Essays on Labor Mobility

by

Parag Mahajan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in The University of Michigan
2020

Doctoral Committee:

Professor John Bound, Co-Chair
Professor Dean Yang, Co-Chair
Professor Hoyt Bleakley
Professor Jagadeesh Sivadasan
Professor Jeffrey A. Smith

Parag Mahajan

paragma@umich.edu

ORCID ID 0000-0002-2020-1964

For Roop and Kavita

ACKNOWLEDGEMENTS

First and foremost, this dissertation would not be possible without the advising from my dissertation committee: John Bound, Dean Yang, Jeff Smith, Hoyt Bleakley, and Jagadeesh Sivadasan. Each committee member has helped me grow as a researcher, professional academic, and adult during my time at Michigan. My committee let me do my thing and let me do it my way. Yet, they always showed up when I needed them to steer me in the right direction.

John has been singularly instrumental in helping me hone and improve my research, particularly Chapter 1. I knew I had something once John was on board. There were times when he was the only person who really understood what I was trying to do, and there were times when he was the only one who could save me from what I was trying to do. I will miss our meetings dearly and thank him for his invaluable advising. Dean has been a constant source of help and advice throughout graduate school. He managed to make me feel like a peer, even as he taught me how to cut through the noise and write a good paper. I thank him for his trust and for everything I have learned from him over the last six years. Jeff's confidence in me has given me confidence in myself since my second year in graduate school. He has been selfless with his time and engagement in advising my work. I cannot thank him enough for agreeing to stay on my committee as an Ann Arbor ex-pat and hosting me in Madison to boot. I have connected with Hoyt as a student, mentee, co-author, and friend over the last six years. I thank him for always pushing me to think about the most interesting question rather than the question that is easiest to answer, and for his sense of humor in the moments when we have exhausted our brains trying to answer it. Jagadeesh was magnanimous in helping me with data access and generous with his time as an advisor, guiding me through concepts in trade despite my lack of formal training. I thank him for all of his help.

I also have to thank Ron Caldwell and Jeff Wenger. Ron made teaching easy, fun, and one of the best parts of my experience as a graduate student. No matter what he says, as far as Dream Teams go, I'd rather be Christian Laettner than Isiah Thomas. I could not have asked for more from a summer internship than what I got at RAND, and that was all thanks to Jeff. I cannot thank him enough for our work together his mentorship since.

Building friendships and professional relationships with my peers has been the most rewarding part of graduate school. I want to thank Chapter 2's co-author, Dhiren Patki, for always keeping my eyes on the prize (as opposed to robustness check #23). I hope this is the start of many joint ventures. I would not have passed prelims, let alone survived the job market, without Dhiren, Ari Binder and Max Risch providing levity and expertise at every step. Jessica Gillooly has been a sounding board, occasional writing tutor, and unwavering ally for seven years. I have also benefited greatly from my friendships with Evan Taylor, (unofficial committee member) Bryan Stuart, Max Gross, Anirudh Jayanti, Meera Mahadevan, Ben Lipsius, Jamie Fogel, Shuqiao Sun, and countless others.

In addition, Michigan's research community—particularly the Labor Economics group—played a large role in my development. Thanks to Charlie Brown, Mel Stephens, and Gabor Kezdi for their instruction, and to all the faculty who were regulars at the Labor Lunch. They treated us all like advisees and always provided useful feedback, regardless of whether our projects were "ready for prime time." I also want to thank colleagues at various seminars throughout Michigan and across the country for their comments and suggestions on my work. Mim Jones, Kathryn Cardenas, Julie Heintz, and Laura Flak all helped build support systems that made sure I never got off track. Finally, I want to thank Joelle Abramowitz and Clint Carter for putting up with my demanding use of the Census RDC. I would be remiss (and a criminal) if I did not say, "any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed."

Completing a dissertation requires inspiration and music. Regarding the former, I want to thank Lebron James. Regarding the latter, I want to thank the creators of the "All-Nighter," "Lo-Fi Beats," and "Brain Food" playlists on Spotify, Above and Beyond for their weekly "Group Therapy" radio show, and Future.

This dissertation was a labor of love to my family. Thank you to my sisters for their support throughout this process and for inspiring me with their unwavering pursuit of improving others' lives. Nothing would be possible without my parents, who traveled half way across the world so that the three of us could pursue our dreams. Living up to this lofty standard has been a blessing, not a burden. Their pride reassures me that the sweat poured into this document was worth it.

Finally, and most importantly, I thank my wife Jarcy. She is the smartest, kindest, and steadiest person I know, and I needed every drop of these qualities over the last seven years. There is no dissertation without her, and I hope I can return some small portion of the love she has provided me with as we move on to the next phase of our partnership.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

This dissertation investigates the determinants and consequences of labor mobility across geographic areas and firms. It is motivated 1) by the singular potential for such mobility to increase welfare; 2) by troubling macroeconomic declines in labor mobility and business dynamism in the U.S.; and 3) by recent literature that allows us to better-understand how differences across firms mediate labor market outcomes.

Its first chapter argues that firm entry and exit play critical roles in determining how immigrant workers are absorbed into and affect local economies. It first documents a positive relationship between immigrant workers and business presence, with inflows driving small-to-medium sized firm creation and reducing exit by older, large firms. It finds that these responses play a dominant role in immigrant worker absorption, accounting for more than two-thirds of immigrant-induced job creation. Using observed proxies for productivity, it then uncovers a critical heterogeneity: while firms are less likely to exit on average, low-productivity firms are more likely to shut down in response to immigrant inflows. The resultant increases in creative destruction are driven by immigrant workers, as opposed to immigrant consumer demand. Placed in the context of a theoretical framework that accounts for firm heterogeneity, these results imply that firm entry and exit drive local production responses to immigration, leading to substantially larger estimates of immigrant-generated economic surplus than canonical models of labor demand.

