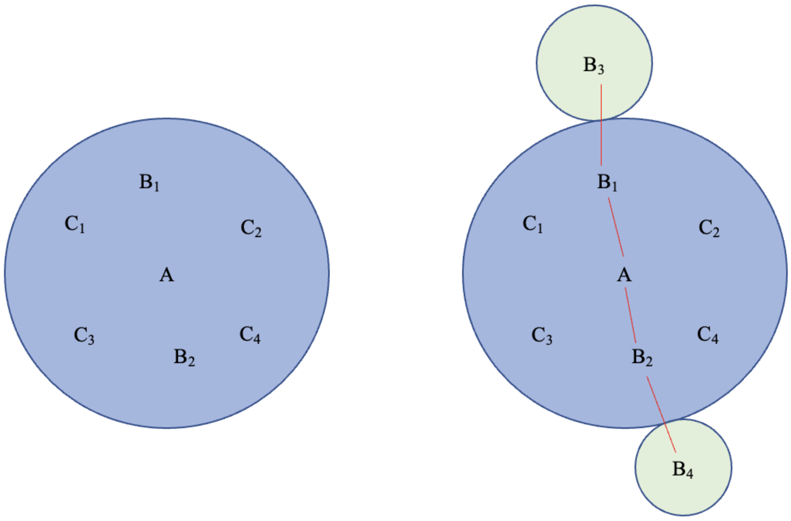
Note: The original wave connected major cities in 1991-1998; the expansion wave connected smaller cities on the existing routes in 1999-2008.

Figure 1.2: Event Time Distribution



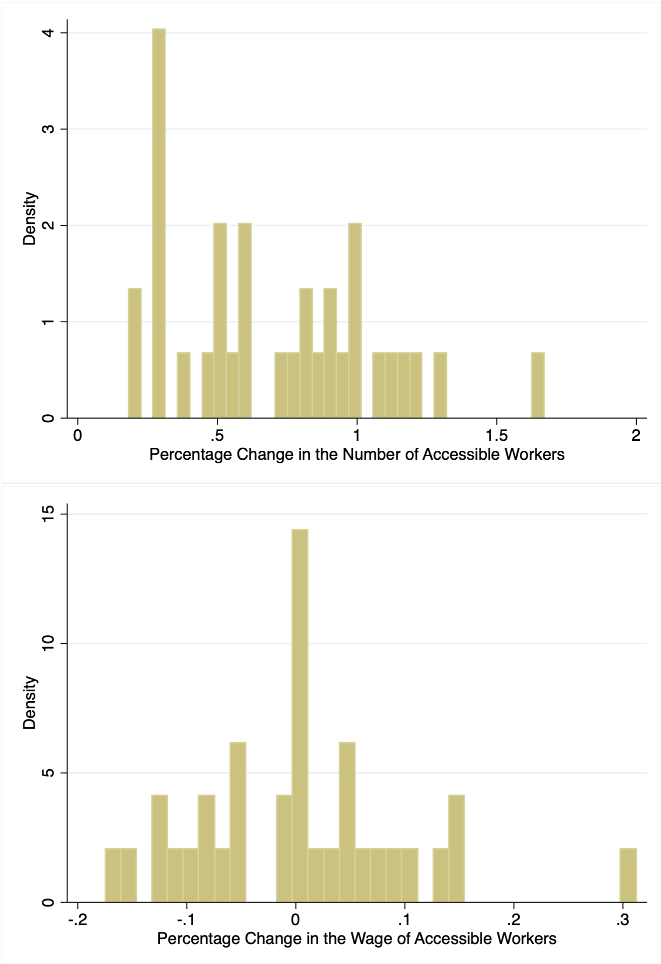
Note: 46 out of 412 districts were connected to the HSR during the original wave (1991-1998). 34 out of 412 districts were connected during the expansion wave (1999-2008) and form the treated group.

Figure 1.3: Accessible Workers Before and After Connection



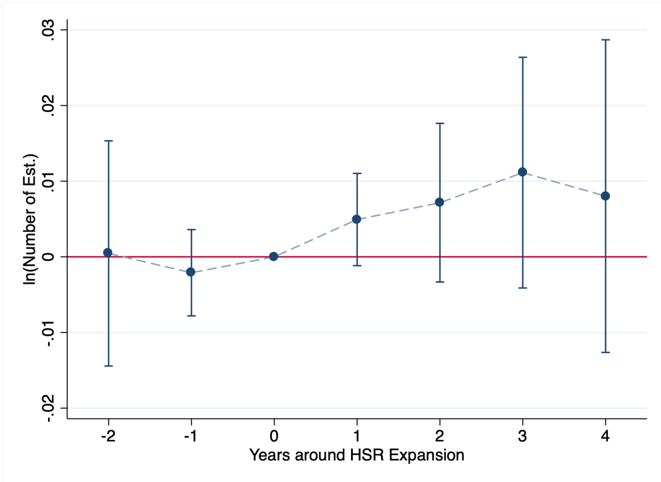
Note: The pool of labor is considered within 150km without HSR and within 400km with HSR.

Figure 1.4: Treatment Intensity



Note: This figure shows the distributions of the two measures of treatment intensity for the treated districts.

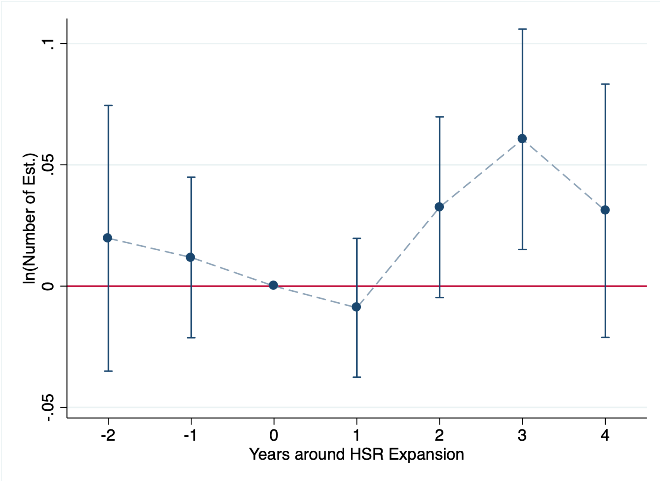
Figure 1.5: Log Number of Establishments in the District



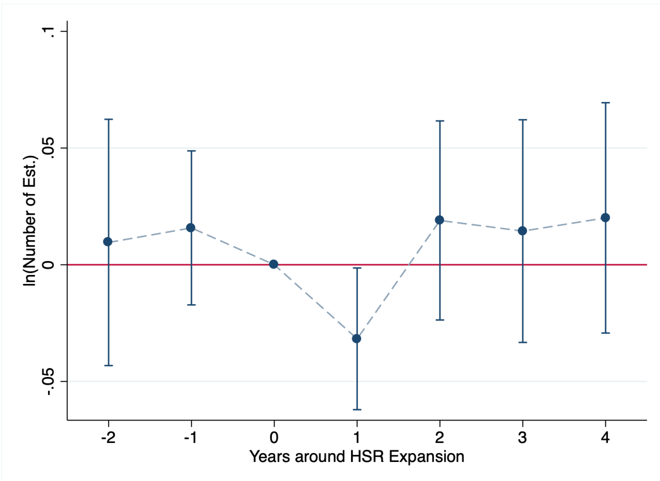
Note: This figure shows the dynamic effects of HSR on the log number of establishments in the district from two years before to four years after the treatment. The average number of establishments in the treated districts in the baseline year is 3311.

Figure 1.6: Log Number of Entry and Exit Establishments in the District

(a) Entry



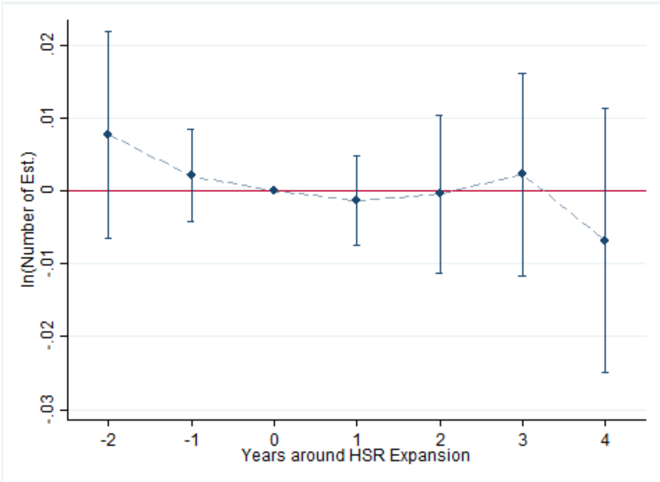
(b) Exit



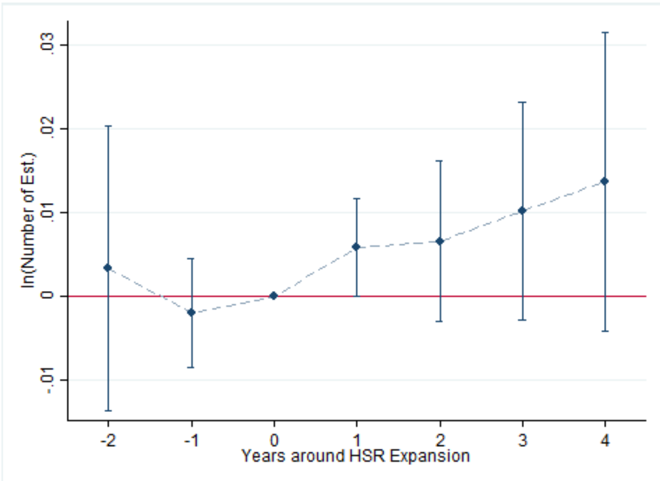
Note: The top and bottom panel shows the dynamic effects of HSR on the log number of entry and exit establishments in the district from two years before to four years after the treatment. The average number of entry establishments in the treated district in the baseline year is 244. The average number of exit establishments in the treated district in the baseline year is 263.

Figure 1.7: Industry Specialization

(a) Low-type Industries



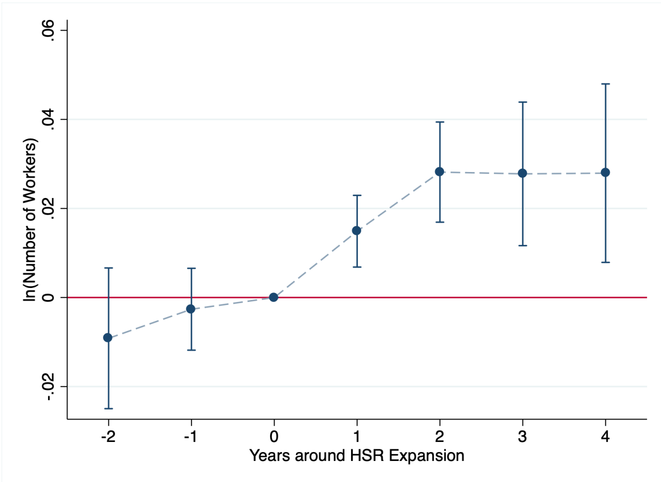
(b) High-type Industries



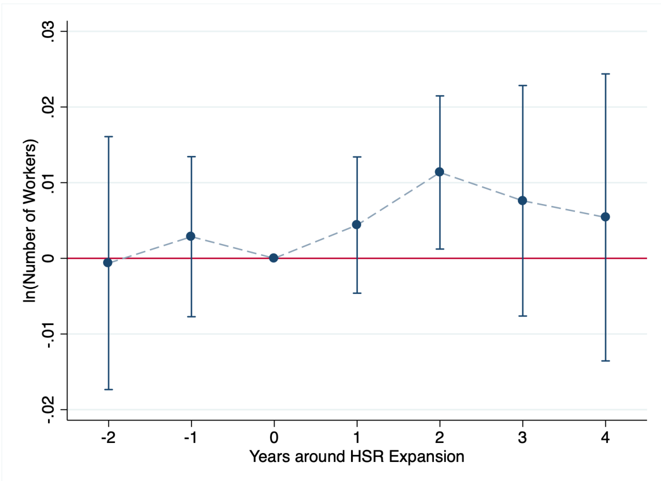
Note: This figure shows the dynamic effects of HSR on the log number of establishments in the two types of industries from two years before to four years after the treatment. The top panel shows the result for the low-type industries and the bottom panel shows the result for the high-type industries.

Figure 1.8: Log Number of Workers in the Establishment

(a) District Fixed Effects



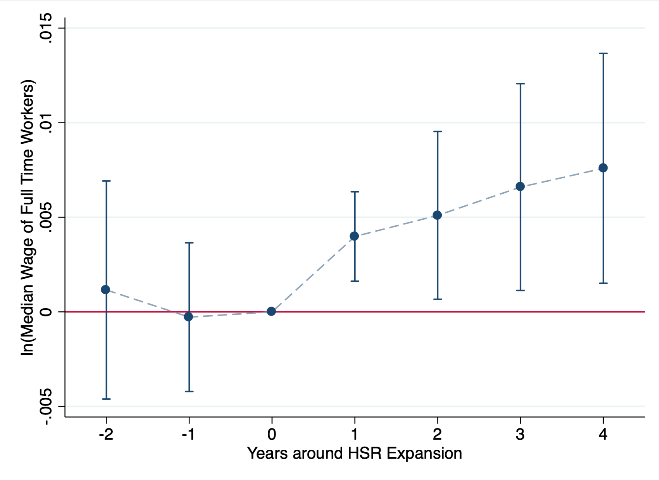
(b) Establishment Fixed Effects



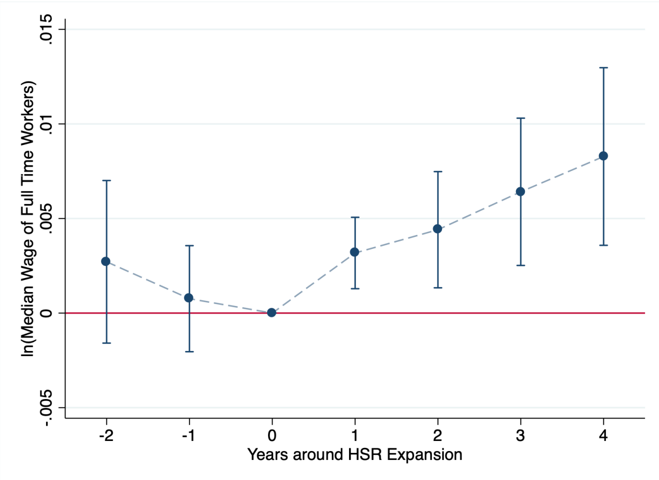
Note: This figure shows the dynamic effects of HSR on the log number of workers in the establishment from two years before to four years after the treatment. The top panel includes district fixed effects and the bottom panel uses establishment fixed effects.

Figure 1.9: Log Median Wage of Full Time Workers in the Establishment

(a) District Fixed Effects



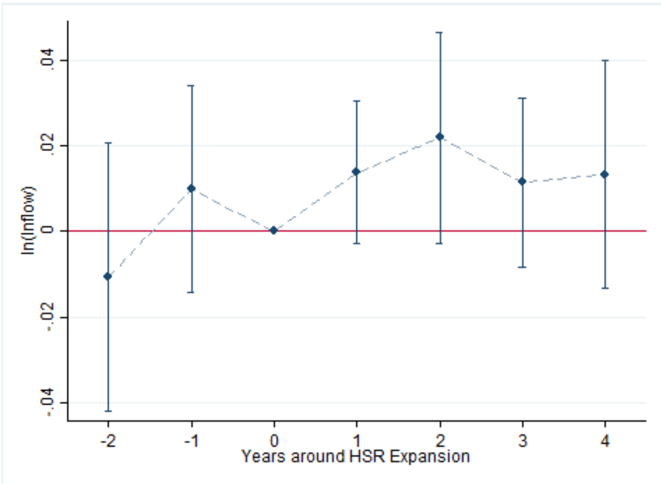
(b) Establishment Fixed Effects



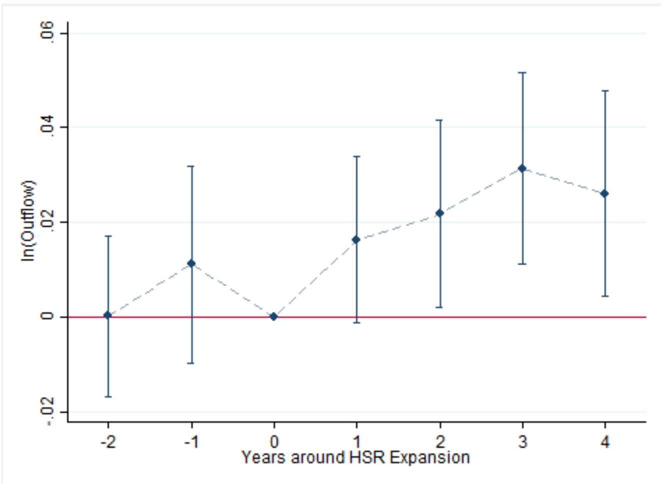
Note: This figure shows the dynamic effects of HSR on the log median wage of the establishment from two years before to four years after the treatment. The top panel includes district fixed effects and the bottom panel uses establishment fixed effects.

Figure 1.10: Worker Inflows and Outflows

(a) Inflows



(b) Outflows



Note: The top and bottom panel shows the dynamic effects of HSR on the log number of inflows and outflows in the establishment from two years before to four years after the treatment.

Table 1.1: Characteristics of Establishments that Enter and that Exit

	(1)	(2)	(3)	(4)	(5)	(6)
	log(number of workers)			log(median wage of full time workers)		
hsr	-0.00514	-0.00422	0.0251***	-0.000373	0.00000556	0.00460*
	(0.00330)	(0.00356)	(0.00773)	(0.00257)	(0.00283)	(0.00245)
entry	-0.384***			-0.141***		
	(0.0120)			(0.00534)		
hsrXentry	0.0668***			-0.0102		
	(0.0149)			(0.0109)		
exit		-0.384***	-0.592***		-0.127***	-0.0489***
		(0.00971)	(0.0211)		(0.00523)	(0.00541)
hsrXexit		0.0292*	-0.245***		-0.00543	0.0138
		(0.0173)	(0.0418)		(0.0104)	(0.00975)
Dist. FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Mean	1.661	1.661	3.519	3.940	3.940	4.201
N	3,224,621	3,229,245	296,164	2,288,823	2,292,225	295,097

Note: *entry/exit* is a dummy variable and equals to 1 if the establishment enters into/exits the district. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.2: Heterogeneous Effects on the Size by Establishment Size and Wage

	(1)	(2)	(3)	(4)
		log(number of workers)		
hsr	0.0251***	0.00670	0.00693	0.00670
	(0.00719)	(0.00803)	(0.00813)	(0.00803)
hsrX(size of the est.)			-0.0233***	
			(0.00697)	
hsrX(wage of the est.)				0.000498
				(0.0140)
Dist. FE	Y			
Est. FE		Y	Y	Y
Year FE	Y	Y	Y	Y
Mean	3.519	3.519	3.519	3.519
N	299,844	299,482	299,482	299,482

Note: *size of the est.* represents the demeaned log size of the establishment at baseline and *wage of the est.* represents the demeaned log median wage of the establishment at baseline. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.3: Heterogeneous Effects on the Wage by Establishment Size and Wage

	(1)	(2)	(3)	(4)
	log(median wage of full time workers)			
hsr	0.00515**	0.00393**	0.00382**	0.00397**
	(0.00254)	(0.00172)	(0.00171)	(0.00162)
hsrX(size of the est.)			0.0101***	
			(0.00120)	
hsrX(wage of the est.)				-0.00765
				(0.0112)
Dist. FE	Y			
Est. FE		Y	Y	Y
Year FE	Y	Y	Y	Y
Mean	4.201	4.201	4.201	4.201
N	298,757	298,378	298,378	298,378

Note: *size of the est.* represents the demeaned log size of the establishment at baseline and *wage of the est.* represents the demeaned log median wage of the establishment at baseline. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.4: Heterogeneous Effects on the Size by District Size and Wage

	(1)	(2)	(3)	(4)	(5)	(6)
			log(number of workers)			
hsr	0.0251***	0.00670	0.0250 ***	0.0403***	0.0252**	0.0103
	(0.00719)	(0.00803)	(0.00738)	(0.0136)	(0.00716)	(0.0100)
hsrX(size of the dist.)			0.0232**			
			(0.00972)			
hsrX(size of the connected dist.)				-0.0204		
				(0.0213)		
hsrX(wage of the dist.)					0.119**	
					(0.0551)	
hsrX(wage of the connected dist.)						-0.160 **
						(0.0766)
Dist. FE	Y		Y	Y	Y	Y
Est. FE		Y				
Year FE	Y	Y	Y	Y	Y	Y
Mean	3.519	3.519	3.519	3.519	3.519	3.519
N	299,844	299,482	299,844	299,844	299,844	299,844

Note: *size of the dist.* and *wage of the dist.* take the demeaned values of log size and wage of the district at baseline. And the size is measured by the population and the wage is measured by per capita GDP of the district. *size of the connected dist.* and *wage of the connected dist.* are percentage changes in population and per capita GDP of the “accessible workers” in the labor markets after the expansion driven by HSR connections. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.5: Heterogeneous Effects on the Wage by District Size and Wage

	(1)	(2)	(3)	(4)	(5)	(6)
	log(median wage of full time workers)					
hsr	0.00515**	0.00393**	0.00513 **	0.00189	0.00514**	-0.000777
	(0.00254)	(0.00172)	(0.00242)	(0.00501)	(0.00253)	(0.00288)
hsrX(size of the dist.)			0.00388			
			(0.00317)			
hsrX(size of the connected dist.)				0.00439		
				(0.00612)		
hsrX(wage of the dist.)					-0.00558	
					(0.00947)	
hsrX(wage of the connected dist.)						-0.0639***
						(0.0231)
Dist. FE	Y		Y	Y	Y	Y
Est. FE		Y				
Year FE	Y	Y	Y	Y	Y	Y
Mean	4.201	4.201	4.201	4.201	4.201	4.201
N	298,757	298,378	298,757	298,757	298,757	298,757

Note: *size of the dist.* and *wage of the dist.* take the demeaned values of log size and wage of the district at baseline. And the size is measured by the population and the wage is measured by per capita GDP of the district. *size of the connected dist.* and *wage of the connected dist.* are percentage changes in population and per capita GDP of the "accessible workers" in the labor markets after the expansion driven by HSR connections. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.6: Characteristics of Entrants

	(1) male	(2) high edu.	(3) log(age)	(4) commuter	(5) log(distance)
hsr	-0.00136	-0.0116**	0.0397***	-0.0144***	-0.0436***
	(0.00327)	(0.00560)	(0.00684)	(0.00445)	(0.0111)
entrant	0.00549	0.0427***	-0.0137***	0.0648***	0.151***
	(0.00518)	(0.00535)	(0.00428)	(0.00444)	(0.0178)
hsrXentrant	0.0126	0.0406**	-0.119***	0.0251**	0.0598***
	(0.00879)	(0.0199)	(0.0153)	(0.0101)	(0.0227)
Est. FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Mean	0.576	0.204	3.682	0.335	3.061
N	1,558,606	1,506,561	1,558,606	1,558,606	1,558,606

Note: *entrant* is a dummy variable and equals to 1 if the worker is an entrant to the establishment. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.7: Characteristics of Leavers

	(1) male	(2) high edu.	(3) log(age)	(4) commuter	(5) log(distance)
hsr	0.000836	-0.000907	-0.00136	0.0102**	0.0104
	(0.00336)	(0.00582)	(0.00305)	(0.00441)	(0.0135)
leaver	0.0148***	0.00913	-0.0234***	0.0412***	0.125***
	(0.00424)	(0.00607)	(0.00665)	(0.00367)	(0.0149)
hsrXleaver	-0.000141	0.0164	-0.0196**	-0.0113	0.00115
	(0.00692)	(0.0150)	(0.00863)	(0.00812)	(0.0260)
Est. FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Mean	0.576	0.204	3.682	0.335	3.061
N	1,725,520	1,660,511	1,725,520	1,725,520	1,725,520

Note: *leaver* is a dummy variable and equals to 1 if the worker is a leaver to the establishment. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.8: Worker Composition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	log number of				fraction of			
	male	female	high edu.	low edu.	male	female	high edu.	low edu.
hsr	0.0105	0.0146	0.0771	-0.0225	0.00259	-0.00259	0.0200*	-0.0183*
	(0.0414)	(0.0562)	(0.0650)	(0.0487)	(0.0126)	(0.0126)	(0.0116)	(0.0108)
Dist. FE			Y				Y	
Year FE			Y				Y	
hsr	0.0140	0.0246	0.0132	0.0169	-0.00131	0.00131	-0.00153	-0.00304
	(0.0190)	(0.0234)	(0.0251)	(0.0290)	(0.00339)	(0.00339)	(0.00326)	(0.00510)
Est. FE			Y				Y	
Year FE			Y				Y	
Mean	3.424	3.115	2.050	3.794	0.562	0.438	0.171	0.754
N	13,117	13,117	13,117	13,117	13,117	13,117	13,117	13,117

Note: Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.9: Worker Composition by Commuting Behavior

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	log number of				fraction of				log average
	commuter	non commuter	male commuter	high edu. commuter	commuter	non commuter	male commuter	high edu. commuter	commuting distance
hsr	0.0397	-0.00919	0.0655	0.0922*	0.0110	-0.0110	0.00992	0.0141*	0.0532*
	(0.0428)	(0.0388)	(0.0452)	(0.0476)	(0.00686)	(0.00686)	(0.00662)	(0.00733)	(0.0290)
Dist. FE			Y				Y		Y
Year FE			Y				Y		Y
hsr	0.0552*	0.00471	0.0749***	0.0338	0.0118**	-0.0118**	0.00923***	0.00173	0.0462***
	(0.0285)	(0.0252)	(0.0273)	(0.0242)	(0.00472)	(0.00472)	(0.00283)	(0.00305)	(0.0169)
Est. FE			Y				Y		Y
Year FE			Y				Y		Y
Mean	2.589	3.808	2.070	1.256	0.268	0.732	0.163	0.061	3.271
N	13,117	13,117	13,117	13,117	13,117	13,117	13,117	13,117	13,117

Note: Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.10: Wage of Stayers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	log(daily wage)							
	all workers				high-educated workers			
hsr_base	0.00521	0.00766	0.0101**	0.00825**	0.0114***	0.0156***	0.0102**	0.00807**
	(0.00683)	(0.00739)	(0.00399)	(0.00371)	(0.00508)	(0.00532)	(0.00456)	(0.00352)
Worker FE	Y	Y	Y		Y	Y	Y	
Dist. FE		Y				Y		
Est. FE			Y				Y	
WorkerXEst. FE				Y				Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Mean	4.215	4.215	4.215	4.215	4.645	4.610	4.610	4.610
4.610								
N	3,311,021	3,311,021	3,255,238	3,033,250	467,089	467,089	459,877	444,424

Note: The analyses for this table follow workers who work in treated and control districts in the baseline. *hsr_base* is defined with their baseline district of work and remains the same even if they change their job to another district. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 1.11: Worker Sorting

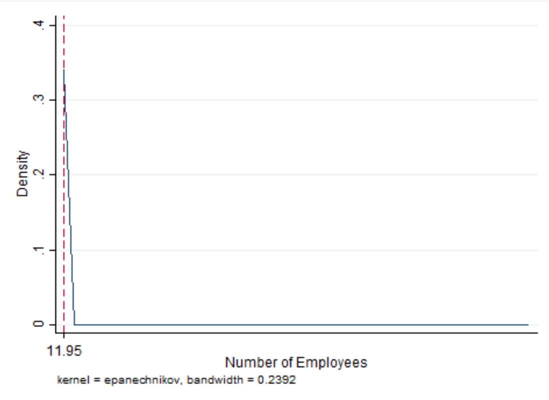
Panel A	(1)	(2)	(3)	(4)
	log(establishment premium)			
hsr	0.0158**	0.00811**	0.0110	0.00959*
	(0.00635)	(0.00391)	(0.00939)	(0.00531)
Dist. FE	Y		Y	
Est. FE		Y		Y
Year FE	Y	Y	Y	Y
Mean	-0.195	-0.195	-0.172	-0.172
N	225,542	204,090	19,004	18,508
Panel B	(1)	(2)	(3)	(4)
	log(establishment premium)			
	all workers		high-educated workers	
hsr_base	-0.000743*	0.000516	-0.000463	0.00413
	(0.000418)	(0.0166)	(0.000420)	(0.00825)
Dist. FE		Y		Y
Worker FE	Y		Y	
Year FE	Y	Y	Y	Y
Mean	-0.00200	-0.00200	0.0370	0.0370
N	1,701,115	1,773,576	267,869	276,752
Panel C	(1)	(2)	(3)	(4)
	log(number of workers)			
	all workers		high-educated workers	
hsr_base	0.00463	-0.0857*	0.000751	-0.0193
	(0.00512)	(0.00500)	(0.00309)	(0.0440)
Dist. FE		Y		Y
Worker FE	Y		Y	
Year FE	Y	Y	Y	Y
Mean	6.116	6.116	6.116	6.116
N	1,705,001	1,7778,929	267,999	276,896

Note: *establishment premium* is measured by the regression-adjusted wages of establishments controlling for observed worker characteristics. Panel B and C use the baseline measures of establishment premium and size. Standard errors are clustered at the district level. The symbols *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

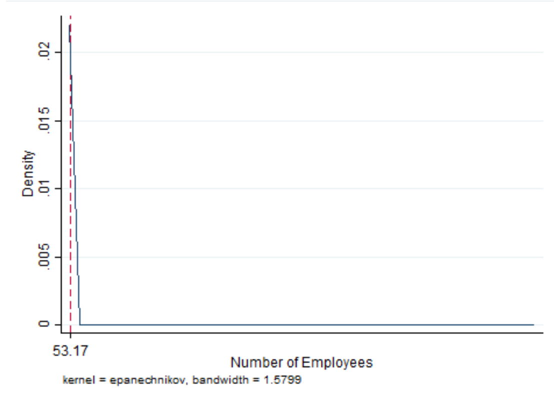
Appendices

Figure 1.A1: Distribution of Establishment Size

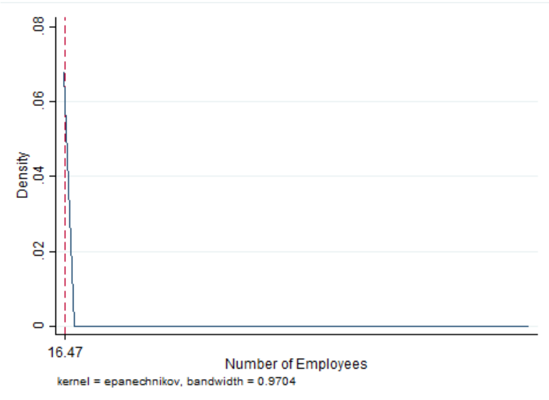
(a) BHP, all sample



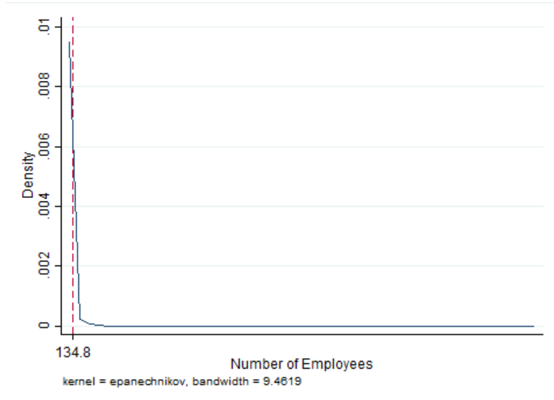
(b) BHP, larger sample



(c) LIAB, with weight



(d) LIAB, without weight



Note: The top left panel of shows the distribution of the size of all establishments and the top right panel shows the distribution of larger establishments that have at least ten employees in 2004. The bottom left panel shows the distribution of the weighted size of establishments and the right panel shows the distribution of the size without weight.

Table 1.A1: List of Treated Cities

Dresden	Leipzig	Aachen	Halle(Saale)
Oberhausen	Lübeck	Erfurt	Rostock
Saarbrücken	Solingen	Regensburg	Ingolstadt
Wolfsburg	Erlangen	Jena	Kaiserslautern
Lüneburg	Bamberg	Aschaffenburg	Weimar
Herford	Neustadt(Weinstr)	Neustadt(Weinstr)	Lutherstadt Wittenberg
Gotha	Eisenach	Stendal	Siegburg
Naumburg(Saale)	Bad Hersfeld	Köthen	Saalfeld(Saale)
Lichtenfels	Montabaur		

Note: These are the 34 cities that got a station during the expansion wave of the HSR from 1999 to 2010.

Table 1.A2: District Population and GDP

	Original Wave		Expansion Wave		No Connection	
Year	Population	GDP	Population	GDP	Population	GDP
2000	435,179	18,131.68	180,288	6,553.87	166,646	3,861.52
2001	435,690	18,799.01	180,222	6,687.70	167,107	3,967.85
2002	437,175	19,062.80	180,364	6,855.44	167,720	3,997.19
2003	438,007	19,133.33	180,527	6,990.17	168,131	4,017.28
2004	438,598	19,491.07	180,870	7,130.03	168,181	4,126.59
2005	439,015	19,684.49	180,854	7,250.93	168,223	4,174.65
2006	440,370	20,464.83	180,995	7,554.97	168,020	4,360.82
2007	441,914	21,402.53	181,084	7,906.41	167,616	4,577.11
2008	443,569	21,743.70	181,015	8,006.43	167,299	4,679.30

Note: GDP is in million euros.

Table 1.A3: Sample Size

	BHP	LIAB	
	Establishment	Establishment	Worker
1998	1,015,567		
1999	1,240,570		
2000	1,262,991	5,252	709,297
2001	1,260,957	5,204	668,506
2002	1,241,909	4,877	618,922
2003	1,248,306	4,949	598,452
2004	1,313,467	5,032	575,181
2005	1,335,830	5,088	545,543
2006	1,363,898	5,305	539,440
2007	1,389,987	5,729	561,461
2008	1,402,199	6,324	583,856
2009	1,424,637		
2010	1,445,083		
2011	1,463,681		
2012	1,476,923		

Note: The left panel shows the sample size of the BHP data, which is a 50% sample of all establishments throughout Germany. The right panel shows the sample size of the LIAB data, which links workers' information on a representative sample of establishments.

Table 1.A4: Summary of Establishments and Workers

	LIAB			
	Establishment		Worker	
	Mean	SD	Mean	SD
Total	102.45	(430.77)		
Commuter	0.24	(0.27)	0.33	(0.47)
Male	0.55	(0.35)	0.61	(0.49)
Higher Education	0.12	(0.21)	0.20	(0.40)
Lower Education	0.73	(0.33)	0.80	(0.40)
Age	41.99	(6.27)	42.22	(9.86)
Daily Wage	65.36	(27.40)	92.99	(39.06)
N	4,963		508,481	

Table 1.A5: Establishment Characteristics by Industry

	Average wage	Fraction of High-educated
Low-type industry		
Mining and quarrying	49.43	0.0602
Manufacturing	71.40	0.1063
Construction	66.10	0.0491
Wholesale and retail trade	63.75	0.0615
Hotels and restaurants	37.54	0.0250
Transport, storage and communication	62.54	0.0501
High-type industry		
Electricity, gas and water supply	116.30	0.1574
Financial intermediation	83.38	0.1457
Real estate, renting and business activities	72.30	0.1683
Public administration and defence	90.25	0.1561
Education	83.86	0.3126
Health and social work	56.65	0.1271
Other community, social and personal service activities	57.19	0.1465

Table 1.A6: Worker Characteristics and Commuting Behavior

	(1) male	(2) high edu.	(3) log(age)	(4) log(wage)
commuter	0.0455***	0.0591***	-0.0220***	0.0550***
	(0.0030)	(0.0093)	(0.0040)	(0.0050)
Est. FE		Y		
Year FE		Y		
log(distance)	0.0331***	0.0440***	-0.0182***	0.0285***
	(0.0016)	(0.0043)	(0.0015)	(0.0041)
Est. FE		Y		
Year FE		Y		
Mean	0.598	0.201	3.691	4.346
N	7,537,941	7,059,032	7,537,941	7,435,884

Chapter 2

The Labor Market Effects of Immigration Enforcement: Evidence from the 2007 Legal Arizona Workers Act (LAWA)

2.1 Introduction

The United States is home to the largest immigrant population in the world, accounting for one-fifth of the world's immigrants as of 2017 (Migration Policy Institute), and accounts for an even greater percentage of the world's undocumented immigrants. Recent years have seen growth of the undocumented population from approximately 3 million in 1990 to 11 million in 2009. (Passel and D'Vera Cohn, 2011). A more recent estimate from the Pew Research Center of the current undocumented immigrants population living in the U.S. was 12 million (Passel, 2013).[1]

For the purpose of legislating and regulating the legal status of undocumented immigrants, enforcement is done at the federal level and state level and takes mainly two forms: police-based and employment-based. The most significant federal legislation was the Immigration Reform and Control Act (IRCA) of 1986, which provided amnesty to undocumented aliens already in the U.S. and started sanctions for knowingly hiring unauthorized aliens. In more recent years, however, local governments started to play a more and more important role in terms of immigration enforcement. For example, the passage of the Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA) of 1996 added Section 287(g), which allowed the U.S. Immigration and Customs Enforcement (ICE) deputy director to enter into agreements with state and local authorities, to the Immigration National Act (INA). The Florida Department of Public Safety became the first state government to sign the contract in 2002, with additional

[1]The Pew Research Center estimate is based on a residual methodology that compares the number of immigrants residing legally in the country with the total number of immigrants measured by a survey.

jurisdictions signing enforcement contracts in subsequent years.