The second chapter is co-authored with Dhiren Patki. Using linked employer-employee administrative data from Germany, we find that cohorts entering the labor market during a recession experience a 4.9 percent reduction in wages cumulated over the first decade of labor market experience. While 40 percent of the wage loss is due to reductions in employer-specific pay, we use a revealed preference-based algorithm to show that fully three-fourths of the losses in employer-specific pay are compensated for by non-pay amenities. The higher non-pay amenities that we associate with recessionary labor market entrants are consistent with the view that employers that hire during business cycle downturns exhibit less cyclically sensitive labor demand and provide greater long-term job security. Our findings show that the welfare cost of labor market entry during recessions can be less severe than pecuniary estimates would imply.

The final chapter is co-authored with Dean Yang. It studies the interplay between negative shocks in origin countries and migrant networks in destination countries. Specifically, we examine the impact of hurricanes on a quarter-century of international migration to the United States and find that hurricanes increase migration to the U.S., with the effect dominated by working-aged individuals and magnitude increasing in the size of prior migrant stocks. We provide new insights into how networks facilitate legal, permanent U.S. immigration in response to origin country shocks, a matter of growing importance as climate change increases natural disaster impacts.

CHAPTER I

Immigration and Local Business Dynamics: Evidence from U.S. Firms

1.1 Introduction

The Census Bureau estimates that by 2030, immigration will overtake natural increase (births minus deaths) as the primary driver of population growth in the United States.[1] A workforce that will increasingly rely on foreign-born workers requires a comprehensive understanding of how they are absorbed into and ultimately shape local economies.[2] Recent advances to data and theory have dramatically expanded our insight into how this process occurs on the firm side of the labor market, with a particular focus on the form and choice of production technique.[3] Nonetheless, most of this burgeoning literature has either implicitly or explicitly restricted its attention to representative firm models of production that, by definition, do not feature differences across firms in input use or total factor productivity. Furthermore, the empirical work that has studied individual firm responses to immigration has largely centered on changes to existing firms in non-U.S. settings.[4]

In contrast, broader study of the U.S. economy—the largest immigrant worker destination in the world—finds that firm entry and exit dynamics are crucial drivers of job creation and productivity growth, particularly when firm entry is accompanied by the exit of less productive firms.[5] Furthermore, immigrant workers appear to be especially active in changing firm entry and shut down decisions: immigrants have a higher propensity to start businesses

[1]See Vespa and Armstrong (2018).

[2]This chapter will use the term workers to encompass both the self-employed and employees.

[3]See, e.g., Lewis (2005, 2012); Clemens et al. (2017); Dustmann and Glitz (2015); Peri (2012); Mitaritonna et al. (2017); Lewis (2011); D'Amuri et al. (2010); Peri and Sparber (2009); Gonzalez and Ortega (2011).

[4]See, e.g., Mitaritonna et al. (2017); Malchow-Moller et al. (2012).

[5]See, e.g., Foster et al., 2008; Bartelsman and Doms, 2000; Baldwin, 1995.

than natives[6], are relatively more likely to work for new businesses[7], and have been found to prevent establishment exit in non-U.S. contexts.[8] When viewed through this lens, the question of how local economies generate enough jobs to absorb immigrant inflows naturally leads us to consider firm entry and firm exit. When we accordingly open the door to differences in productivity and factor use across firms, such entry and exit introduces a new channel through which immigrants alter a local economy: by changing its firm composition.

In this chapter, I thus argue that immigrant absorption into local labor markets and the subsequent effects immigrant workers have on local economies are predicated on firm entry and firm exit. I use several empirical analyses to illustrate this proposition. As a proof of concept, I first pool together two sources of plausibly exogenous variation using a method proposed in Dube and Zipperer (2015)—the Mariel Boatlift and the Legal Arizona Workers Act—and document a causal, positive relationship between immigrant inflows and increased business presence.

To more fully characterize the sources and consequences of this relationship, I next turn to a detailed commuting zone-industry-decade panel that utilizes administrative data from the Census Bureau's Longitudinal Business Database (LBD) and the universe of survey responses to Long-Form Decennial Censuses. To resolve endogeneity concerns, I develop a new instrument for geography-by-industry immigrant presence that exploits data on bilateral stocks of emigrants in non-U.S. OECD member nations to isolate exogenous migration pushes from sending countries. I then use previous compatriot locational choices to distribute pushed immigrants into specific commuting zones, and compatriot industry choices in other Census regions to distribute pushed immigrants into specific industries. I once again find a robust positive effect of immigrant inflows on firm presence within commuting zone-industry pairings over time. Each immigrant added roughly 0.05 additional firms to the commuting zone and industry in which they worked in the decades spanning 1980 through 2010, driven by the entry of small-to-medium sized firms and the prevention of exit among large, older firms.

These margins are at the crux of two novel empirical results on immigrant absorption in U.S. economies. First, firm entry and the prevention of firm exit account for more than two-thirds of immigrant-induced job creation, while the expansion of continuing firms plays a minimal role over the span of a decade. Second, the prevention of firm exit, on average, masks a key moderator: using a set of productivity correlates and following a panel of over four million firms over time from 2000 to 2010, I find that increased exposure to immigrant workers culls lower-productivity firms from the market while making higher-productivity

[6] See, e.g., Borjas (1986); Fairlie and Lofstrom (2015); Kerr and Kerr (2018); Hunt (2011).
[7] See, e.g., Kerr and Kerr (2016).
[8] See, e.g., Mitaritonna et al. (2017).

firms far less likely to shut down. These results suggest a productive reallocation through creative destruction. A second panel that follows over 500,000 firms covered by the 2007 Survey of Business Owners adds nuance to this primary finding: shut-down probabilities in response to immigrant inflows are lower for all high-productivity firms, regardless of ownership nativity; however, among lower-productivity firms, those owned by natives are much more likely to shut down than those owned by immigrants. These results are highly suggestive of ties between immigrant entrepreneurs and employees.