Located on the border with Mexico, Arizona has many documented and undocumented immigrants and has passed the strictest state laws in terms of undocumented immigrants.[2] The 2007 Legal Arizona Workers Act (LAWA) required all employers to verify the identity and work eligibility of all new hires using the federal E-Verify system. Subsequently, the 2010 Arizona SB1070 made it a crime to apply for or hold a job in Arizona without legal authorization, required police officers to check the immigration status of anyone they believed may be in the country illegally, and allowed police to stop and arrest anyone they had reason to believe lacks proper immigration papers. Such immigration policies determined the number of immigrants allowed in, the selection criteria used to admit them, and the level of resources devoted to controlling undocumented immigration. Hence, understanding the role that immigration enforcement plays in the U.S. as well as the effects they have on the economy and local natives is important for researchers and policy makers.

There has been a great deal of research on the determinants and effects of immigration as a whole (e.g., Card, 2001; Borjas, 1999, 2006; Lewis and Peri, 2015), but less focus on policies affecting undocumented immigrants. The existing literature focusing on immigration enforcement mainly analyze their effects on local population composition (Bohn et al., 2014; Amuedo-Dorantes and Lozano, 2015), location choice (Bohn and Pugatch, 2015) or flow (Hoekstra and Orozco-Aleman, 2017). Some recent studies also focus on the effects on economic resources of children with undocumented

[2] Mexico has historical significance as a source of U.S. immigration, both authorized and unauthorized.

parents (Amuedo-Dorantes et al., 2018), on health and mental health outcomes of Latino immigrants living in the U.S. (Wang and Kaushal, 2018), as well as on political outcomes (Mayda et al., 2018). As for labor market outcomes, Chassamboulli and Peri (2015) theoretically show that reducing the number of undocumented immigrants could weaken low-skilled labor markets and increase unemployment of native low-skilled workers. The first empirical study of the economic impacts of the local immigration regulation is by Pham and Van (2010), who study the effects of county level implementation of Section 287(g) on employment and payroll. They also conduct analyses of different industries, especially those with high immigrant composition. Kostandini et al. (2013) specifically look at the effects of Section 287(g) on U.S. agriculture in dozens of counties.

This paper aims at finding more empirical evidence on the effects of state level enforcement on the presence of undocumented immigrants in the labor force, and hence on labor market outcomes by using the enactment of such strict laws in Arizona as described above. Due to lack of information on legal status of immigrants, most studies focus on the Hispanic noncitizens, the population that is most likely to be undocumented. Bohn et al. (2014) verify a significant exit of Hispanic noncitizens from Arizona after the passage of LAWA in 2007, whereas Amuedo-Dorantes and Lozano (2015) find that the effect of SB1070 in 2010 has been minimal. This paper further explores the effects of LAWA on the composition of the labor force as well as industry heterogeneity in Arizona. I first show that the fraction of Hispanic noncitizens decreases significantly by around 2.3 percentage points in the labor force, and then verify that the change is not due to a decrease in the overall labor force, but rather the replacement of Hispanic noncitizen workers by other subgroups. In order to see which subgroups fill the jobs, I focus on the fraction of Hispanic citizens and Nonhispanic whites as well as the fraction of workers of different education levels. I also look into the change in unemployment rates for each of these subgroups. Since LAWA affects new hires, I also look into the new hire rate and separation rate of firms and find that they tend to reduce both, indicating that they are firing fewer workers instead of hiring more authorized workers in response to LAWA.

This paper uses a difference-in-differences design based on a synthetic control approach. I show that the synthetic control has similar concurrent economic trends with Arizona, using aggregate data from the American Community Survey (ACS, 2001-2016 waves).

The paper proceeds as follows: Section 2 provides background about trends in the undocumented population in the U.S. and the implementation of LAWA in Arizona. Section 3 reviews related literature regarding both theoretical and empirical analyses about undocumented immigrants. Sections 4 and 5 describe the data and identification strategy used in the paper. Sections 6 and 7 present the results and robustness checks. Section 8 concludes and provides direction for future analyses.

2.2 Related Literature

This part introduces two strands of literature regarding the analyses of the labor market effects of immigration, especially targeting at undocumented immigrants in the United States. The first strand includes the theoretical modeling of the effects of immigrants on local natives. Borjas (2003) adopts the neoclassical labor demand-supply approach and introduces skill cells, and finds negative effects on the wage of less educated natives. Ottaviano and Peri (2012) then expand it by estimating the substitutability between natives and immigrants of similar education and experience levels, and find small positive effects on the wage of average natives. In terms of the undocumented immigrants, Liu (2010) uses a dynamic general equilibrium model with labor market frictions and finds that an increase in undocumented immigration can generate significant welfare gains for the natives. Chassamboulli and Peri (2015) set up a two-country model with search in the labor market and feature documented and undocumented immigrants among the low skilled in order to figure out the labor market effects of reducing the number of undocumented immigrants. According to their results, the unskilled immigrants receive lower pay and generate higher surplus for the firm than unskilled native workers because of worse outside options. This in turn pushes firms to create more jobs per unemployed when there are more immigrants, improving the tightness of the labor market and reducing the unemployment rate of the natives.

The second strand includes the corresponding empirical analyses. This paper adds to the "natural experiment" approach of studying the effects of immigration, pioneered by Card (1990). The other often used approach is the Bartik (1991) style instrument for demand shocks. Pham and Van (2010) is the first study towards the economic impacts of local anti-immigration laws. They use the County Business Patterns (CBP) data set and find a 1 to 2 percent drop in employment for both authorized and unauthorized workers. According to their results, the laws hurt some industries, such as restaurant while helping others, such as grocery and liquor store industry. Since agriculture in the United States is highly reliant on immigrant workers (Seid 2006; Levine 2009), Kostandini et al. (2013) take a closer look at the effects of local immigration enforcement on U.S. agriculture in dozens of U.S. counties, utilizing a quasi-experiment provided by local variation in the timing of adopting 287(g) programs as well as combining individual level data from the American Community Survey (ACS) and the county-level tabulations of farm survey data from the U.S. Census of Agriculture. Their study yields evidence that county enforcement reduces immigrant presence in adopting jurisdictions and the wages of farm workers, patterns of farm labor use, output choices and farm profitability are found to be effected in a manner consistent with farm labor shortage.

There are also empirical analyses specifically towards Arizona. Bohn et al. (2014) study the aggregate population movement caused by the 2007 Legal Arizona Workers Act (LAWA) and document a notable and statistical significant reduction in the proportion of the Hispanic noncitizen population in Arizona. Their analysis is based on the monthly Current Population Survey (CPS) data sets and the synthetic control method developed by Abadie et al. (2010). Amuedo-Dorantes and Lozano (2015) use the similar data and method to study the effects of the other anti-immigration law Arizona SB 1070 and conclude a minimal effect on the stock of Hispanic noncitizen. Hoekstra and Orozco-Aleman (2017) then employ a unique data set from the Survey of Migration to the Northern Border (EMIF) and claim that the passage of SB 1070 reduce the flow of undocumented immigrants into Arizona by 30 to 70 percent.

2.3 Background

2.3.1 Immigrants in the U.S. and Arizona

According to the American Community Survey (ACS) data, more than 42.4 million documented and undocumented immigrants resided in the United States in 2014, accounting for 13.3 percent of the total U.S. population of 318.8 million.[3] In addition to immigrants, the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) reported that there were 16.2 million U.S.-born minor (¡18) children with at least one immigrant parent in 2014, for a total of 58.6 million immigrants and their children. In 2014, an estimated 11.1 million undocumented immigrants lived in the U.S., unchanged since 2009 and down from a peak of 12.2 million in 2007 (Pew Research Center). Mexico had by far the largest immigrant population in the country, with 11.7 million documented and undocumented Mexican immigrants living in the U.S. in 2014. There were 5.6 million Mexican undocumented immigrants living in the U.S. in 2015 and 2016, down from 6.4 million in 2009.

According to reports from the American Immigration Council, 914,400 immigrants comprised 13.4 percent of the Arizona's population in 2015. Among all the immigrants, roughly 325,000 undocumented immigrants comprised 35 percent of the immigrant population and 4.9 percent of the total state population in 2014. As for education levels, around 36.9 percent of all immigrants held less than a high-school diploma, while the number for natives was 9 percent in Arizona in 2015. Besides, more than a

[3]The Census Bureau refers to all immigrants as foreign-born.

quarter-million U.S. citizens in Arizona were living with at least one family member who is undocumented.

Based on the above statistics, by requiring E-verify of the identity for all new hires, LAWA would be expected to cause the exit of both documented and undocumented immigrants in Arizona, as well as the decrease in the incoming immigrants to the state. Due to the large component of Mexican immigrants, the Hispanic noncitizen population would be the group of people that were mostly affected by this regulation.

2.3.2 LAWA

Enacted in September 1996, the Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA) first established the Basic Pilot Program to test the feasibility of electronically verifying the work-authorization status of all newly hired employees. The E-Verify Program started from a web version of the Basic Pilot Program in June 2004, and was voluntary for employees other than some federal agencies. However, state legislation has expanded the mandatory use of E-verify and the most comprehensive one was the Legal Arizona Workers Act (LAWA), enacted on July 2, 2007, requiring all Arizona employers to verify new employees through E-Verify starting on January 1, 2008. The law prohibits businesses from knowingly hiring an "unauthorized alien", defined as "an alien who does not have the legal right or authorization under federal law to work in the United States." Violation of the law would induce suspension of business licenses, but there was no penalties for failing to use E-Verify.

The implementation of LAWA has seen immediate effects on the hiring process.

The number of employers registered with E-verify in Arizona increased from only a few in March 2007 to around one quarter of all employers in January 2010, which accounts for one-third of the nationwide registrations (Westat, 2009). And half of all new hires between October 2008 and September 2009 in Arizona were run through E-verify (Berry, 2010).

2.4 Data Description

The main source of data for this study is from the U.S. Census Bureau's American Community Survey (2001-2016 waves), tabulated from the public-use file obtained from the Integrated Public-Use Microdata Series or IPUMS (Ruggles et al., 2015). Both of the two U.S. Census Bureau surveys, the Current Population Survey, or CPS, and the American Community Survey, or ACS, ask people their Hispanic origin as well as citizenship status, which can identify the most ”likely unauthorized” subgroup of population. In order to estimate the number of immigrants in the U.S., the Pew Research Center uses the ACS data for 2005 and later years and the CPS data for before 2005. The reason is that starting from 2005, ACS interviews about 3 million people a year (1% of the entire country). The density of the ACS from 2001 to 2004 waves are around 0.4% of entire population. However, the CPS only interviews about 55,000 households a month and increases the sample size to 80,000 every March, which can bring sizable margins of error. An important concern of using the survey data to estimate the number of the "likely unauthorized" is that the response rate of the this group of people may be lower than the other groups. Though it may not be an issue if there is no difference in the response rates across the country in this empirical setting, prior research by the Department of Homeland Security and others indicate that some 90 percent of undocumented immigrants respond to the ACS. The remaining issue is LAWA would lower the response rate in Arizona, making it to be different with the rates in the control group. Then the estimate will be a lower bound of the true effect. Since LAWA is employment-based instead of police-based, I assume that it will not change the response rate of the Hispanic noncitizen in Arizona. The data only includes information on employment status as well as specific industries. Based to the 1990 classification, I focus on the following ten industries: Agriculture, Construction, Manufacturing, Utilities, Wholesale and retail trade, Finance, Business, Leisure, Professional services, Public Administration.

The second source of data is from the Census Bureau's Building Permit Survey, which provides data on the number of new housing units authorized by building permits. I collect the 2001-2016 waves state level information on the total number of new residential housing units, divided by the total population in the state for the same year. The goal is to check whether Arizona was hit differently by the 2008 economic downturn than the control group.

The third source of data is the public-use Quarterly Workforce Indicators (QWI) data, which is from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD). The QWI includes information on employment levels and flows, which can shed light on whether changes in employment is due to changes in hiring or separation. So I look at the new hire rate and separation rate of workers by Hispanic origin in Arizona and the control group.[4] The data is collected from 2004 to 2015 waves.[5]

2.5 Empirical Methodology

Similar with Bohn et al. (2014) and Amuedo-Dorantes and Lozano (2015), this paper uses the synthetic control method developed by Abadie et al. (2010). Synthetic control builds upon the standard difference-in-differences model but allows for time-varying state-specific heterogeneity. It takes a data driven approach to select a group of states that can form a counterfactual post-LAWA path for Arizona, which shares similar trends in terms of the outcome and other characteristics. The setting of this paper is similar with the original synthetic control model in Abadie et al. (2010), which has a single treated unit with multiple controls.

Suppose there are $J+1$ states over $t = 1, ...T$ periods, with the first state being treated and the states $2, ...J+1$ being unaffected. An intervention occurs at period $T_0 + 1$, $1 < T_0 + 1 < T$ and affects the first state only. Suppose $Y_{i,t}^N$ is the outcome that would be observed for state i at time t in the absence of the intervention, and $Y_{it}^I$is the outcome that would be observed for state i at time t if state i is exposed to the intervention in periods $T_0 + 1$ to T. The aim is to estimate the effect of the intervention on the treated state $(\alpha_{1T_0+1}, ..., \alpha_{1T})$, where $\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N$

[4]The data does not include information on citizenship status.

[5]The data for Arizona starts from 2005, and for Massachusetts starts from 2011.

for $t > T_0$. In order to estimate Y_{1t}^N, I first assume that Y_{it}^N is given by a factor model:

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it}$$

where δ_t is an unknown common factor, Z_i is a vector of observed covariates, λ_t is a vector of unobserved common factors, and μ_i is a vector of unknown factor loadings. The term $\lambda_t \mu_i$ represents heterogeneous responses to multiple unobserved factors and the basic idea is to find a convex combination of the J states in the donor pool that matches some pre-treatment outcomes plus additional covariates predictive of the outcome, then μ_i is automatically matched. So let X_1 be a $K \times 1$ vector that includes the pre-intervention values of the outcome and covariates for the treated state. Let X_0 be a K matrix as a collection of comparable data vector for each of the J states in the donor pool. Then define a $J \times 1$ vector $W = (w_1, w_2, ...w_J)'$ and W^* is chosen to minimize some distance between X_1 and $X_0 W$, $\parallel X_1 - X_0 W \parallel$, subject to $\sum_{j=1}^{J} w_j = 1$ and $w_j \geq 0$ for $j = (1, ..., J)$. The distance is measured by $\parallel X_1 - X_0 W \parallel_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$, where V is some $(K \times K)$ diagonal and positive-definite matrix providing the relative weights for the contribution of the variables in X_1 and X_0 to minimize the mean squared prediction error of the outcome variable in the pre-intervention periods.[6] Once the optimal W^* is chosen, the "synthetic control" is formed, and the post-intervention counterfactual outcome of the treated can be calculated using the weighted average of the values from the states with positive weights. And the

[6]The choice of V could be based on subjective assessments of the predictive power of X, data-driven, cross-validation, etc. The Stata procedure developed by Abadie et al. (2010) uses as the default a regression-based measure of V that assigns more weight to the matching variables strongly predictive of the dependent variable and the elements of V are normalized to sum to 1.

difference-in-differences estimate can also be obtained with the following regression:

$$y_{it} = \beta_t + \beta_1 Arizona_{it} + \beta_2 post_{it} + \beta_3 Arizona_{it} * post_{it} + u_{it}$$

The inference techniques associated with synthetic control is based on a placebo study or permutation test, when the large sample asymptotic framework for assessing the properties of estimators are not well suited. Specifically, for each state in the donor pool, I repeat the minimization procedure to identify a synthetic procedure as if these states also faced the same intervention. Then the distribution of the these place estimates provides the equivalent of a sampling distribution for the original estimate of the treated.

In the analysis of LAWA as an intervention that was implemented in the beginning of 2008 in Arizona, I first select the donor pool of states that may potentially become part of the control for Arizona. In order that Arizona is the only treated, the states that passed similar policy that restricts the employment of the unauthorized including Mississippi, Rhode Island, South Carolina, and Utah are excluded from the donor pool. I also omit D.C and the remaining 45 states form the donor pool. The main dependent variable is the fraction of Hispanic noncitizen in the labor force.

Table 2.1: Means of Matching Variables

Variables	Arizona	Synthetic Control
Fraction of Hispanic noncitizen in labor force, 2002	0.120	0.127
Fraction of Hispanic noncitizen in labor force, 2005	0.139	0.125
Fraction of Hispanic noncitizen in population	0.097	0.098
	0.108	0.104
Labor force*	2,532	2,800
	2,825	3,042
Labor force participation rate	0.633	0.652
	0.633	0.655
Unemployment rate	0.074	0.071
	0.059	0.062

*Measured in thousands.

Table 2.2: State Weights in the Synthetic Control

State	Weight
California	0.036
Florida	0.174
Nevada	0.598
New Mexico	0.191

Table 2.1 shows the means of the main matching variables of Arizona and the synthetic control. (The other variables used for matching are listed in the Appendix Tables 2.A1 and 2.A2.) Table 2.2 displays the optimal weights chosen based on the choice of the dependent variable and matching variables. Since the data is from 2001 to 2016 and the intervention was in 2008. I first match on two lagged dependent variables in 2002 and 2005. The result is similar when I use 2003 and 2006 as two lagged variables to match. Bohn et al. (2014) use all the pre-intervention values and find the inclusion of

other covariates does not change the result. However, suggested by Kaul et al. (2015), using all outcome lags as separate predictors renders all other covariates irrelevant and threatens the estimator's unbiasedness. So I include only two years in the matching variables. Since I am also focusing on industry heterogeneity, I also match on the number of workers as well as fractions of Hispanic noncitizen in different industries. The matching variables (other than the lagged variables) are averaged from 2001 to 2003 (in the first row) and 2004 to 2006 (in the second row). The result is also robust to the inclusion of average values of the proportion of the state population in each of four broad educational attainment categories (less than high school, high school graduate, some college, and college or more).