In order to synthesize these results and evaluate their consequences in general equilibrium, I develop a model of the U.S. economy that incorporates elements from workhorse labor and trade models (e.g., Melitz, 2003; Ottaviano and Peri, 2008). The production side of the economy features fixed operating costs and entrepreneurs with heterogeneous ability while the consumption side demands variety, giving rise to a non-trivial productivity distribution across firms. The production function allows for imperfect substitutability across immigrant and native workers within the same education group. In the spirit of Bustos (2011), I then incorporate another, crucial heterogeneity: some firms pay additional fixed costs to access a technology that allows them to better utilize immigrant employees. Because these firms must pay an additional cost, they are positively selected on productivity. When immigrant exposure increases, these immigrant-heavy firms see larger reductions in labor costs than their lower productivity, immigrant-light counterparts. The resulting increase in competition forces the lowest productivity firms to shut down. I show that the effect of immigration on native welfare—the "immigration surplus"—in this model hinges on firm entry and firm reallocation. Immigrant-induced changes to the productivity composition and mass of firms generate first-order welfare benefits. In comparison, second-order effects that also arise in canonical, representative firm models of production are relatively muted.

The rest of this chapter is organized as follows: Section 1.2 provides a brief literature review. Section 1.3 documents the positive relationship between immigrant inflows and increased business presence and culminates by showing how this relationship drives immigrant absorption. Section 1.4 analyzes firm shut down decisions, stratified by correlates of initial firm productivity. Section 1.5 synthesizes the results from Sections 1.3 and 1.4 and discusses welfare implications using a theoretical model. Section 1.6 concludes.

1.2 Related Literature

The empirical analyses presented in this chapter add to a growing literature that utilizes advances in data availability to study firm-level responses to immigration. Three important, related papers contain results regarding the effect of immigration or foreign-born workers on the number of establishments in local labor markets. Olney (2013) presents evidence that low-skilled immigrants generate increased establishment presence in the 30 largest U.S. metropolitan areas using a yearly panel that covers 1998 through 2008. These findings are consistent with those presented in Section 1.3, and help motivate the expanded set of empirical analyses I present below. Section 1.3 documents a broad-based relationship between immigrant workers and net firm entry, including in non-tradable industries where Olney (2013) finds no effect. Beerli and Peri (2015) and Ruffner and Siegenthaler (2016) both find that the abolition of restrictions on cross-border commuters in Switzerland led to increased establishment presence in areas most affected by the policy. Unlike this study, the workers they analyze are not immigrants since they continue to live in foreign countries and were mostly high-skilled. Nonetheless, their results illuminate another context in which the effect of foreign workers on establishment creation is not likely due to consumer demand, given that they still live abroad. This chapter makes a similar argument regarding inflows of immigrants into the U.S. that are relatively low-skilled.

Two studies utilize employer-employee linked data to study how firms mediate and are ultimately affected by immigrant absorption. Dustmann and Glitz (2015) use employer-employee linked data from Germany to test the importance of alternative immigrant absorption mechanisms. As in this chapter, they focus on adjustment mechanisms other than wage changes. They find that 15 percent of immigrant absorption in the tradable sector can be accounted for by net firm creation. This result provides important motivation for the decomposition exercise presented in Section 1.3.5, though results are not directly comparable because my analysis includes non-tradable industries, is conducted fully within-sector, and takes place in the U.S. Mitaritonna et al. (2017) study the effect of immigration on existing firm productivity in France. They find that exposure to immigrants increases firm productivity and decreases the probability of firm exit, the latter of which is also found in Section 1.4 here. However, Mitaritonna et al. (2017) do not find that exit prevention is stratified by firm productivity. Once again, this discrepancy may be caused by their focus on the manufacturing sector, differences between the French and U.S. economies as a whole, the fact that immigrant inflows were skill-intensive in their setting, or differences in how initial firm productivity is measured. Future work reconciling these differences will help clarify the external validity of the results contained in this chapter.

Two aforementioned papers are also related to Section 1.4. Peri (2012) analyzes a state-decade panel and finds that immigrant inflows into the workforce generate increases in total factor productivity, resulting in an immigration surplus that is augmented by endogenous technological change. At least half of this technological change is mediated by task reallocation and increased specialization by native workers in response to immigrant inflows. In conjunction with the aforementioned literature on the importance of firm entry and exit to productivity gains, the empirical results presented in this chapter suggest that some of this task specialization may occur through reallocations that are enabled by firm entry. Hong and McLaren (2015) find that immigrant inflows lead to spillover job creation—specifically, that each immigrant generates 1.2 jobs in local labor markets that they arrive in. They also find that effects are primarily driven by the nontradable industries, which leads them to conclude that immigrant consumer demand is responsible for local spillover job creation. In contrast to their motivation (but not their results), my empirical specification attempts to control for the effect of immigrant-specific consumer demand, instead focusing on adjustments to immigration that are made on the supply side of the product market.

In general, my results also suggest that entrepreneurship plays an important mediating role in how immigration affects a local economy's dynamics. This notion is broadly in line with Beaudry et al. (2018), who reintroduce the importance of entrepreneurship to labor demand using a search theoretic framework. It also relates to recent work by Karahan et al. (2019) and Hopenhayn et al. (2018), who model and test for the importance of population growth declines in explaining declining entrepreneurship in the U.S. economy. The results contained in Section 1.3.1 imply that immigration can help alleviate the declining start-up rate simply by replenishing population growth. However, this chapter's focus is on the unique extensive margin labor demand responses to immigration shocks, motivated by unique features that immigrants bring to local economies. For example, results in Section 1.3.4 indicate that immigrant inflows increase firm presence, even when controlling for overall employment growth. Additionally, results in Section 1.4 imply that immigrant-owned firms benefit the most from access to immigrant employees, consistent with previous evidence that immigrant employees disproportionately work at immigrant-owned firms (Garcia-Perez, 2009). This chapter is thus also related to a budding literature on immigrant entrepreneurship in general (Kerr and Kerr, 2016, 2018; Fairlie and Lofstrom, 2015), which has recently broadened its previous focus from only the high-tech sector and the sciences (e.g., Kerr 2013; Hunt 2011; Hunt and Gauthier-Loiselle 2010; Peri et al. 2015; Kerr et al. 2016). However, as discussed in Section 1.3, the magnitude of the effect of immigrant presence on firm presence means that immigrant entrepreneurship alone cannot be the only mechanism. Thus, this chapter also relates to literature on how the presence of immigrant workers can spur native entrepreneurship (e.g., Duleep et al., 2012).

Section 1.5 develops a theoretical model in which heterogeneous firms distinguish between immigrant and native workers. Simply allowing for these heterogeneities clarifies the large role individual firms play in determining the effects immigrant workers have on local economies. In particular, increased firm presence benefits native consumers by increasing product variety, and reallocation from less to more productive firms (through culling of the former and entry by the latter) reduces prices. Seminal work by George Borjas (see, e.g., Chapter 7 in Borjas, 2014) motivated these "immigrant surplus" calculations—where immigrant surplus is defined as the effect of immigration on native welfare. As in his calculations, immigrants in my model generate increases in average native welfare when they arrive with a different skill mix than the incumbent population. Ottaviano and Peri (2012), meanwhile, show that immigrants can further increase native welfare when they are imperfect substitutes in production for native workers of the same skill and experience. I show that both of these results are amplified when we account for heterogenous firm responses. Waugh (2017) also studies firm dynamics in response to immigration in a model with monopolistic competition. While his focus is high-skilled immigration and the H1-B visa program, he similarly finds that immigrant inflows result in net firm creation. However, the mechanisms through which this occur are increased scale and an endogenous relationship between productivity and high-skilled labor. Motivated by Bustos (2011)—who introduces endogenous technological change to the Melitz framework by allowing a subset of firms to pay a higher fixed cost to access a better production technology—and Blaum et al. (2018)—who show how heterogeneity in imported input use can dramatically alter the effect of shocks on firm prices—I instead allow for heterogeneity in the ability of firms to utilize immigrant labor. This introduces an alternate channel through which immigrants spur net firm creation and increase productivity. Resulting price decreases that increase consumer welfare have been studied by Cortes (2008) (through a wage reduction channel) and di Giovanni et al. (2014) (through an increased variety channel). The productive reallocation channel is a novel addition to this theoretical literature.

1.3 Immigration, Firms, and Local Labor Markets

A bevy of previous literature indicates that firm entry is integral to job creation in the overall U.S. economy (e.g., Decker et al., 2014), and that immigrant workers—both as entrepreneurs and employees—tend to be over-represented in new firms (e.g, Kerr and Kerr, 2016). This section probes the relationship between immigrants and firm presence, then details its characteristics and its importance to immigrant-induced job creation.

1.3.1 Evidence from Case Studies

In order to establish a causal empirical relationship between immigration and business presence, I separately analyze, then pool together, two well-known case studies from the immigration literature—the 1980 Mariel Boatlift in Miami and the 2008 enactment of the

Legal Arizona Workers Act (LAWA). The structure of the analyses presented in this sub-section can be thought of as mimicking an indirect least squares approach to instrumental variable estimation. In each case, I establish a "first stage," in which an event brought an economically significant inflow (or outflow) of immigrants into (or out of) a local geography, either using existing estimates in the literature or through my own estimation. I then estimate "reduced form" effects of these events on the counts of establishments in those geographies using the Synthetic Control Method (SCM) (Abadie et al., 2010). Finally, using a pooling method proposed by Dube and Zipperer (2015), I combine the case studies and generate harmonized estimates of the dynamic effects of immigration on net firm presence. The case studies that I focus on broadly represent two distinct push factors for immigration that are common in the U.S.—shocks originating in sending countries and policies meant to deter undocumented immigration. Continuing the analogy to instrumental variables, these case studies will thus estimate different local average treatment effects, and pooling them together will generate an average of these different effects.

Despite its numerous benefits, because of the wide range of covariates that can be used in a given SCM model, the methodology can be susceptible to cherry-picking (Ferman et al., 2016). In order to circumvent these concerns, I choose the predictor variables for each synthetic control model for establishments per worker using a cross-validation procedure, also proposed by Dube and Zipperer (2015). Briefly, the procedure selects the set of control variables that best predicts donor unit outcomes in the post-treatment period.[9] The utility of this methodology can be seen in selection of sector shares for inclusion in the model. Geographic areas, including Arizona, that had high concentrations in the construction industry because of the housing bubble were harder hit by the Great Recession. These shares, then, serve as key control variables because they ensure that Arizona and its metropolitan areas are compared to other areas that were hard-hit by the Great Recession for the same underlying reason. Additional outflows of immigrants that occurred in Arizona after 2008, relative to these other hard-hit areas, are then more credibly claimed to be the result of LAWA rather than the Great Recession. The cross-validation procedure selects sector shares for inclusion precisely because they help accurately predict post-recession (post-2008) local economic paths in donor geographies. Section A.1.1 provides a more detailed overview of SCM, while Section A.1.2 provides additional details of the cross-validation procedure.