2.6 Results

2.6.1 Validity of Synthetic Control

The main concern of the effects on labor market is the coincidence of LAWA with the Great Recession in 2008, which could also drive changes in employee composition. In order that the post-intervention outcomes of synthetic control displayed in Table 2.2 can serve as the counterfactual Arizona in the absence of LAWA, I show that Arizona was hit by the downturn similarly with the synthetic control. As the recession was precipitated by a housing crisis, especially causing stagnant on the new housing construction. I compare the trends of the number of new residential housing units per-capita constructed during the same time period, 2001 to 2016, in Arizona and the synthetic control.The results are shown in Figure 2.1. The top panel shows the trends in Arizona and the four states in the donor pool that have positive weights. The bottom panel shows the trends in Arizona and the synthetic control calculated by the weighted average of these states. In terms of the decrease in the number of new constructions, Florida and Nevada were hit at lease as heavily as Arizona. California and New Mexico were less severely affected. The synthetic control, which gives higher weighs to Florida and Nevada and lower weights to California and New Mexico, has similar trend with Arizona.

2.6.2 Composition in the Labor Force

As LAWA directly affects undocumented immigrants, the main analysis focus on the composition of Hispanic noncitizen, the group of population that are mostly likely to be undocumented, in the labor force. Figure 2.2 shows the trends of the fraction of Hispanic noncitizen in the labor force. The top panel indicates Arizona and all 45 states in the donor pool, where Arizona displays a notable decrease. And the middle panel shows Arizona and the four states that form the synthetic control. The bottom panel compares the trends for Arizona with the synthetic control, where there is similar pre-intervention trends and a significant decrease in Arizona after the intervention. In order to get the magnitude and significance level of the effect, I run the difference-in-differences regression, which estimates the change in the average fraction of Hispanic noncitizen in the labor force in Arizona as a result of LAWA. The time window that I use is from 5 years before year 2008 to 5 years after year 2008. The result is shown in the first column of Table 2.3. There is a 2.3 percentage points decrease in the fraction of Hispanic noncitizen in the labor force, which ranks the first among the all the 46 placebo tests, so the p-value from a one-tailed test of the likelihood of observing an estimate at least as negative as Arizona is around 0.022.

In order to see how the employment of Hispanic noncitizen is replacement by other subgroups. I look into the fractions of other main groups of population including the Hispanic citizen and the Nonhispanic white in the labor force and the results are shown in columns 2 and 3 of Table 2.3. The compositional change in the labor force in Arizona is driven by the replacement of Hispanic noncitizen by Hispanic citizen (0.5 percentage point and rank the second) and Nonhispanic white (2.1 percentage points and rank the third). Then columns 4 and 5 of Table 2.3 show the replacement by education levels. The low education refers to non-college degree and the low-educated workers seem to replace more of the jobs (1.9 percentage points and rank the first) than the high-educated workers with a college degree (0.4 percentage point and rank the seventh). Since I don't match on the pre-treatment values of the fraction of subgroups, I check the pre-trends between Arizona and the synthetic control. Appendix Figures 2.A1 and 2.A2 show that the fraction of all the four subgroups in the labor force have similar pre-trends between Arizona and the synthetic control, so the replacement results above are caused by the intervention.

To dig into the mechanisms of the effects on worker composition. I first look at the change in the fraction of Hispanic noncitizen in the population. Figure 2.4 shows that it decreases significantly in Arizona, which means the policy not only causes the Hispanic noncitizen to leave the labor market but to leave the state. I then look at changes in labor force participation rates and unemployment rates across different subgroups. Table 2.5 shows that there is no significant change in labor participation across different workers. Figure 2.5 shows a drop in the overall unemployment rate in Arizona, whereas the first column of Table 2.6 indicates that the change is not significant. The other columns of Table 2.6 indicate that the unemployment rate for all the subgroups tend to decrease, but none of them is significant. The drop in the fraction of Hispanic noncitizen in the labor force is probably driven by the exit of this group of population, especially those that are unemployed.

2.6.3 Industry Heterogeneity

Figure 2.3 and Table 2.4 show the changes of the fraction of Hispanic noncitizen by ten industries, including agriculture, construction, manufacturing, utility, trade, finance, business, leisure, professional and public. I find that there are no significant changes in the overall employment in all the ten industries, but nearly all industries experience significantly drop in fraction of workers that are Hispanic noncitizen (from around 10 percent to more than 20 percent). The replacement of these workers by other subgroups are shown in Tables 2.7 to 2.16. The industry heterogeneity could be explained by the likelihood of the Hispanic noncitizen to be unauthorized in each industry. It could also be related with the difference in worker turnovers, which are shown in Section 6.4.

2.6.4 Worker Turnover

Furthermore, by looking into the stock and flow of workers with the QWI data, I investigate through what process (hiring or separation) does the policy affect the composition in the labor force. Since the data has no information on citizenship status, I look into the new hire rate as well as the separation rate by Hispanic origin. The results are displayed in Figures 2.6 to 2.8, where the top panels show the new hire rate and the bottom panels show the separation rate, and the left panels show the absolute values and the right panels show the corresponding demeaned values. Overall, there are both lower new hire rate and separation rate in Arizona after the policy, and the drop in new hire rate is especially larger among the Hispanic. Both Hispanic and Nonhispanic experience lower separation rates. So LAWA affects worker turnovers and as a result, firms tend to hire less new workers and reduce the separation rate to maintain the employment level. Accordingly, by looking into the age distribution of workers in the labor market, Appendix Figure 2.A3 indicates that there is an increase in the proportion of workers that are older than 30 years old in Arizona.

2.7 Robustness

First, the results may change due to the definition of "undocumented immigrant". The legal status of immigrants is hard to detect and verify, hence researchers have been working on the methodology to credibly identify and enumerate the size of undocumented immigrants based on the residual approach framework first advanced by

Warren and Passel (1987). Some recent improvements in the methodology (Passel et al., 2014) led to the creation of a "likely unauthorized" identifier to the Annual Social and Economic Supplement (ASEC) files of the Current Population Survey (CPS) since 2012 (Borjas, 2017). The use of survey data or census data from Mexico will also serve as a check of the estimation of undocumented immigrants in the U.S.

Second, there is concern about the spillover of the Hispanic noncitizen to other states. Westat (2010) conducts a case study consisting of a stakeholders meeting, on-site visits in Arizona, and analysis of data in the E-Verify Transaction Database and employer database, in order to make the evaluation report to the U.S. Department of Homeland Security. In response to the question regarding whether and where do the workers move or plan to move as a result of LAWA, around one third of the workers choose Mexico, around one third choose the states that form the synthetic control (California, Nevada, Florida and New Mexico), and the other one third choose the other states such as Colorado, Utah and Texas. We may worry about the fact that Hispanic noncitizen move to the control states, which will bias the results. However, since Arizona has a relatively smaller population than the control states, and I don't observe an increase in the fraction of Hispanic noncitizen in the synthetic control, this concern becomes less important.

Third, the synthetic control method has a relatively subjective way of choosing the matching variables, so several robustness checks need to be done to show that the estimates are robust to the exclusion and inclusion of some variables. I mentioned some of them in Section 5. Instead, the dependent variable plays an important role when choosing the weights, and I am using the same weight chosen by using the fraction of Hispanic noncitizen in the labor force as the dependent variable for other outcomes as well. Appendix Figure 2.A4 and Appendix Table 2.A3 show the results of industry heterogeneity when using the fractions in the ten industries as dependent variables, and the results are similar.

2.8 Discussion

This paper verifies a significant drop in the fraction of Hispanic noncitizen in the labor force in Arizona as a result of LAWA. It looks into industry heterogeneity, explores the indirect effects on the other subgroups of the population, and identifies possible mechanisms that are related with the findings, which include the exit of Hispanic noncitizen, the deterrence in the incoming immigrants, becoming self-employed as well as becoming unemployed. And more precisely, the changes could happen to either unemployed or employed, either new workers or existing workers. The future work includes more comprehensive analyses towards the mechanisms and relating the empirical results to the theoretical models.

References

A. Abadie, A. Diamond, and J. Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

C. Amuedo-Dorantes and F. Lozano. On the effectiveness of sb1070 in arizona. *Economic inquiry*, 53(1):335–351, 2015.

C. Amuedo-Dorantes, E. Arenas-Arroyo, and A. Sevilla. Immigration enforcement and economic resources of children with likely unauthorized parents. *Journal of Public Economics*, 158:63–78, 2018.

T. J. Bartik. Who benefits from state and local economic development policies? 1991.

S. Bohn and T. Pugatch. Us border enforcement and mexican immigrant location choice. *Demography*, 52(5):1543–1570, 2015.

S. Bohn, M. Lofstrom, and S. Raphael. Did the 2007 legal arizona workers act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics*, 96(2):258–269, 2014.

G. J. Borjas. The economic analysis of immigration. In *Handbook of labor economics*, volume 3, pages 1697–1760. Elsevier, 1999.

G. J. Borjas. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The quarterly journal of economics*, 118(4): 1335–1374, 2003.

G. J. Borjas. Native internal migration and the labor market impact of immigration. *Journal of Human resources*, 41(2):221–258, 2006.

G. J. Borjas. The labor supply of undocumented immigrants. *Labour Economics*, 46: 1–13, 2017.

D. Card. The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43 (2):245–257, 1990.

D. Card. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64, 2001.

A. Chassamboulli and G. Peri. The labor market effects of reducing the number of illegal immigrants. *Review of Economic Dynamics*, 18(4):792–821, 2015.

M. Hoekstra and S. Orozco-Aleman. Illegal immigration, state law, and deterrence. *American Economic Journal: Economic Policy*, 9(2):228–52, 2017.

A. Kaul, S. Klößner, G. Pfeifer, and M. Schieler. Synthetic control methods: Never use all pre-intervention outcomes as economic predictors. *Unpublished. URl: http://www. oekonometrie. uni-saarland. de/papers/SCM_ Predictors. pdf*, 2015.

G. Kostandini, E. Mykerezi, and C. Escalante. The impact of immigration enforcement on the us farming sector. *American Journal of Agricultural Economics*, 96(1):172–192, 2013.

E. Lewis and G. Peri. Immigration and the economy of cities and regions. In *Handbook of regional and urban economics*, volume 5, pages 625–685. Elsevier, 2015.

X. Liu. On the macroeconomic and welfare effects of illegal immigration. *Journal of Economic Dynamics and Control*, 34(12):2547–2567, 2010.

A. M. Mayda, G. Peri, and W. Steingress. The political impact of immigration: Evidence from the united states. Technical report, National Bureau of Economic Research, 2018.

G. I. Ottaviano and G. Peri. Rethinking the effect of immigration on wages. *Journal of the European economic association*, 10(1):152–197, 2012.

J. S. Passel. D'vera cohn, and ana gonzalez-barrera. 2012. *Net Migration from Mexico Falls to Zero—and Perhaps Less*, 2013.

J. S. Passel and S. W. D'Vera Cohn. *Unauthorized immigrant population: National and state trends, 2010.* Pew Hispanic Center Washington, DC, 2011.

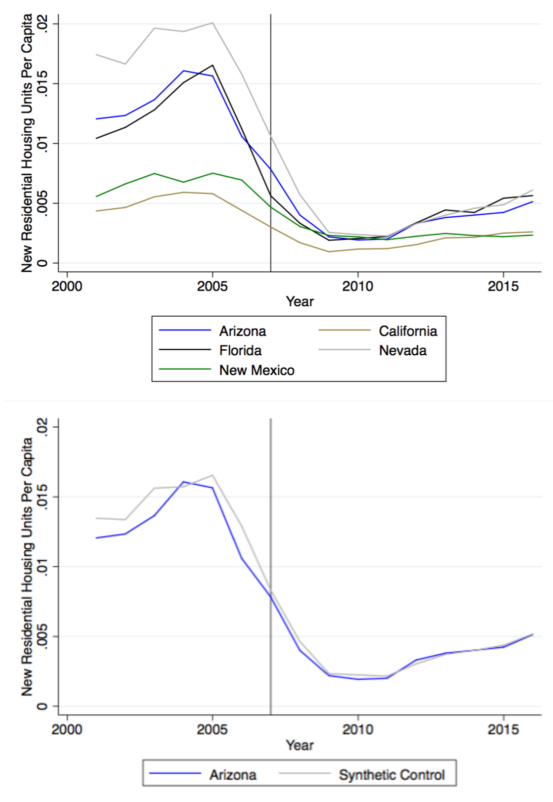
J. S. Passel, D. Cohn, J. M. Krogstad, and A. Gonzalez-Barrera. As growth stalls, unauthorized immigrant population becomes more settled. *Washington: Pew Research Center Hispanic's Trend Project*, pages 1–25, 2014.

H. Pham and P. H. Van. Economic impact of local immigration regulation: an empirical analysis. *Immigr. & Nat'lity L. Rev.*, 31:687, 2010.

J. S.-H. Wang and N. Kaushal. Health and mental health effects of local immigration enforcement. Technical report, National Bureau of Economic Research, 2018.

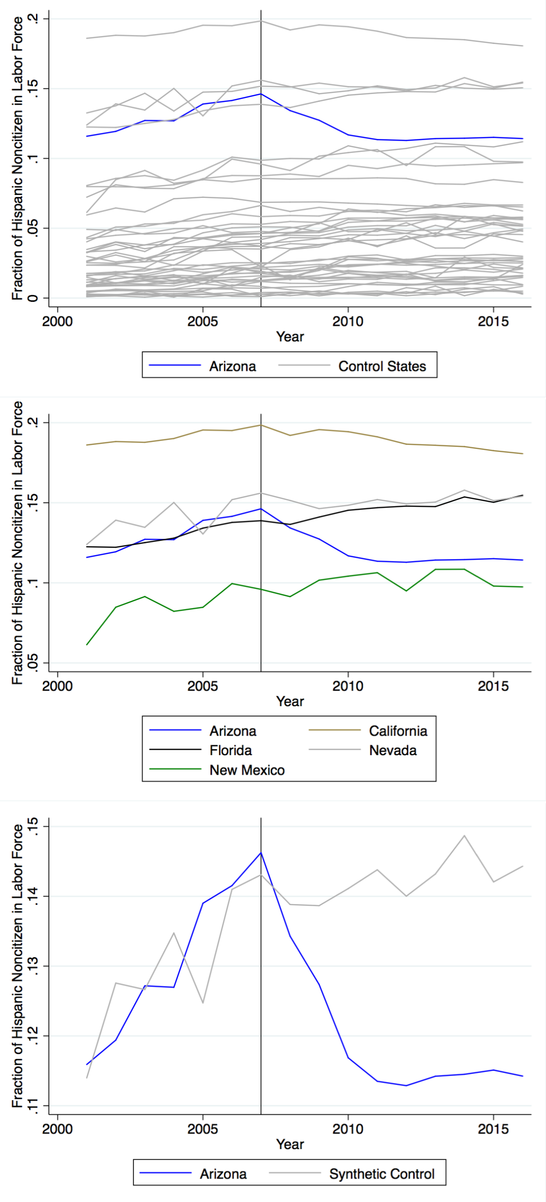
Figures and Tables

Figure 2.1: Trends in the Number of New Residential Housing Units Per-Capita



Note: The top panel shows the trends in Arizona and the four states in the donor pool that have positive weights. The bottom panel shows the trends in Arizona and the synthetic control.

Figure 2.2: Trends in the Fraction of Hispanic Noncitizen in the Labor Force



Note: The top panel shows the trends in Arizona and all 45 states in the donor pool. The middle panel shows the trends in Arizona and the four states in the donor pool that have positive weights. The bottom panel shows the trends in Arizona and the synthetic control.

Figure 2.3: The Percentage Decrease in the Fraction of Hispanic Noncitizen by Industry

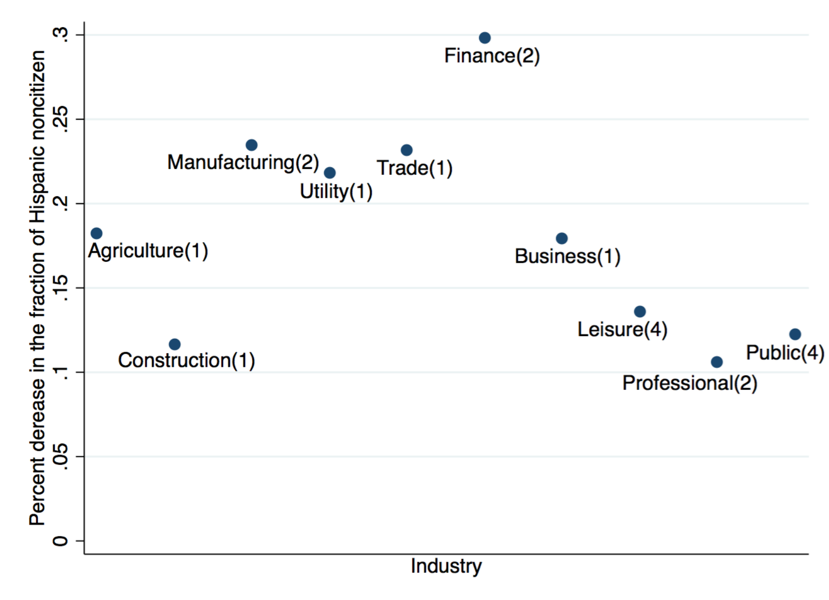
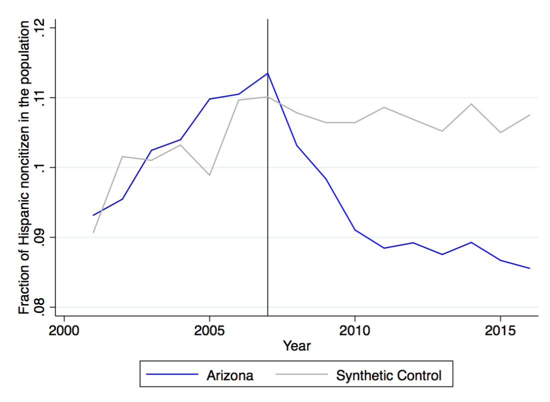


Figure 2.4: Trends in the Fraction of Hispanic Noncitizen in the Population



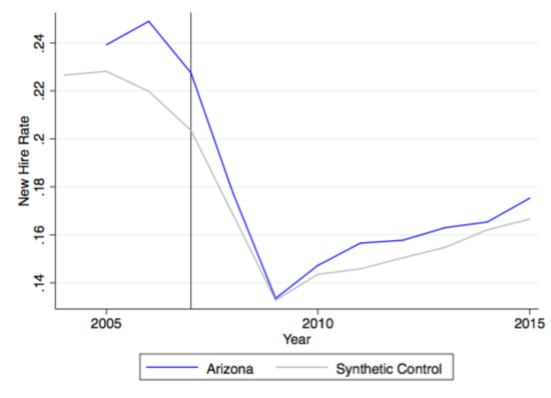
Note: The figure shows the trends in Arizona and the synthetic control.