[9] "Donor" units are all geographies other than the treated unit. Each receives a non-negative weight. The weighted sum of donor unit outcomes, using these weights, is the "synthetic control" outcome.

1.3.1.1 Background and Data

The first case study I examine is the Mariel Boatlift, representing variation generated by push factors that occur abroad. The Mariel Boatlift saw roughly 125,000 Cubans immigrate to Miami between May and September of 1980 after Fidel Castro announced that he would open the ports of Mariel, Cuba to those seeking to leave the country. Beginning with the seminal work of Card (1990), the literature has gone back and forth regarding the consequences of this labor supply shock on the wages of substitutable workers in Miami, where a majority of "Marielitos" ended up settling (Peri and Yasenov, 2015; Borjas, 2017; Clemens and Hunt, 2017). Taking advantage of this existing literature, I rely on previous estimates of the size of the Mariel inflow shock rather than estimating it myself: the "first stage" effect of the Mariel Boatlift was to increase Miami's foreign-born workforce by 6.6 percent relative to initial employment (Card, 1990). However, I follow a more recent literature in examining how the Mariel Boatlift affected alternate adjustment channels, beyond wage changes (e.g., Anastasopoulos et al., 2018). I seek to extend this work by examining its effect on establishment presence in the Miami metropolitan statistical area.

The second case study I consider is the introduction of the Legal Arizona Workers Act (LAWA) in 2008, representing variation generated by sub-national legislation meant to curb the presence of undocumented immigrants in local areas. LAWA attempted to achieve this aim by requiring all employers to verify the eligibility of their new hires using the national E-Verify system. Bohn et al. (2014) previously studied LAWA, and found that it produced a robust and statistically significant drop in the number of Hispanic residents with less than a high-school degree in Arizona—the group most likely to contain undocumented immigrants. Unlike the Mariel case, I estimate the "first stage" affect of LAWA on overall immigrant presence since direct analogs have not been previously estimated.[10] This first stage model is estimated using the same covariates selected for the "reduced form," establishment outcome. In order to keep geographies consistent across cases, I specifically study Arizona's largest metropolitan statistical area, Phoenix-Mesa-Scottsdale, and drop the remaining Arizona cities from the analysis. Full results for the Arizona case at the state level can be seen in Section A.1.4.

The data for the case study analyses come from IPUMS-USA and the County Business Patterns. Using IPUMS-USA (Ruggles et al., 2019a), I construct measures of immigrant

[10]Bohn et al. (2014) study the change in the proportion of residents that are Hispanic with less than a high-school degree—a proxy for the proportion of residents that are undocumented. In order to retain consistency with the Mariel case and with the independent variable studied in Section 1.3, I study the change in the overall immigrant workforce relative to 2007 employment in Arizona. Nonetheless, the results of this exercise are wholly consistent with Bohn et al. (2014)—both find that LAWA generated a severe drop in immigrant presence.

presence for 1980, 1990, and 2001 through 2004 at the state level and 2000 and 2005 through 2014 for both the state and MSA level. Data for the years 1980, 1990, and 2000 are from public-use Decennial Long-Form Census microdata, while data for 2005 onward are from public-use American Community Survey microdata. The County Business Patterns (CBP) contains counts of establishments by industry and size at the county level going back to the 1960s. For consistent coverage, I use the CBP from 1970 forward. The CBP also generates key control variables: log employment as a proxy for market size and sector shares. These data were accessed directly from the Census Bureau and from the University of Michigan's Inter-university Consortium for Political and Social Research (ICPSR). These counts, divided by employment in the year before treatment ($t^* - 1$) serve as the outcomes in the analysis described below.

1.3.1.2 Individual Case Results

Figure 1.1 shows standard SCM results for the two events examined in this chapter. Along with plotting the treated unit's outcome ($y_{gt,\text{treated}}$ for a given outcome variable y and treated geography g) and the synthetic control outcome ($y_{gt,\text{synthetic}}$) in the top row, the bottom row also shows the evolution of the difference-in-differences estimates between the two ($[y_{gt,\text{treated}} - y_{gt,\text{synthetic}}] - [y_{g,t^*-1,\text{treated}} - y_{g,t^*-1,\text{synthetic}}]$, where $t^* - 1$ is the year before treatment) compared to this same difference for each potential donor unit, after constructing a "placebo" synthetic control for each donor unit in its bottom row. Comparing the thick black line (which plots the evolution of the treated unit) to the gray lines (one for each donor unit) generates a visual test of how unique the treated unit is in its behavior both before and after the event. The predictor variables include every-five-year pre-treatment averages of the outcome[11], along with log employment, sector shares, the under-40 year-old share, the self-employed share, and the college share at time $t^* - 1$. These variables were selected using the cross-validation procedure described above.

As found in Bohn et al. (2014) for Arizona, Phoenix shows a stark decrease in its immigrant workforce after the passage of LAWA in Figure 1.1. This is true both for the state as a whole and for its largest metropolitan area, Phoenix-Mesa-Scottsdale.[12] Note that immigrant workers can be employees or self-employed, so this outflow captures actual decreases in the number of immigrants relative to the counterfactual, not just the also-sizable response of switching to self-employment documented in Bohn and Lofstrom (2012).

[11] i.e., 1970-1974 and 1975-1979 for Miami and 1970-1974, 1975-1979,...,1995-1999, and 2000-2007 for Phoenix/Arizona.

[12] See Figure A.1 for Arizona as a whole. Results do not change qualitatively when considering Phoenix compared to Arizona as a whole. From this point forward, I will use "Phoenix" and "Phoenix-Mesa-Scottsdale" interchangeably.

For the Phoenix cases, I define

$$\delta^{1S} \equiv (I_{\text{treated},t^*+5} - I_{\text{synthetic},t^*+5}) - (I_{\text{treated},t^*-1} - I_{\text{synthetic},t^*-1})$$

where I_{gt} is the immigrant worker count in year t and geography g divided by the overall year $t^* - 1$ worker count in g. I use five years later as a benchmark to define permanent changes to the labor force induced by LAWA and to retain consistency with similar estimates for the Mariel Boatlift in Card (1990). For Phoenix, Table 1.1 shows these estimates along with "empirical" p-values, generated after conducting exact inference based on the rank of the δ^{1S} for the treated unit relative to the rank of the same parameter estimated after creating a synthetic control for all donor units (see Abadie et al., 2010 for details of this inference procedure). This manner of inference formalizes the aforementioned visual placebo tests in the bottom rows of Figures 1.1.

Miami's first stage is not included in Figure 1.1 because the Mariel Boatlift occurred before yearly estimates of immigrant populations are available for the U.S.[13] However, as mentioned above, the number of Marielitos and total employment in Miami in 1979 are well-documented in Card (1990). These allow me to construct an estimate of δ^{1S} for Miami, shown in Table 1.1. As a numerator, I use Card's estimate that the Mariel Boatlift generated a permanent increase of 45,000 foreign-born workers in Miami. As a denominator, I use his tabulated estimate of the 16-61 year-old Miami labor force in 1979, 678,200.[14] Comparing the first column in Figure 1.1 to the second two columns reveals that both Miami and Phoenix appear to feature a positive relationship between firm presence and immigrant inflows. This is despite the fact that the origins of these inflows are substantively different, as described above.

1.3.1.3 Pooling Cases

Despite the suggestive nature of the point estimates, the reduced form results presented in the second and third columns Figure 1.1 come from distinct first stages—both in terms of size and sign. In order to harmonize these analyses and double my statistical power, I pool cases and conduct an exercise similar to two-stage least squares. This approach is motivated by Dube and Zipperer (2015), who pool SCM cases to study the effects of minimum wage changes. As a first and aforementioned step, I focus on the largest metropolitan area in Arizona—Phoenix—rather than the state as a whole and drop the other Arizona cities from

[13] Peri and Yasenov (2015), for example, use the "Hispanic" variable in the CPS to get around this.

[14] Note that in Card (1990), Miami is only defined as Miami-Dade County, and both the numerator and denominator here reflect that county only. My outcome and control variables include Ft. Lauderdale and West Palm Beach—the entire Miami commuting zone.

the analysis.[15] This harmonizes the geographies being studied to the MSA level while keeping the number of events being studied to two.

I then define

$$\delta_t^{RF} \equiv (y_{\text{treated},t} - y_{\text{synthetic},t}) - (y_{\text{treated},t^*-1} - y_{\text{synthetic},t^*-1})$$

where y_{gt} stands for the number of establishments in year t and geography g divided by year t^*-1 employment in g. Where δ^{1S} can be thought of as a "first stage" in instrumental variable parlance—the causal effect of a given event on immigrant inflows five years after—we can similarly think of δ_t^{RF} as a "reduced form"—the causal effect of that same event on net establishment entry in year t.

I can then utilize the indirect least squares formula and define

$$\beta_t^e \equiv \frac{\delta_t^{RF}}{\delta^{1S}}$$

where e stands for a given event (i.e., $e \in \{\text{Mariel Boatlift,LAWA}\}$). Then, following Dube and Zipperer (2015), I estimate an average across events to recover a single estimate of the effect of immigrant inflows on net firm creation in year t

$$\beta_t \equiv \frac{1}{E} \sum_e \beta_t^e$$

where E is the total number of events ($E = 2$ here). I estimate β_t for $t \leq 5$ to retain consistency with δ^{1S} and to avoid confounding factors that can arise over longer horizons. The interpretation of β_5 is then the number of firms associated with each exogenously pushed, permanent immigrant over the five year period after the initial shock. β_1 through β_4 show us how the generation of β_5 evolves.

As described above, individual SCM results generally utilize exact inference based on the rank of the difference-in-difference estimator for the treated unit relative to the rank of the same parameter estimated after creating a synthetic control for all donor units. This percentile rank has a uniform distribution, enabling simple calculation of empirical p-values. Dube and Zipperer (2015) extend this concept to pooled case studies, where inference of the mean of these difference-in-difference estimates (such as β_t) estimated by SCM is facilitated by the fact that the sum of uniform random variables has a known (Irwin-Hall) distribution. Further details on inference used in this section can be seen in Appendix Section A.1.3 and

[15] An alternate strategy is to expand our focus from Miami to the state of Florida and pool case studies at the state level. The results from this pooling exercise can be seen in Section A.1.4.

Dube and Zipperer (2015).[16]

The primary result of this harmonized, pooled SCM exercise at the metropolitan level can be seen in Figure 1.2. Pooled estimates from Miami and Phoenix in pre-treatment years hover around and are not statistically distinguishable from zero, passing the standard event study pre-trends test. Five years after the immigration shock, the estimated effect of immigrant inflows on establishment presence is roughly 0.08. This number is economically meaningful: if we were to apply it to overall immigrant inflows into the U.S. between 1980 and 2010, it would indicate that immigrants were responsible for 40 percent of net establishment creation in that period despite only being responsible for 30 percent of net workforce growth. As expected, this number is also greater than the estimated effects found below in Section 1.3, where any effects of consumer demand are limited by comparing across industries within commuting zones. However, it is strikingly similar to estimates in Section 1.3 that do not control for consumer demand. Unlike the decade-by-decade analysis featured in Section 1.3, however, Figure 1.2 also sheds light on the dynamic responses of local labor markets to immigration shocks. These responses start relatively rapidly, but continue to show an upward trend throughout the analysis window.