Figure 2.5: Trends in the Unemployment Rate

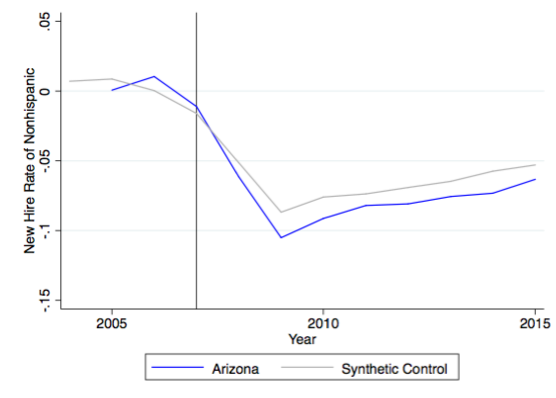
Note: The top panel shows the trends in Arizona and the four states in the donor pool that have positive weights. The bottom panel shows the trends in Arizona and the synthetic control.

Figure 2.6: Trends in the Overall New Hire Rate and Separation Rate

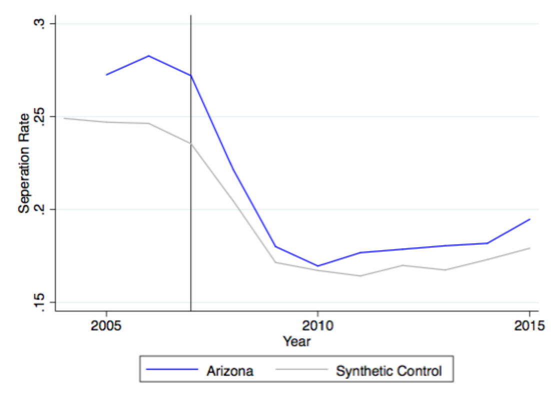
(a) New Hire Rate, absolute values



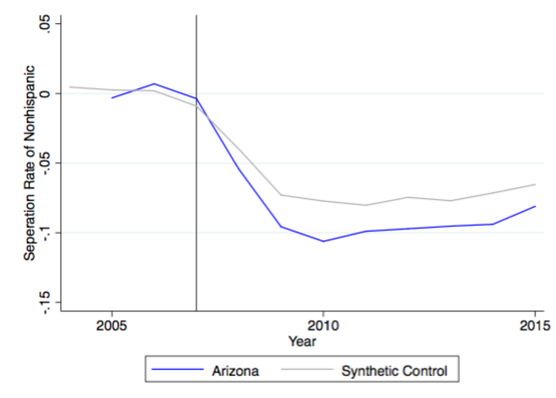
(b) New Hire Rate, demeaned values



(c) Separation Rate, absolute values



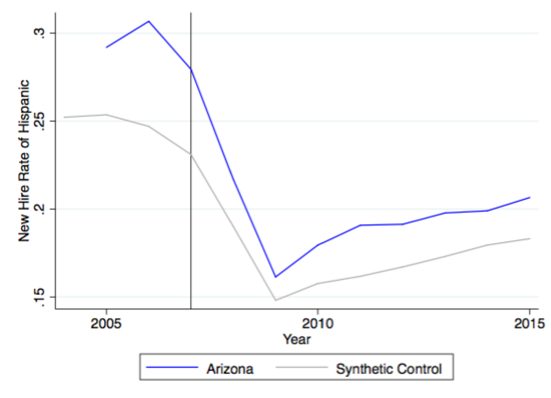
(d) Separation Rate, demeaned values



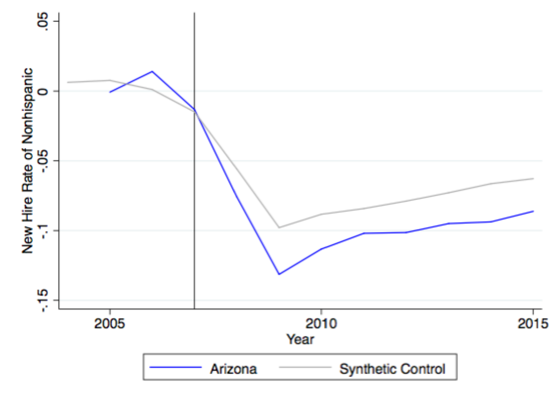
Note: The top panels show the new hire rate and the bottom panels show the separation rate, and the left panels show the absolute values and the right panels show the corresponding demeaned values.

Figure 2.7: Trends in the Overall New Hire Rate and Separation Rate of Hispanic

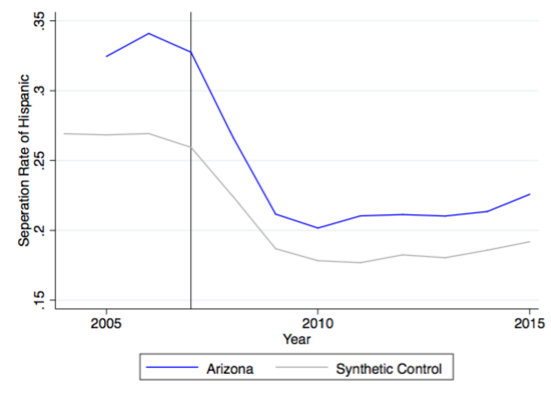
(a) New Hire Rate, absolute values



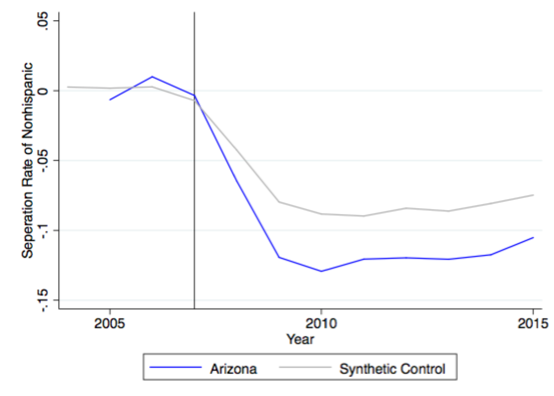
(b) New Hire Rate, demeaned values



(c) Separation Rate, absolute values



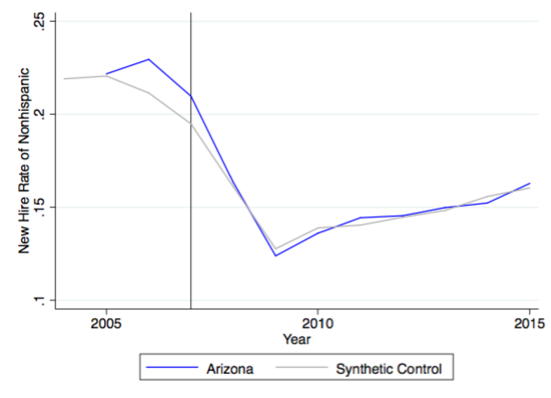
(d) Separation Rate, demeaned values



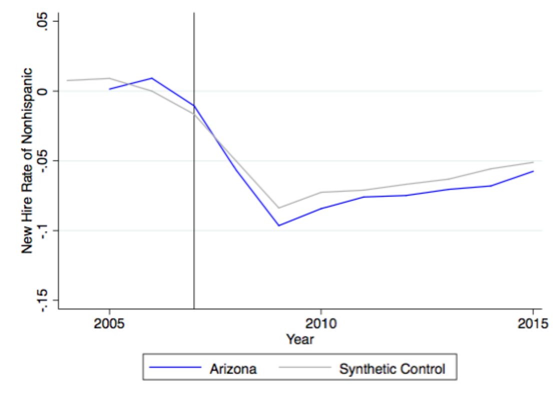
Note: The top panels show the new hire rate and the bottom panels show the separation rate, and the left panels show the absolute values and the right panels show the corresponding demeaned values.

Figure 2.8: Trends in the Overall New Hire Rate and Separation Rate of Nonhispanic

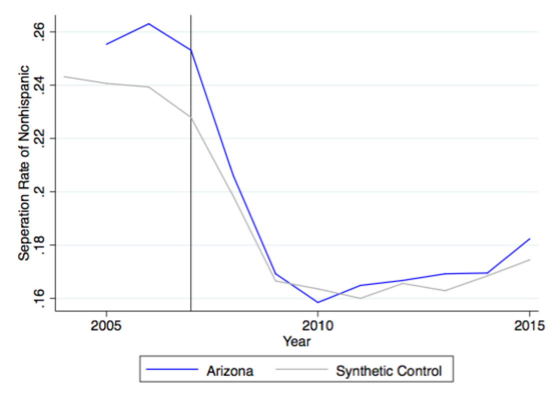
(a) New Hire Rate, absolute values



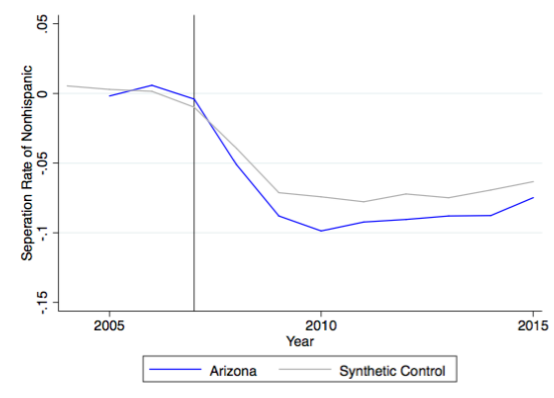
(b) New Hire Rate, demeaned values



(c) Separation Rate, absolute values



(d) Separation Rate, demeaned values



Note: The top panels show the new hire rate and the bottom panels show the separation rate, and the left panels show the absolute values and the right panels show the corresponding demeaned values.

Table 2.3: Changes in Composition in the Labor Force

	Fraction in the labor force				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.002	0.006	0.047	0.027	-0.0293
	(0.006)	(0.026)	(0.018)	(0.007)	(0.005)
Arizona_post	-0.023	0.005	0.021	0.004	0.019
	(0.008)	(0.034)	(0.022)	(0.009)	(0.006)
Observations	55	55	55	55	55
R-Squared	0.256	0.042	0.420	0.526	0.365
Mean	0.136	0.131	0.633	0.238	0.626
Rank	1/46	2/46	3/46	7/46	1/46

Standard errors in parentheses

Table 2.4: Changes in Composition in the Labor Force by Industry

	Fraction of Hispanic noncitizen			Fraction in the population		
	Arizona_post*	Mean	Rank	Arizona_post*	Mean	Rank
Agriculture	-0.103(18.20%)	0.566	1/46	0.001(4.00%)	0.025	15/46
Construction	-0.038(11.59%)	0.328	1/46	-0.002(1.96%)	0.102	20/46
Manufacturing	-0.037(23.42%)	0.158	2/46	-0.003(3.49%)	0.086	18/46
Utility	-0.017(21.79%)	0.078	1/46	-0.001(1.52%)	0.066	23/46
Trade	-0.031(23.13%)	0.134	1/46	-0.009(4.25%)	0.212	3/46
Finance	-0.014(29.79%)	0.047	2/46	0.004(5.19%)	0.077	4/46
Business	-0.027(17.88%)	0.151	1/46	-0.001(1.41%)	0.071	14/46
Leisure	-0.022(13.58%)	0.162	4/46	-0.002(3.33%)	0.06	10/46
Professional	-0.006(10.53%)	0.057	2/46	0.012(5.13%)	0.234	3/46
Public	-0.005(12.20%0	0.041	4/46	-0.000(0.00%)	0.052	23/46

*Arizona_post reports the difference-in-differences estimates based on the synthetic control.

Table 2.5: Changes in the Labor Participation Rate

	Labor participation rate					
	Overall	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	-0.023	-0.028	-0.013	-0.021	-0.027	-0.024
	(0.008)	(0.012)	(0.010)	(0.008)	(0.006)	(0.008)
Arizona_post	-0.008	-0.028	-0.010	-0.005	0.003	-0.006
	(0.011)	(0.017)	(0.013)	(0.011)	(0.009)	(0.011)
Observations	55	55	55	55	55	55
R-Squared	0.519	0.466	0.533	0.480	0.537	0.537
Mean	0.631	0.677	0.678	0.616	0.711	0.597
Rank	5/46	9/46	13/46	11/46	21/46	8/46

Standard errors in parentheses

Table 2.6: Changes in the Unemployment Rate

	Unemployment rate					
	Overall	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	-0.002	0.009	-0.002	-0.004	-0.003	-0.002
	(0.001)	(0.005)	(0.005)	(0.001)	(0.002)	(0.002)
Arizona_post	-0.010	-0.012	-0.002	-0.011	-0.010	-0.009
	(0.003)	(0.006)	(0.007)	(0.004)	(0.003)	(0.003)
Observations	55	55	55	55	55	55
R-Squared	0.951	0.896	0.899	0.881	0.857	0.955
Mean	0.062	0.067	0.081	0.050	0.027	0.074
Rank	9/46	15/46	21/46	8/46	4/46	11/46

Standard errors in parentheses

Table 2.7: Changes in Composition in Agriculture

	Agriculture				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.163	0.010	-0.147	-0.032	-0.132
	(0.026)	(0.020)	(0.014)	(0.008)	(0.028)
Arizona_post	-0.103	0.022	0.075	0.012	0.091
	(0.035)	(0.024)	(0.019)	(0.013)	(0.035)
Observations	55	55	55	55	55
R-Squared	0.516	0.170	0.688	0.331	0.441
Mean	0.566	0.099	0.292	0.076	0.358
Rank	1/46	2/46	1/46	9/46	1/46

Standard errors in parentheses

Table 2.8: Changes in Composition in Construction

	Construction				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.033	0.002	-0.019	0.000	-0.034
	(0.018)	(0.027)	(0.021)	(0.003)	(0.018)
Arizona_post	-0.038	0.001	0.035	0.010	0.028
	(0.020)	(0.037)	(0.030)	(0.004)	(0.020)
Observations	55	55	55	55	55
R-Squared	0.486	0.039	0.082	0.632	0.388
Mean	0.328	0.119	0.492	0.075	0.597
Rank	1/46	13/46	1/46	7/46	1/46

Standard errors in parentheses

Table 2.9: Changes in Composition in Manufacturing

	Manufacturing				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	-0.009	0.012	0.016	0.094	-0.086
	(0.006)	(0.025)	(0.027)	(0.006)	(0.007)
Arizona_post	-0.037	0.005	0.030	0.009	0.028
	(0.011)	(0.033)	(0.033)	(0.012)	(0.010)
Observations	55	55	55	55	55
R-Squared	0.431	0.031	0.124	0.843	0.632
Mean	0.158	0.128	0.608	0.286	0.556
Rank	2/46	8/46	6/46	13/46	3/46

Standard errors in parentheses

Table 2.10: Changes in Composition in Utility

	Utility				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	-0.005	0.016	0.075	0.019	-0.014
	(0.009)	(0.028)	(0.025)	(0.008)	(0.012)
Arizona_post	-0.017	-0.006	0.014	0.010	0.007
	(0.012)	(0.040)	(0.031)	(0.011)	(0.016)
Observations	55	55	55	55	55
R-Squared	0.167	0.034	0.504	0.444	0.191
Mean	0.078	0.140	0.691	0.184	0.738
Rank	1/46	8/46	6/46	8/46	21/46

Standard errors in parentheses

Table 2.11: Changes in Composition in Trade

	Trade				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.010	0.002	0.057	0.009	-0.020
(0.004)	(0.028)	(0.022)	(0.003)	(0.006)	
Arizona_post	-0.031	0.008	0.023	-0.000	0.032
	(0.007)	(0.036)	(0.027)	(0.005)	(0.009)
Observations	55	55	55	55	55
R-Squared	0.323	0.065	0.436	0.367	0.235
Mean	0.134	0.147	0.634	0.121	0.746
Rank	1/46	2/46	4/46	22/46	1/36

Standard errors in parentheses

Table 2.12: Changes in Composition in Finance

	Finance				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	-0.012	-0.007	0.080	0.036	-0.023
	(0.009)	(0.023)	(0.016)	(0.009)	(0.012)
Arizona_post	-0.014	-0.005	-0.003	0.015	-0.001
	(0.011)	(0.033)	(0.021)	(0.014)	(0.016)
Observations	55	55	55	55	55
R-Squared	0.348	0.066	0.644	0.533	0.316
Mean	0.047	0.114	0.765	0.323	0.630
Rank	2/46	16/46	17/46	8/46	21/46

Standard errors in parentheses

Table 2.13: Changes in Composition in Business

	Business				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.003	-0.002	0.047	0.036	-0.039
	(0.009)	(0.027)	(0.017)	(0.008)	(0.009)
Arizona_post	-0.027	0.012	0.009	0.001	0.025
	(0.012)	(0.035)	(0.022)	(0.012)	(0.012)
Observations	55	55	55	55	55
R-Squared	0.271	0.061	0.426	0.423	0.321
Mean	0.151	0.119	0.622	0.200	0.649
Rank	1/46	1/46	9/46	16/46	7/46

Standard errors in parentheses

Table 2.14: Changes in Composition in Leisure

	Leisure				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	-0.017	-0.012	0.103	0.007	0.010
	(0.013)	(0.026)	(0.012)	(0.007)	(0.010)
Arizona_post	-0.022	0.013	0.012	-0.007	0.028
	(0.016)	(0.034)	(0.018)	(0.009)	(0.014)
Observations	55	55	55	55	55
R-Squared	0.362	0.061	0.738	0.241	0.355
Mean	0.162	0.113	0.601	0.133	0.705
Rank	4/46	3/46	11/46	13/46	2/46

Standard errors in parentheses

Table 2.15: Changes in Composition in Professional

	Professional				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.006	0.009	0.035	0.001	-0.006
	(0.005)	(0.026)	(0.025)	(0.006)	(0.007)
Arizona_post	-0.006	0.004	0.015	-0.001	0.006
	(0.007)	(0.035)	(0.030)	(0.007)	(0.008)
Observations	55	55	55	55	55
R-Squared	0.131	0.029	0.295	0.395	0.456
Mean	0.057	0.123	0.702	0.434	0.509
Rank	2/46	7/46	7/46	24/46	10/46