1.3.1.4 Discussion

The results in this section imply a causal link between immigrant inflows and establishment presence. They also raise a new set of questions regarding its nature. In particular, that more immigrants lead to more firms may be the result of increased overall consumer demand, immigrant-specific consumer demand, general labor cost reductions, or immigrant-specific characteristics from the supply side of the product market. Due to data limitations, they are also unable to shed light on whether the link between immigration and firm presence comes from firm entry, the prevention of firm exit, or a combination of the two.[17] Thus, while the results contained in this section provide clear identification, they ultimately serve

[16] Note that I only use δ^{1S} as a scaling factor here—I do not account for variance in its estimation in the pooling exercise. The confidence intervals should be understood as only applying to the reduced form.

[17] Previous drafts of this chapter used the U.S. Census Bureau's Business Dynamics Statistics (BDS) at the metropolitan statistical area (MSA) level for this analysis. The BDS has substantial advantages over the CBP: it counts firms as well as establishments, and counts firm entry and exit separately, allowing researchers to study the flows that change the firm stock. However, two disadvantages regarding the Miami case study render it currently ill-suited for use here. First, the synthetic control method is suited for cases with many pre-treatment years, whereas the Mariel Boatlift occurred in 1980 and the BDS starts in 1977. Second, the BDS appears to have a significant discrepancy with the CBP regarding establishment and employment counts for the Miami MSA. In 1980, the BDS measures 524,350 employees, whereas the Decennial Census measures 1,057,240 and the CBP measures 982,983. Future drafts plan to resolve this discrpenacy and combine information from the CBP and BDS to both give the SCM enough pre-treatment years to properly find a synthetic control unit for Miami, and to be able to take advantage of the rich data contained in the BDS.

more as a proof of concept. The rest of this section, then, utilizes detailed data to probe 1) whether the effect found here is broad-based beyond the specific cases of Arizona and Miami; 2) whether some of it can be accounted for by the effect immigrants have on the production side of the economy, as workers and entrepreneurs as opposed to as consumers; 3) its heterogeneous nature in terms of firm age, size, and other characteristics; and 4) how much it matters for immigrant absorption.

1.3.2 A Broader View: Data and Sample

These further analyses are facilitated by access to confidential data from the U.S. Census Bureau's Longitudinal Business Database (LBD), which contains survey responses that are mandated by law from each U.S. non-farm, employee-hiring, private-sector establishment. Establishments are assigned unique, consistent identifiers that can be linked over time to create a true panel. Crucially, the LBD also contains unique firm identifiers, which allows me to aggregate establishments to their owning firms. The focus on firms is distinct from a focus on establishments because it carries the possibility of a new "variety" from the perspective of the consumer and an additional outside option from the perspective of a worker.[18] Fort and Klimek (2018) provide consistent North American Industrial Classification System (NAICS) codes over time for these firms which I then map to industry groups described below. This facilitates within-industry analyses that more credibly isolate the effect of immigrant workers on firm entry and exit decisions.

The LBD allows me to go beyond the publicly-available, overall counts of establishments in a given geography—studied in Section 1.3.1—by allowing for firm counts within geography *and* industry group *by* various firm characteristics. The most important of these characteristics is firm age. Haltiwanger et al. (2013) find that much of the public attention paid to small businesses and their role in job creation should be transferred to young businesses, which are generally but incidentally small. Furthermore, as documented in Decker et al. (2014), new firm creation also appears to play a particularly crucial role in both job creation and productivity. That the longitudinal linkages in the LBD facilitate the construction of within sector-geography firm counts by age bin allows me to understand whether immigrant inflows are preventing continuing firm deaths, spawning new (under 10 years of age) firm activity, or both. In addition, the LBD also allows for within industry group-geography firm counts in size-by-age bins. These splits facilitate a more complete characterization of the increased firm presence that arises with immigrant inflows and lead to specific hypotheses regarding its consequences that are tested in Sections 1.3.5 and 1.4. A firm is counted in a given commuting

[18] For additional reasons to favor the study of firms over establishments in a similar context, see Haltiwanger (2015).

zone and industry if it owns at least one operating establishment in that commuting zone and industry. This means the same firm can be counted multiple times, if it operates in multiple locations and/or industries.

In order to study the effect of immigrant presence on these outcomes, I also exploit restricted-access U.S. Census Bureau demographic data from the 1980, 1990, and 2000 Long-Form Decennial Censuses and the 2005 through 2012 American Community Survey. These data allow for unusually precise measures of immigrant inflows, not just into geographies, but into relatively detailed industry groups within geography by country of origin. These elements are crucial to the identification strategy presented in Section 1.3.3. Because industry classifications differ between the Census and NAICS, I construct aggregated industry groups using the 1990 Decennial Census industry codes as a bridge between Census industry codes in other years and 3-digit NAICS codes contained in the LBD.[19] In some cases, the 1990 industry classification corresponds to more than one 3-digit NAICS code, and in some cases a 3-digit NAICS code corresponds to more than one 1990 industry classification. The industry groups I use therefore generally represent the smallest possible mutually-exclusive sets of industry classifications.[20] Some additional aggregations are made to ensure that industry groups do not vary excessively in size. The agriculture, mining, and public sectors are dropped from the analysis due to relatively less reliable coverage in the LBD. The final set of industry groups can be seen in Table 1.4.