Standard errors in parentheses

Table 2.16: Changes in Composition in Public

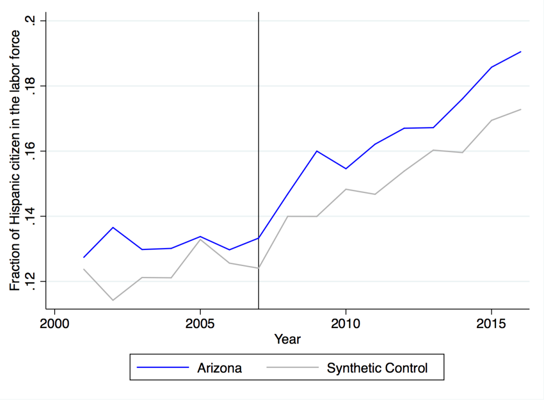
	Public				
	Hispanic noncitizen	Hispanic citizen	Nonhispanic white	High education	Low education
Arizona	0.012	0.049	-0.037	-0.017	0.005
	(0.004)	(0.033)	(0.036)	(0.011)	(0.013)
Arizona_post	-0.005	0.007	0.016	-0.005	0.010
	(0.005)	(0.045)	(0.047)	(0.012)	(0.014)
Observations	55	55	55	55	55
R-Squared	0.357	0.093	0.079	0.515	0.452
Mean	0.041	0.184	0.618	0.308	0.652
Rank	4/46	9/46	10/46	19/46	13/46

Standard errors in parentheses

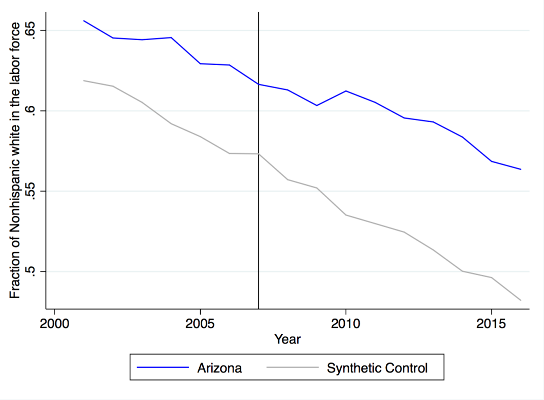
Appendices

Figure 2.A1: Trends in the Fraction in the Labor Force by Race

(a) Hispanic Citizen



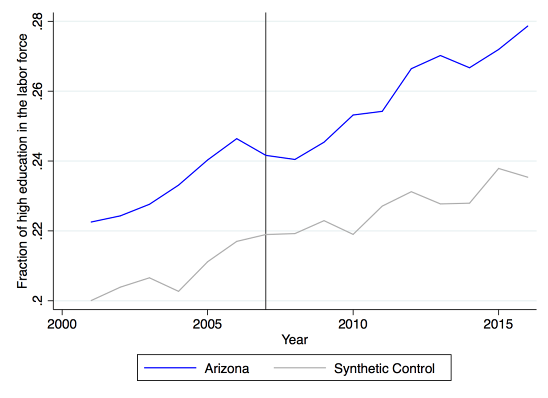
(b) Nonhispanic White



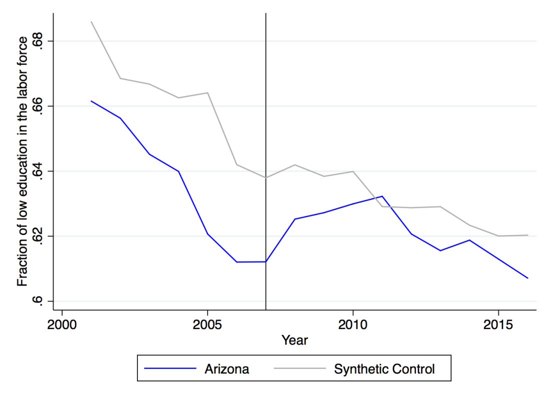
Note: The top panel shows the trends for the fraction of Hispanic citizen and the bottom panel shows the trends for the fraction of Nonhispanic white.

Figure 2.A2: Trends in the Fraction in the Labor Force by Education

(a) High-educated Workers



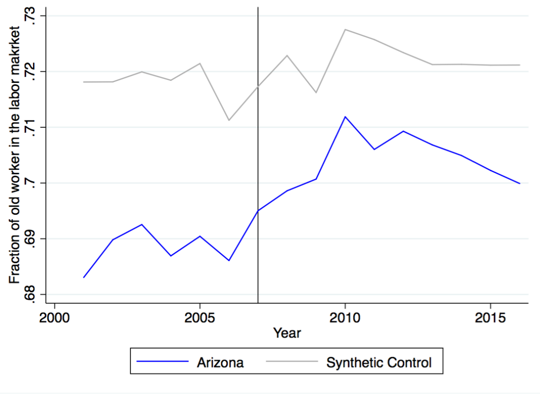
(b) Low-educated Workers



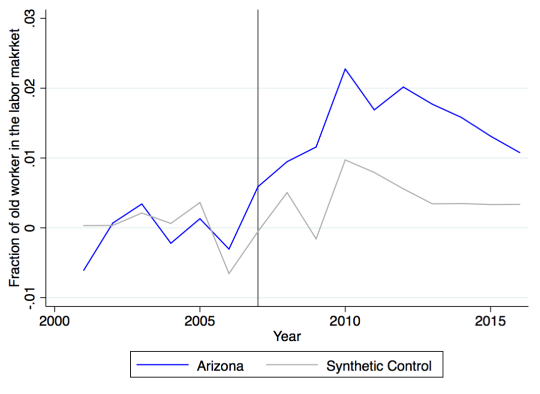
Note: The top panel shows the trends for the fraction of high-educated workers and the bottom panel shows the trends for the fraction of low-educated workers.

Figure 2.A3: The Fraction of Workers More Than 30 Years Old in the Labor Market

(a) absolute values



(b) demeaned values



Note: The top panel shows the absolute values and the right panel shows the corresponding demeaned values.

Figure 2.A4: The Percentage Decrease in the Fraction of Hispanic Noncitizen by Industry

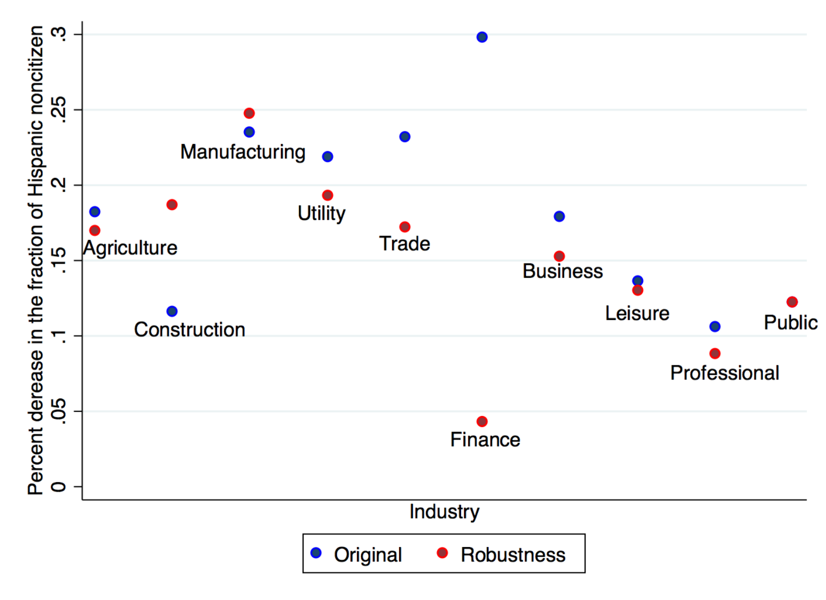


Table 2.A1: Means of Matching Variables (Continued)

Fraction of Hispanic noncitizen by industry	Arizona	Synthetic Control
Agriculture	0.565	0.365
	0.563	0.425
Construction	0.275	0.266
	0.340	0.295
Manufacturing	0.140	0.133
	0.157	0.168
Utility	0.059	0.072
	0.079	0.083
Trade	0.127	0.121
	0.129	0.121
Finance	0.041	0.056
	0.049	0.060
Business	0.134	0.137
	0.143	0.144
Leisure	0.135	0.164
	0.166	0.175
Professional	0.056	0.044
	0.056	0.053
Public	0.036	0.027
	0.037	0.025

Table 2.A2: Means of Matching Variables (Continued)

Labor force by industry	Arizona	Synthetic Control
Agriculture	60752	68203
	66802	70126
Construction	225186	231004
	299435	297358
Manufacturing	244943	219007
	239378	219273
Utility	177084	197945
	188633	208356
Trade	543229	610821
	600923	651940
Finance	186333	187326
	219146	216401
Business	181860	190937
	198502	199264
Leisure	156262	268866
	16646	284721
Professional	574899	643733
	653296	694711
Public	137287	132975
	147049	142501

Table 2.A3: Changes in Composition in the Labor Force by Industry (Robustness)

	Fraction of Hispanic noncitizen		Fraction in the population	
	Arizona_post*	Mean	Arizona_post*	Mean
Agriculture	-0.096(16.96%)	0.566	-0.001(4.00%)	0.025
Construction	-0.061(18.60%)	0.328	-0.011(10.78%)	0.102
Manufacturing	-0.039(24.68%)	0.158	0.001(1.16%)	0.086
Utility	-0.015(19.23%)	0.078	0.001(1.52%)	0.066
Trade	-0.023(17.16%)	0.134	0.000(0.00%)	0.212
Finance	-0.002(4.26%)	0.047	0.004(5.19%)	0.077
Business	-0.023(15.23%)	0.151	-0.001(1.41%)	0.071
Leisure	-0.021(12.96%)	0.162	-0.005(8.33%)	0.06
Professional	-0.005(8.77%)	0.057	0.011(4.70%)	0.234
Public	-0.005(12.20%)	0.041	0.002(3.85%)	0.052

*Arizona_post reports the difference-in-differences estimates based on the synthetic control.

Chapter 3

Labor Market Power and Wage Determination in China

3.1 Introduction

A growing literature has emphasized the existence of labor market monopsony power stemming from employer concentration within local labor markets, which deviates from the conventional view of labor markets as perfectly competitive. With a small number of employers bargaining with workers, the elasticity of the labor supply curve faced by a single employer may be finite and wages may be set below perfectly competitive rates. Such wage-setting power of employers has also been shown to contribute to stagnant wage growth and individual worker's welfare loss. It is hence crucial to understand to what extent do employers suppress wages as well as the role of imperfect competition among employers when making policy responses.

Card (2022) discusses the recent literature about the idea of monopsonistic wage setting, among which a growing number of papers study the relationship between labor market concentration measured by the Herfindahl-Hirschman Index (HHI) of firm employment on worker wages and show a negative effect of higher concentration on wages, with elasticities between the HHI and wages on the order of -0.05 to -0.15. In this paper, we use firm-level data from the Chinese Annual Survey of Industrial Firms (CASIF) to add to evidence of the relationship between employer concentration and wage behavior, and to study labor market power and wage determination in China over the period 1998 to 2013. Our data includes all state-owned and non-state owned industrial firms with annual revenues above 5 million Yuan. We define a local labor market using different measures of regions and industries, and construct the HHI measures of firm employment at both the prefecture-industry-year level and the county-industry-year level. The underline assumption is that there is limited mobility of jobs across our definitions of regions and industries, due to the fact that the job search is largely local and workers switch to different jobs mostly within the same industry.

Existing studies that document a negative effect of local labor market concentration on wages exploit the cross-sectional and time-series variations of the HHI of firm employment (Benmelech et al., 2022; Rinz, 2022; Lipsius, 2018; Hershbein et al., 2018; Qiu and Sojourner, 2019), online job vacancies (Azar et al., 2022; Dube et al., 2020; Hershbein et al., 2018), or wage bills (Berger et al., 2019). Most of the papers explore either local labor markets or online labor markets in the United States, focusing on either the whole economy or certain industries such as manufacturing. The U.S. labor markets are usually measured at the county-industry-year level or the commuting zone-industry-year level, while some papers substitute industry measures with occupation or task measures. Our paper is closely related to Benmelech et al. (2022), which show an increasing trend of employment concentration in the manufacturing sector over the past decades that could be associated with greater import competition in the U.S. from China. It is hence interesting and meaningful to add more evidence from the flip side of the coin, i.e. concentration in local labor markets in China that experience export expansion at the same time.

We first verify that in contrast to the increasing trend found in the U.S. labor markets, the average HHI decreases in Chinese labor markets over the time period 1998 to 2013, which crossed China's entry into the WTO in 2001. Then we exploit OLS models with different sets of fixed effects and show a negative relationship between labor market concentration and wages both at the market level and at the firm level. The results indicate a 1% increase in the HHI is associated with an upto 0.177% decrease in worker wages at the market level and an upto 0.033% decrease in worker wages at the firm level. Further in order to deal with the endogeneity problem of the HHI, we construct the instrumental variable (IV) using the inverse number of employers in other geographical markets for the same industry and year, hence exploiting variations only driven by national-level changes instead of endogenous changes in productivity of a particular area. The estimated coefficients with IV regressions remain significantly negative, and become larger in magnitude, indicating that a 1% increase in the HHI reduces worker wages by upto 0.244% at the market level and upto 0.107% at the firm level.

To our knowledge, this paper is the first to analyze wage determination in China from the view of labor market power measured by the HHI. We find comparable and consistent results with existing studies and verify the existence of imperfect competition among firms in China. It also provides a potential explanation for the wage growth in China in recent decades and emphasizes the importance of considering the role of labor monopsony in policy evaluations. For example, the existence of labor market monopsony has also been indicated in the minimum wage literature that has found no evidence of dis-employment effects, following the early work of Card and Krueger (1994) (e.g., Dube et al., 2010; Giuliano, 2013). In contrast, some other studies have shown different levels of negative employment effects (e.g., Fang and Lin, 2015; Meer and

West, 2016). A recent paper by Azar et al. (2019a) empirically tests the employment effects of minimum wages interacted with the level of concentration of labor markets. They provide empirical support of the monopsony explanation of the lack of negative employment effects by showing that more concentrated labor markets experience less negative minimum wage-induced employment changes.

The rest of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the data and summary statistics. Section 4 provides the methodology used for analyses. Section 5 shows the empirical results. Finally, Section 6 Concludes.

3.2 Related Literature

We first complement the recent literature on market power of firms in the product market (e.g., Barkai, 2020; Autor et al., 2020; Covarrubias et al., 2019) that has found that the increase in industry concentration is associated with the decline in the labor share of GDP in the United States. Similar with product market power, a growing literature tend to depart from the textbook perfectly competitive labor markets and analyze causes and consequences of such labor market power (e.g., Manning, 2003, 2011; Hirsch et al., 2010), among which this paper fits into the empirical studies on the effects of labor market concentration on wages. For example, Azar et al. (2022) use data from an online job posting platform in the United States and measure labor market concentration with the HHI of job vacancies. They define local labor markets by a combination of commuting zones and occupations, and find that higher concentration is associated with lower posted wages. Benmelech et al. (2022) focus on U.S. manufacturing industries and find a negative relationship between local-level employer concentration measured by the HHI of firm employment and wages. With similar administrative data, Rinz (2022) suggests that increased local concentration reduces earnings and increases inequality while Lipsius (2018) shows that declining average local labor market concentration since 1980 is an implausible driver of the falling labor share in the United States. Qiu and Sojourner (2019) distinguish labor market concentration with product market concentration and suggest that higher product market concentration exacerbates the negative effects of labor market concentration on labor compensation. As for a more convincing research design in terms of the causality between concentration and wages, some studies exploit the merger-induced changes in concentration which also increase labor market power and analyze the wage effects. Prager and Schmitt (2021) examine hospital mergers and find evidence of reduced wage growth associated with the increase in concentration induced by large mergers. They further verify that the reduced wage growth is attenuated in markets with strong labor unions. Arnold (2019) finds similar results using matched employer-employee data form the U.S. Census and also implies negative spillovers on other firms in the same labor market. They also use data on job-to-job mobility patterns to account for substitutability across industries by extending a simple Cournot model of labor market competition.

This paper is also related with studies that use structural analysis to estimate the elasticity of the labor supply curve to an individual firm. Azar et al. (2019b) estimate the elasticity of job applications to posted wages as a proxy for the elasticity of the labor supply curve and relate their estimation to measures of labor market concentration. Dube et al. (2020) use an on-demand labor platform with both observational and experimental variations in wages. Their estimation of labor supply elasticity is around 0.1. Jarosch et al. (2019) develop a model of size-based labor market power to study the effects of labor market concentration on wages in Austria. They find that larger firms pay less due to workers' worse outside options and wages are lower in more concentrated markets since firms compete less for workers. Berger et al. (2019) develop a general equilibrium oligopsony model and exploit the quasi-experiment of state corporate tax changes in the United States as an identified labor demand shock. They then use their model to measure the welfare loss from labor market power as well as the effects of minimum wages and merger experiments.

Lastly, this study is also related to the local labor markets approach that takes the regional economy as the unit of analysis. Manning and Petrongolo (2017) propose a spatial job search model accounting for overlaps and interdependencies of labor markets and estimate that labor mobility across regions is limited and that the job search is largely local. Furthermore, in evaluating labor demand shocks, researchers exploit facts that regional adjustment across regions are slow and incomplete and that mobility of labor between sectors is costly. For example, by exploiting cross-market variations in import exposure stemming from initial differences in industry specialization, Autor et al. (2013) study the effects of import competition from China on U.S. local labor markets.

3.3 Data and Summary Statistics

3.3.1 Data and Main Variables

The main dataset for this analysis is the firm-level data from the Chinese Annual Survey of Industrial Firms (CASIF), which has been conducted by the National Bureau of Statistics of China (NBSC). It contains all the state-owned industrial firms and non-state-owned industrial firms that have annual revenues above 5 million Yuan. Our main analyses use all available years in the dataset, which are from 1998 to 2013. However, for some analyses, we focus on the period 1998-2007 due to issues of missing sample, missing key variables or poor data quality afterwards.[1] The number of firms varies from 145,966 in 1998 to 336,730 in 2007, and to 344,875 in 2013.

The key variables include the total annual wage and the total employment at the firm level, the region firm is located, and the industry classification. To decrease the influence of outliers in the data we winsorize the top 1% and bottom 1% of the total annual wage. The average wage of each firm is then calculated using the total wage divided by the total employment throughout the years, and is used as our main outcome variable. We use the average output per worker calculated by firm-level output divided by the total employment as the measure for worker productivity. Panel A of Table 3.2 reports descriptive statistics for these variables of firm observations over the period 1998-2013. The average total wage in a firm is 6,953.9 (in thousands) Yuan and the average total employment is approximately 300. The average annual wage per

[1] We also run our main analyses using data from 1998 to 2007 as robustness checks.

worker is 17.3 (in thousands) Yuan and the average productivity of a firm is 620.9 (in thousands) Yuan. Moreover, we group all firms by different ownership types according to the definition of the NBSC, and distinguish between local state-owned firms (i.e. firms in which the state holds a majority; hereafter, SOE), local nonstate-owned firms (i.e. collectively owned and private firms; hereafter, nonSOE), and firms with investors from Hong Kong, Macao or Taiwan as well as foreign countries (hereafter, foreign).

Our measures of regions are both prefectures and counties, where prefecture is the geographical division between province and county in China. However, due to changes in administrative boundaries in China, the consequent codes of regions may not be comparable during the time period of analysis. We hence convert the region codes of all years to the benchmark system based on 2002 National Standard of Administration (GB/T 2260-2002). According to China's Ministry of Civic Affairs, there are 2,859 counties within the 337 prefectures in the 31 provinces of mainland China in 2002. The industry classification uses China Standard Industrial Classification (CSIC) at the 4-digit level. We concert the industry code of all years based on the 2011 national standard (GB/T 4754-2011). We use measures of industries with both 4-digit CSIC code and 2-digit CSIC code. According to the definition of the NBSC, industrial firms that are included in the CASIF data refer to firms with the CSIC code starting from 06 to 46, 95% of which are manufacturing firms with the CSIC code starting from 13 to 43. Table 3.1 provides a list of the 41 distinct 2-digit CSIC codes as well as the percentage of all observations that belongs to each industry category. Firms are from 581 distinct 4-digit CISC code industries in the data.

3.3.2 Labor Market Concentration

The idea of constructing labor market concentration with the HHI is borrowed from the industrial organization literature that uses the HHI to measure product market concentration. The Federal Trade Commission (FTC) and the U.S. Department of Justice (DOJ) also use HHI thresholds to make horizontal merger guidelines, where the HHI above 2500 is regarded as "highly concentrated" and pre/post merger changes in the HHI of more than 200 in highly concentrated markets are likely to be challenged. While these guidelines are for evaluating whether a merger is likely to enhance market power on the selling side of the market, the antitrust agencies also state that the evaluation of whether a merger will enhance market power on the buying side of the market including the labor market is also part of the legal framework.

Due to the fact that labor mobility is limited across regions and industries and the job search is largely local, the emerging literature employs the local labor market approach that defines a labor market by a combination of regions, industries, and years. Hence, the HHI in region r, industry i, and year t is constructed as follows:

$$HHI_{rit} = \sum_{f=1}^{N} s_{frit}^2$$

where s_{frit} is the employment share of firm f in region r, industry i, and year t measured by the employment level of firm f divided by the total employment of all firms in the labor market and can be expressed as follows:

$$s_{frit} = \frac{emp_{frit}}{\sum_{f=1}^{N} emp_{frit}}$$

Hence similar with the HHI in product market that is calculated by summing squared market shares of all firms competing in a market and captures product market power, the HHI used in labor market exploits employment shares of all firms that hire from the same pool of workers in a market and captures labor market power. And the definition of a "labor market" is based on different measures of regions and industries. Panels B to E of Table 3.2 show summary statistics of HHI measures defined over different combinations of regions and industries over the time period 1998-2013. We report means of the HHI and create dummies indicating markets where there is only one firm (hence HHI=1). It is as expected that HHI measures are higher when we define local employer concentration at the county level than at the prefecture level and when we define local employer concentration with the 4-digit CSIC code than with the 2-digit CSIC code. The means of the HHI vary from 0.357 to 0.732, and the means of the dummies vary from 0.139 to 0.572, indicating that from 13.9% to 57.2% of all local labor markets has only one firm depending on the definition of the market. So local labor markets are concentrated and a few firms dominate hiring in the market. The standard deviations of HHI measures and dummies have the magnitude at 0.333 to 0.495, exhibiting substantial cross-sectional variations in local employer concentration.

The associated logs of total employment at the market level are also reported, with means varying from 5.454 to 7.466 and standard deviations varying from 1.412 to 1.865. Figure 3.1 shows trends of the average HHI in the local labor market from 1998 to 2013. These summary statistics are comparable with Benmelech et al. (2022), who use plant level data from the U.S. Census Bureau over the period 1978 to 2016 to construct HHI measures in the U.S. manufacturing sector.

3.4 Methodology

In theory, when the labor market becomes more concentrated, which means a smaller number of firms compete among workers within the local labor market, workers' bargaining position in wage determination will decrease and firms will have greater power to set wages. The negative relationship between labor market concentration and worker wages found in previous literature underscores the important role of labor market power in wage determination. In order to examine the relationship of labor market concentration and worker wages in China, we employ both the market level regressions similar with Azar et al. (2022) and firm level regressions similar with Benmelech et al. (2022). To deal with endogeneity of the HHI, we exploit the instrumental variable strategy similar with the one used in Azar et al. (2022) and Rinz (2022).

3.4.1 Market Level

We start our analyses with the following baseline regression at the market level:

$$lnW_{rit} = \alpha_t + \gamma_{ri} + \beta_1 HHI_{rit} + X'_{rit}\beta_2 + \mu_{rit}$$

where lnW_{rit} is the log average wage of all firms in region r, industry i, and year t. $lnHHI_{rit}$ is the corresponding log HHI and X'_{rit} is set of controls at the market by year level. We also include year fixed effects α_t and market fixed effects γ_{ri}. Standard errors μ_{rit} are clustered at the market level. The average wage for each local labor market is calculated using the average wage of all firms weighted by employment level of each firm within the local market, hence representing the average wage faced by each individual worker of the local market. We also include the log of total employment at the market-by-year level to consider the size of the local labor market.

Our specifications include different market fixed effects according to different definitions of local labor markets, and exploit variations in the HHI within local labor markets. We also include province-specific year trends to control for any time-varying province-specific characteristics that may drive the results. Furthermore, we also regress the market-level average wage on one-year lagged log HHI and one-year lagged log total employment to allow for the deferred effects of labor market power on wages.

3.4.2 Firm Level

We then conduct firm-level analyses using the following baseline regression:

$$lnW_{frit} = \alpha_f + \gamma_{rt} + \phi_{it} + \beta_1 HHI_{rit} + X'_{frit}\beta_2 + Z'_{rit}\beta_3 + \mu_{frit}$$

where lnW_{frit} is the log wage of firm f in region r, industry i, and year t. $lnHHI_{rit}$ is the corresponding log HHI, and X'_{frit} is a set of controls at the firm-by-year level and Z'_{rit} is a set of controls at the market-by-year level. Standard errors μ_{rit} are clustered at the market level.

Our different specifications include different sets of fixed effects. The first specification includes region fixed effects, industry fixed effects, year fixed effects, and firm fixed effects separately. The second specification considers time-varying region-specific and industry-specific characteristics by including region-by-year fixed effects γ_{rt} and industry-by-year fixed effects ϕ_{it}, together with firm fixed effects α_f. So the design exploits within-firm variation in the HHI over the years, and further controls for region trends and industry trends that may drive the results. The third specification further adds the log market-by-year level employment and the log firm-by-year level productivity as controls. Lastly, to deal with the concern of product market concentration that correlates with both local labor market concentration and worker wages (Autor et al., 2020), we include the log national-level HHI as a proxy for national-level product market competition, which is calculated by:

$$HHI_{it} = \sum_{f=1}^{N} s_{fit}^2$$

where s_{fit} is the employment share of firm f in industry i and year t measured by the employment level of firm f divided by the total employment of all firms in that industry of the same year and can be expressed as follows:

$$s_{fit} = \frac{emp_{fit}}{\sum_{f=1}^{N} emp_{fit}}$$

3.4.3 IV

To address the issue of the endogeneity problem of HHI measures, we adopt the instrumental variable strategy as in Azar et al. (2022), which is to instrument for the HHI of a local labor market with the average of log(1/N) in other geographical regions for the same industry and year (where N refers to the number of firms in the labor market). Hence, it identifies the effects of concentration on worker wages using variations in the HHI that are driven by national changes in the same industry and year, but are not related with time-varying market specific changes, especially the time-varying market-specific productivity changes that are correlated with the HHI and the average wage.

3.5 Results

3.5.1 Market Level

Tables 3.3-3.5 represent results for market-level regressions using data from 1998 to 2013. Columns 1-4 use prefectures as regional boundaries, and columns 5-8 use counties as regional boundaries. Columns 1-2 and 5-6 use 2-digit CSIC codes to define industries, and columns 3-4 and 7-8 use 4-digit CSIC codes to define industries. The odd columns report results with contemporaneous concentration measures and controls, and the even columns report results with lagged measures for concentration and control variables. All specifications include market fixed effects and year fixed effects. Table 3.3 reports the baseline results, where the coefficients on the log HHI are negative and significant across four different definitions of local labor markets, with the magnitude varying from 0.002 to 0.034, indicating a 1% increase in the HHI is associated with a 0.002% to 0.034% decrease in worker wages.

Table 3.4 includes the market-level employment as control and the magnitude of coefficients on the log HHI increases to 0.026 to 0.176. Table 3.5 includes province-specific year trends and the magnitude increases to 0.027 to 0.177, indicating the results are not driven by province-specific characteristics. Overall, the results are similar with Azar et al. (2022), which has the magnitude at about 0.03 to 0.1. Except when using prefectures and 2-digit CSIC codes to define a local market, the coefficients on one-year lagged log HHI are significantly negative with smaller magnitude, indicating there is less response of worker wages to lagged concentration measures. The coefficients on log market-level employment are negative and significant, contradicting the prediction of agglomeration effects, which suggests that wages should be higher in local markets with more workers. However, by construction, our outcome variable is negatively related with contemporaneous market employment but should not be related with lagged values. And our positive coefficients on lagged log market employment are hence consistent. The results verify that labor market monopsony measured by concentration has a detriment effect on worker wages.

3.5.2 Firm Level

Tables 3.6-3.9 represent results for four different specifications at the firm level using data from 1998 to 2013. Table 3.6 shows results with firm and year fixed effects, together with region and industry fixed effects. The coefficients on the log HHI remain negative, but become less significant. The magnitude reduces to about 0.001 to 0.008, indicating most of the effects at the market level are driven by cross-firm variations. The coefficients on one-year lagged log HHI are more significant and larger in magnitude, varying from 0.006 to 0.012. To further control for time-varying region and industry trends, Table 3.7 shows results with firm fixed effects, together with region-by-year and industry-by-year fixed effects, and the coefficients on the log HHI become positive and the coefficients on one-year lagged log HHI are similar. With the same set of fixed effects, Table 3.8 shows results when adding market-level and firm-level controls, which are log market employment and log firm productivity, and the coefficients are significantly negative with magnitude varying from 0.005 to 0.033. As expected, the coefficients on log productivity are positive, indicating a positive relation between labor productivity and wages. Lastly, by further controlling for national HHI, the results shown in Table 3.9 remain similar, indicating estimated effects are not driven by product market concentration.

3.5.3 IV

Table 3.10 represents results based on the instrument variable strategy at the market level, using the same specification with Table 3.5. Panel A shows results for the first stage and indicates that the log HHI in the local labor market is positively correlated with the average of log (1/N) in other markets. So the relevance condition is satisfied. The coefficients on the log HHI shown in Panel B remain significantly negative, but become larger in magnitude. A 1% increase in the HHI reduces worker wages by 0.109% to 0.244%.

Table 3.11 represents IV results at the firm level, using the same specification with Table 3.8. The coefficients on the log HHI shown in Panel B are significantly negative and the magnitude increases to 0.034 to 0.107.

3.5.4 Ownership

According to the definition of the NBSC, around 13% of all firms in the data are state-owned firms (SOE), 73% of all firms are nonstate-owned firms (nonSOE), and 14% of all firms are owned by investors from Hong Kong, Macao or Taiwan as well as foreign countries (foreign). Figure 3.2 shows trends of the average HHI of firm employment measured with prefectures and two-digit CSIC codes for the three types of firms by their ownership structure. There are sharp changes in the trends in 2010 due to issues of missing key variables (the ownership variable is missing in 2009; the employment variable is missing in 2011) and inconsistency of data coverage after 2008. The average HHI of nonSOE and foreign firms decreases over time, with the former experiencing slightly more changes. However, the average HHI of SOE increases during the time period, which could be explained by mergers and acquisitions of SOE since around 2000. In our firm-level analyses based on ownership, we focus on the period 1998 to 2007 and use the same specification with Table 3.8. Panels A and B of Table 3.12 show results using prefectures and counties to define regions, respectively. We show separately the results for SOE, nonSOE and foreign firms. SOE should response less to concentration compared with the nonSOE. However, the coefficients associated with SOE are the largest in absolute values, which could be explained by mergers and acquisitions that affect both the HHI and the average firm wage.

3.6 Conclusion

In conclusion, this paper adds to evidence on labor monopsony using firm-level data from China. We first show decreasing trends of local labor market concentration measured by the HHI over the years from 1998 to 2013, which crossed China's entry into the WTO. We then find consistent results with the previous literature which finds that more concentrated labor markets are associated with lower wages. Our results indicate a 1% increase in the HHI measure of local labor market concentration significantly reduces worker wages by 0.177% at the market level and 0.033% at the firm level using OLS models, and by 0.244% at the market level and 0.107% at the firm level using IV models. The results are consistent using different definitions of local labor markets. It also sheds light on the importance of the role of imperfect competition in the labor market when making policy evaluations.

References

D. Arnold. Mergers and acquisitions, local labor market concentration, and worker outcomes. *Local Labor Market Concentration, and Worker Outcomes (October 27, 2019)*, 2019.

D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2): 645–709, 2020.

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

J. Azar, E. Huet-Vaughn, I. Marinescu, B. Taska, and T. Von Wachter. Minimum wage employment effects and labor market concentration. Technical report, National Bureau of Economic Research, 2019a.

J. Azar, I. Marinescu, and M. Steinbaum. Measuring labor market power two ways. In *AEA Papers and Proceedings*, volume 109, pages 317–21, 2019b.

J. Azar, I. Marinescu, and M. Steinbaum. Labor market concentration. *Journal of Human Resources*, 57(S):S167–S199, 2022.

S. Barkai. Declining labor and capital shares. *The Journal of Finance*, 75(5):2421–2463, 2020.

E. Benmelech, N. K. Bergman, and H. Kim. Strong employers and weak employees how does employer concentration affect wages? *Journal of Human Resources*, 57(S): S200–S250, 2022.

D. W. Berger, K. F. Herkenhoff, and S. Mongey. Labor market power. Technical report, National Bureau of Economic Research, 2019.

D. Card. Who set your wage? Technical report, National Bureau of Economic Research, 2022.

D. Card and A. B. Krueger. Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. *American Economic Review*, 84 (4):772–93, 1994.

M. Covarrubias, G. Gutiérrez, and T. Philippon. From good to bad concentration? us industries over the past 30 years. Technical report, National Bureau of Economic Research, 2019.

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

A. Dube, J. Jacobs, S. Naidu, and S. Suri. Monopsony in online labor markets. *American Economic Review: Insights*, 2(1):33–46, 2020.

T. Fang and C. Lin. Minimum wages and employment in china. *IZA Journal of Labor Policy*, 4(1):22, 2015.

L. Giuliano. Minimum wage effects on employment, substitution, and the teenage labor supply: Evidence from personnel data. *Journal of Labor Economics*, 31(1):155–194, 2013.

B. Hershbein, C. Macaluso, and C. Yeh. Concentration in us local labor markets: evidence from vacancy and employment data. Technical report, Working paper, 2018.

B. Hirsch, T. Schank, and C. Schnabel. Differences in labor supply to monopsonistic firms and the gender pay gap: An empirical analysis using linked employer-employee data from germany. *Journal of Labor Economics*, 28(2):291–330, 2010.

G. Jarosch, J. S. Nimczik, and I. Sorkin. Granular search, market structure, and wages. 2019.

B. Lipsius. Labor market concentration does not explain the falling labor share. *Available at SSRN 3279007*, 2018.

A. Manning. *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press, 2003.

A. Manning. Imperfect competition in the labor market. In *Handbook of labor economics*, volume 4, pages 973–1041. Elsevier, 2011.

A. Manning and B. Petrongolo. How local are labor markets? evidence from a spatial job search model. *American Economic Review*, 107(10):2877–2907, 2017.

J. Meer and J. West. Effects of the minimum wage on employment dynamics. *Journal of Human Resources*, 51(2):500–522, 2016.

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

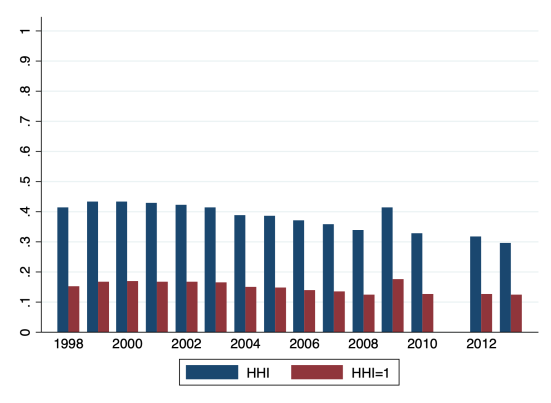
Y. Qiu and A. Sojourner. Labor-market concentration and labor compensation. *Available at SSRN 3312197*, 2019.

K. Rinz. Labor market concentration, earnings, and inequality. *Journal of Human Resources*, 57(S):S251–S283, 2022.

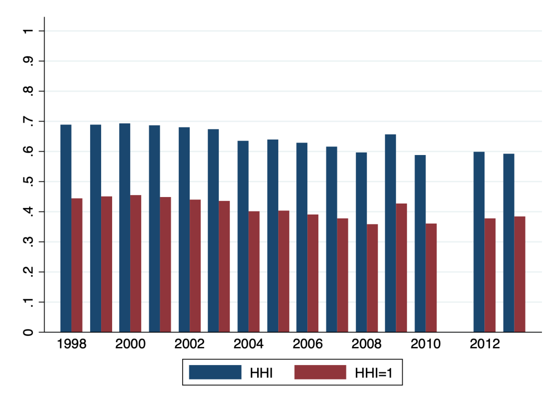
Figures and Tables

Figure 3.1: Trends in Average Local Labor Market Concentration

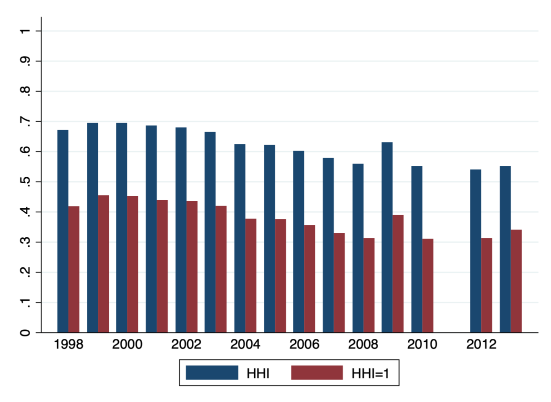
(a) Prefecture and two-digit CSIS code



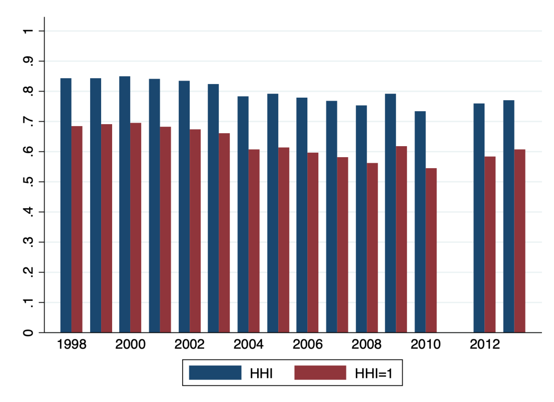
(b) Prefecture and four-digit CSIS code



(c) County and two-digit CSIS code

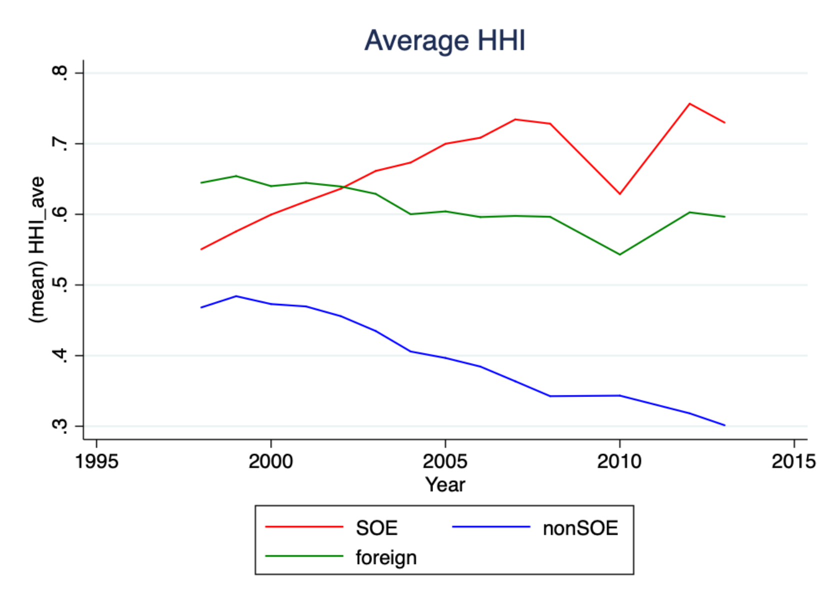


(d) County and four-digit CSIS code



Note: The figure plots trends of labor market concentration that is measured by the HHI of firm employment for each combination of different measures of districts and industries. The index is the average across district-industry-year triads. It shows the degree of labor market concentration that an average employer is in.

Figure 3.2: Trends in Average Local Labor Market Concentration by Ownership



Note: The figure plots trends of the average HHI of firm employment that is measured with prefecture and two-digit CSIC code for three types of firms by their ownership structure. SOE refers to state-owned firms, nonSOE refers to nonstate-owned firms, and foreign refers to firms that are owned by investors from Hong Kong, Macao or Taiwan as well as foreign countries.

Table 3.1: List of 2-digit CSIC Codes

2-digit CSIC Codes	Description	Percentage
Mining		
06	Coal Mining and Dressing	1.99
07	Petroleum and Natural Gas Extraction	0.03
08	Ferrous Metals Mining and Dressing	0.75
09	Nonferrous Metals Mining and Dressing	0.56
10	Nonmetal Minerals Mining and Dressing	0.92
11	Mining auxiliary activity	0.04
12	Other Mining and Dressing	0.00
Manufacturing		
13	Food Processing	5.78
14	Food Production	2.17
15	Beverage Production	1.49
16	Tobacco Processing	0.08
17	Textile Industry	7.71
18	Garments and Other Fiber Products	4.14
19	Leather, Furs, Down and Related Products	2.69
20	Timber Processing, Bamboo, Cane, Palm Fiber and Straw Products	2.12
21	Furniture Manufacturing	2.17
22	Paper making and Paper Products	2.50
23	Printing and Record Medium Reproduction	1.70
24	Cultural, Educational and Sports Goods	2.51
25	Petroleum Processing and Coking	0.66
26	Raw Chemical Materials and Chemical Products	6.75
27	Medical and Pharmaceutical Products	1.86
28	Chemical Fiber Products	0.50
29	Rubber and Plastic Products	5.06
30	Nonmetal Mineral Products	7.81
31	Smelting and Pressing of Ferrous Metals	3.30
32	Smelting and Pressing of Nonferrous Metals	1.68
33	Metal Products	5.60
34	Ordinary Machinery Manufacturing	6.09
35	For Special Purposes Equipment Manufacturing	3.98
36	Automobile Manufacturing	2.89
37	Railway, Watercraft, Aerospace and Other Transport Equipment	1.50
38	Electric Equipment and Machinery	5.80
39	Electronic and Telecommunications Equipment	3.28
40	Instruments and Meters Machinery	1.17
41	Other Manufacturing	0.49
42	Comprehensive Utilization of Waste Resources	0.19
43	Repair of Metal Products, Machinery and Equipment	0.10
Production and Supply		
44	Production and Supply of Electric Power, Steam and Hot Water	1.95
45	Production and Supply of Gas	0.21
46	Production and Supply of Tap Water	0.77

Table 3.2: Summary Statistics on Firm Observations

	Mean	STD
Total firm wage (000)	6,953.9	74,420.97
Total employees	299.5	1,350.08
Avergage wage (winsorized, 000)	17.3	18.78
Productivity (output per employee, 000)	620.9	23,981.46
HHI (prefecture ind2 year)	0.357	0.333
HHI (prefecture ind2 year)=1	0.139	0.346
log(employment, prefecture ind2 year)	7.466	1.865
HHI (prefecture ind4 year)	0.594	0.370
HHI (prefecture ind4 year)=1	0.377	0.485
log(employment, prefecture ind4 year)	6.028	1.619
HHI (county ind2 year)	0.576	0.368
HHI (county ind2 year)=1	0.351	0.477
log(employment, county ind2 year)	6.144	1.608
HHI (county ind4 year)	0.732	0.352
HHI (county ind4 year)=1	0.572	0.495
log(employment, county ind4 year)	5.456	1.412

Table 3.3: Local Employer Concentration and Wages, Market Level I

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	-0.002		-0.021***		-0.025***		-0.034***	
	(0.004)		(0.002)		(0.003)		(0.002)	
log(lag HHI)		0.012***		-0.005**		-0.011***		-0.013***
		(0.004)		(0.002)		(0.003)		(0.002)
Region-industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Province-specific year trends								
Observations	118,549	97,825	537,068	413,274	373,617	296,089	857,283	612,352
R-Squared	0.780	0.772	0.692	0.693	0.706	0.701	0.686	0.689
Mean	2.411	2.411	2.393	2.393	2.327	2.327	2.376	2.376

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Local Employer Concentration and Wages, Market Level II

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	-0.026***		-0.101***		-0.095***		-0.176***	
	(0.005)		(0.003)		(0.003)		(0.003)	
log(employment)	-0.037***		-0.084***		-0.079***		-0.135***	
	(0.004)		(0.002)		(0.002)		(0.002)	
log(lag HHI)		0.017***		-0.000		-0.007***		-0.020***
		(0.004)		(0.003)		(0.003)		(0.003)
log(lag tot. employment)		0.008***		0.005***		0.005***		-0.007**
		(0.003)		(0.002)		(0.002)		(0.002)
Region-industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Province-specific year trends								
Observations	118,549	97,825	537,068	413,274	373,617	296,089	857,283	612,352
R-Squared	0.781	0.772	0.697	0.693	0.710	0.701	0.696	0.690
Mean	2.411	2.411	2.393	2.393	2.327	2.327	2.376	2.376

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Local Employer Concentration and Wages, Market Level III

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	-0.027***		-0.100***		-0.100***		-0.177***	
	(0.005)		(0.003)		(0.003)		(0.003)	
log(employment)	-0.035***		-0.081***		-0.077***		-0.132***	
	(0.004)		(0.002)		(0.002)		(0.002)	
log(lag HHI)		0.017***		0.000		-0.010***		-0.022***
		(0.004)		(0.003)		(0.003)		(0.003)
log(lag tot. employment)		0.011***		0.009***		0.007***		-0.003**
		(0.003)		(0.002)		(0.002)		(0.002)
Region-industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Province-specific year trends	Y	Y	Y	Y	Y	Y	Y	Y
Observations	118,549	97,825	537,068	413,274	373,617	296,089	857,283	612,352
R-Squared	0.786	0.777	0.701	0.698	0.714	0.706	0.700	0.694
Mean	2.411	2.411	2.393	2.393	2.327	2.327	2.376	2.376

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Local Employer Concentration and Wages, Firm Level I

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	-0.008**		-0.001		-0.005		-0.002	
	(0.004)		(0.002)		(0.003)		(0.002)	
log(lag HHI)		-0.006*		-0.008***		-0.012***		-0.010***
		(0.004)		(0.002)		(0.003)		(0.002)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Region-year FE								
Industry-year FE								
Observations	2,997,396	2,440,345	2,997,396	2,383,891	2,207,820	1,742,449	2,207,820	1,635,057
R-Squared	0.691	0.679	0.691	0.682	0.689	0.676	0.689	0.684
Mean	2.510	2.510	2.510	2.510	2.478	2.478	2.478	2.478

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Local Employer Concentration and Wages, Firm Level II

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	0.001		0.002		0.007***		0.006***	
	(0.002)		(0.001)		(0.002)		(0.001)	
log(lag HHI)		-0.006***		-0.009***		-0.006***		-0.007***
		(0.002)		(0.001)		(0.002)		(0.001)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE								
Region FE								
Industry FE								
Region-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,997,390	2,440,215	2,997,376	2,383,671	2,207,462	1,741,313	2,207,424	1,633,763
R-Squared	0.706	0.697	0.710	0.704	0.723	0.715	0.726	0.726
Mean	2.510	2.510	2.510	2.510	2.478	2.478	2.478	2.478

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Local Employer Concentration and Wages, Firm Level III

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	-0.000		-0.010***		-0.005***		-0.033***	
	(0.002)		(0.001)		(0.002)		(0.002)	
log(tot. employment)	-0.013***		-0.024***		-0.026***		-0.049***	
	(0.002)		(0.001)		(0.001)		(0.001)	
log(productivity)	0.347***	0.300***	0.344***	0.296***	0.357***	0.306***	0.351***	0.304***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
log(lag HHI)		-0.002		0.001		0.003		0.003*
		(0.002)		(0.001)		(0.002)		(0.002)
log(lag tot. employment)		-0.000		0.002**		-0.000		0.000
		(0.001)		(0.001)		(0.001)		(0.001)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE								
Region FE								
Industry FE								
Region-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,982,044	2,429,283	2,982,031	2,373,005	2,198,560	1,735,779	2,198,522	1,628,625
R-Squared	0.765	0.742	0.767	0.747	0.781	0.760	0.784	0.768
Mean	2.510	2.510	2.510	2.510	2.478	2.478	2.478	2.478

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: Local Employer Concentration and Wages, Firm Level IV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Region:	Prefecture				County			
Industry:	2-digit CSIC		4-digit CSIC		2-digit CSIC		4-digit CSIC	
Dep. Var.:	log(ave. wage)							
log(HHI)	-0.000		-0.010***		-0.005***		-0.033***	
	(0.002)		(0.001)		(0.002)		(0.002)	
log(national HHI)	-0.164		-0.689		0.597		-0.078	
	(266.106)		(313.226)		(158.951)		(192.273)	
log(tot. employment)	-0.013***		-0.024***		-0.026***		-0.048***	
	(0.002)		(0.001)		(0.001)		(0.001)	
log(productivity)	0.346***	0.300***	0.344***	0.296***	0.357***	0.306***	0.351***	0.304***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
log(lag HHI)		-0.002		0.001		0.003		0.003*
		(0.002)		(0.001)		(0.002)		(0.002)
log(lag national HHI)		0.018		-0.015**		0.004		0.006
		(0.023)		(0.007)		(0.012)		(0.005)
log(lag tot. employment)		0.002***		0.002*		-0.000		0.000
		(0.001)		(0.001)		(0.001)		(0.001)
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE								
Region FE								
Industry FE								
Region-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,982,044	2,429,283	2,982,031	2,373,005	2,198,560	1,735,779	2,198,522	1,628,625
R-Squared	0.765	0.742	0.767	0.747	0.781	0.760	0.784	0.768
Mean	2.510	2.510	2.510	2.510	2.478	2.478	2.478	2.478

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: Local Employer Concentration and Wages (IV), Market Level

Panel A: First Stage

	(1)	(2)	(3)	(4)
Region:	Prefecture		County	
Industry:	2-digit CSIC	4-digit CSIC	2-digit CSIC	4-digit CSIC
Dep. Var.:	log(HHI)			
Average log(1/N)	0.634***	0.576**	0.727***	0.559***
	(0.016)	(0.008)	(0.016)	(0.010)
log(employment)	-0.243***	-0.238***	-0.261***	-0.217***
	(0.004)	(0.001)	(0.002)	(0.001)
Region-industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Province-specific year trends	Y	Y	Y	Y
Observations	136,282	619,071	429,548	994,700
R-Squared	0.893	0.838	0.837	0.794
Mean	7.736	8.520	8.474	8.850

Panel B: Second Stage

	(1)	(2)	(3)	(4)
Region:	Prefecture		County	
Industry:	2-digit CSIC	4-digit CSIC	2-digit CSIC	4-digit CSIC
Dep. Var.:	log(ave. wage)			
log(HHI)	-0.109***	-0.112***	-0.157***	-0.244***
	(0.018)	(0.013)	(0.023)	(0.020)
log(employment)	-0.058***	-0.085***	-0.093***	-0.148***
	(0.006)	(0.004)	(0.007)	(0.005)
Region-industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Province-specific year trends	Y	Y	Y	Y
Observations	118,549	537,068	373,617	857,283
R-Squared	0.784	0.701	0.713	0.699
Mean	2.411	2.393	2.327	2.376
Kleibergen-Paap F-stat	1423.070	5069.540	1878.054	2865.099

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11: Local Employer Concentration and Wages (IV), Firm Level

Panel A: First Stage

	(1)	(2)	(3)	(4)
Region:	Prefecture		County	
Industry:	2-digit CSIC	4-digit CSIC	2-digit CSIC	4-digit CSIC
Dep. Var.:	log(HHI)			
Average log(1/N)	-175.485***	-23.657***	-250.615***	-27.745***
	(7.838)	(0.871)	(20.081)	(0.849)
log(employment)	0.214***	-0.196***	-0.099***	-0.271***
	(0.021)	(0.006)	(0.013)	(0.003)
log(productivity)	0.005***	-0.011***	-0.002**	-0.033***
	(0.001)	(0.001)	(0.001)	(0.001)
Firm FE	Y	Y	Y	Y
Year FE				
Region FE				
Industry FE				
Region-year FE	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y
Observations	3,508,068	3,508,050	2,595,635	2,595,596
R-Squared	0.961	0.949	0.955	0.949
Mean	5.873	7.052	7.055	7.865

Panel B: Second Stage

	(1)	(2)	(3)	(4)
Region:	Prefecture		County	
Industry:	2-digit CSIC	4-digit CSIC	2-digit CSIC	4-digit CSIC
Dep. Var.:	log(ave. wage)			
log(HHI)	-0.034***	-0.061***	-0.061***	-0.107***
	(0.003)	(0.004)	(0.004)	(0.005)
log(employment)	-0.020***	-0.041***	-0.041***	-0.074***
	(0.002)	(0.002)	(0.002)	(0.002)
log(productivity)	0.347***	0.343***	0.357***	0.348***
	(0.002)	(0.001)	(0.002)	(0.001)
Firm FE	Y	Y	Y	Y
Year FE				
Region FE				
Industry FE				
Region-year FE	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y
Observations	2,982,044	2,982,027	2,198,560	2,198,522
R-Squared	0.764	0.766	0.781	0.783
Mean	2.510	2.510	2.478	2.478
Kleibergen-Paap F-stat	415.084	733.548	163.706	2865.099

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12: Local Employer Concentration and Wages by Ownership, Firm Level

Panel A:

	(1)	(2)	(3)	(4)	(5)	(6)
Region:			Prefecture			
Industry:		2-digit CSIC			4-digit CSIC	
Ownership:	SOE	non SOE	Foreign	SOE	non SOE	Foreign
Dep. Var.:			log(ave. wage)			
log(HHI)	-0.009**	-0.004*	0.001	-0.035***	-0.008***	-0.012***
	(0.004)	(0.002)	(0.004)	(0.005)	(0.002)	(0.003)
log(tot. employment)	-0.035***	0.000	-0.003	-0.064***	-0.011***	-0.013***
	(0.004)	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)
log(productivity)	0.229***	0.239***	0.254***	0.226***	0.238***	0.255***
	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE						
Region FE						
Industry FE						
Region-year FE	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y
Observations	244,382	1,401,543	269,299	244,102	1,401,497	268,999
R-Squared	0.821	0.737	0.741	0.827	0. 739	0.747
Mean	1.990	2.297	2.688	1.990	2.297	2.688

Panel B:

	(1)	(2)	(3)	(4)	(5)	(6)
Region:			Prefecture			
Industry:		2-digit CSIC			4-digit CSIC	
Ownership:	SOE	non SOE	Foreign	SOE	non SOE	Foreign
Dep. Var.:			log(ave. wage)			
log(HHI)	-0.033***	-0.007***	0.008*	-0.097***	-0.018***	-0.011**
	(0.006)	(0.002)	(0.004)	(0.008)	(0.002)	(0.004)
log(tot. employment)	-0.071***	-0.009***	-0.001	-0.131***	-0.021***	-0.019***
	(0.004)	(0.002)	(0.004)	(0.005)	(0.001)	(0.003)
log(productivity)	0.236***	0.240***	0.254***	0.224***	0.238***	0.253***
	(0.004)	(0.002)	(0.005)	(0.004)	(0.002)	(0.004)
Firm FE	Y	Y	Y	Y	Y	Y
Year FE						
Region FE						
Industry FE						
Region-year FE	Y	Y	Y	Y	Y	Y
Industry-year FE	Y	Y	Y	Y	Y	Y
Observations	179,384	1,032,637	156,198	178,977	1,032,564	155,724
R-Squared	0.840	0.758	0.742	0.849	0.760	0.752
Mean	1.943	2.264	2.607	1.943	2.264	2.607

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$