The analyses presented in Section 1.3.4 and Section 1.4 are thus based on immigrant exposure in commuting zone-industry pairs over time, where industry is defined in Table 1.4. Commuting zone groupings are provided by David Dorn, as used in Autor and Dorn (2013). In order to limit the impact of measurement error and take full advantage of the precision allowed for by the restricted-access demographic data, I make two sample restrictions. First I keep commuting zone-industry pairs with at least 100 workers in 1980. Second, I regress LBD-measured employee counts on measured private sector employee counts from the demographic data across decades at the commuting zone-industry group level. These two measures should correspond closely, and outliers indicate reason for caution in combining the data sources for further regression analysis. I thus eliminate commuting zones-industry pairs with maximum squared residuals (across decade) above the 75th percentile.

[19] Crosswalks provided by IPUMS-USA between the 1990 and other Census year classifications, as well as between the 1990 Census industry classifications and NAICS codes, were crucial to this process.

[20] For example, 1990 Census Industry classification code 132 is "Knitting mills" and corresponds to NAICS Codes 315 and 313. However, NAICS code 313 also covers "Yarn, thread, and fabric mills," which is 1990 Census Industry classifiction code 142. Additionally, NAICS code 315 also includes manufacturing of "Apparel and accessories other than knitting." Manufacturing of apparel and accessories, knitting mills, and yarn, thread, and fabric mills are therefore all covered in the same industry grouping in my analysis.

1.3.3 A Broader View: Identification

The primary specification for the analyses in Section 1.3.4 is

$$\Delta y_{gkt} = \alpha + \beta\left(\Delta I_{gkt}\right) + \Gamma X_{gkt} + \alpha_{gt} + \alpha_{kt} + \varepsilon_{gkt} \tag{1.3.1}$$

where g indexes a commuting zone, k indexes an industry group[21], t indexes a year, and the Δ operator represents a ten-year change within a commuting zone and industry group (change within gk). For example, in Columns 2 and 3 of Table 1.5, Δy_{gkt} is the change in firm presence in commuting zone g and industry group k during a given decade, divided by the start-of-decade workforce in gk. The independent variable of interest, ΔI_{gkt}, is the change in immigrant worker stock in gk between year $t-10$ and t, divided by the start-of-decade workforce in gk. In that specification, the coefficient of interest β measures the number of firms created, on net, per immigrant—the same interpretation as the results from Section 1.3.1. X_{gkt} is a vector of control variables that can include 1980 commuting zone-industry characteristics interacted with year dummy variables and a Census region-industry-year fixed effect. The analyses covered by this specification span three decades (1980-2010), in which there were large immigrant inflows to the U.S., and covers nearly the entire geography of the U.S. (more than 700 commuting zones). Thus, it dramatically expands on the external validity and policy relevance of the results found in Section 1.3.1.

Even with the rich fixed effect structure contained in Equation (1.3.1), however, endogeneity concerns regarding immigrant industry choices within geographies and geographical choices within industry remain. Immigrant employees, for example, may choose to work in booming industries, generating biased estimates of β. Immigrant employees may also be more adept than native workers at locating to areas that are booming, as found in Cadena and Kovak (2016), even if they work in the same industry. Meanwhile, if immigrant entrepreneurs are attracted to geographies where they face less competition or if immigrant employees are linked to large firms in more concentrated markets, ordinary least squares (OLS) estimates of β with firm presence as an outcome may contain a downward bias. Measurement error may also play a role here, even with our sample restrictions, given that immigrant inflows can be small (and thus estimated from relatively few unweighted sample observations) within commuting zone-industry group and that Equation (1.3.1) is a panel model. This could generate substantial attenuation bias in any OLS estimate of β. In short, even in a relatively saturated model, isolating exogenous variation that pushes immigrants into commuting zone-industry pairings substantially strengthens causal interpretation of β at the cost of

[21] When noted, k may index a 1-digit SIC Sector. However, when not noted, it references the industry groups listed in Table 1.4.

reducing estimate precision.

A standard shift-share instrument for immigrant stock I_{gt} in a panel with commuting zone-year observations would take the following form:

$$z_{gt} \equiv \frac{1}{E_{g,1980}} \sum_{o} \pi_{go,1980} \times I_{o(-g)t}$$

where 1980 serves as the base year and o indexes a worker's origin country. $\pi_{go,1980}$ is the share of origin country o's stock of immigrants in 1980 that was located in commuting zone g, $I_{o(-g)t}$ is the overall stock of immigrants from country o in year t for all commuting zones other than g, and $E_{g,1980}$ is the 1980 workforce in commuting zone g. Δz_{gt} would then serve as an instrument for ΔI_{gt}, the overall change in immigrant presence in commuting zone g between $t-10$ and t, by distributing immigrant inflows to commuting zones based on network effects operating through initial stocks $\pi_{go,1980}$.

Previous literature has recognized that replacing $\Delta I_{o(-g)t}$ with variables that capture exogenous push factors from sending country o in Δz_{gt} can make the exclusion restriction more plausible. This notion accords with recent work by Borusyak et al. (2018), who demonstrate that when shift components are as good as randomly assigned conditional on shares, shift-share instruments do not violate the exclusion restriction. Llull (2017) thus uses conflict, natural disasters, changes in per capita income, and changes to political regimes as aggregate push factors. He takes the additional step of replacing $\pi_{go,1980}$ with distance because he works in a setting with cross-country migrant destinations.[22] Monras (2015) hones in on one sending country, Mexico, and uses the Peso Crisis of 1995 as an exogenous push factor, interacting it with 1980 state shares of Mexican immigrants to generate exogenous variation in Mexican inflows. Angrist and Kugler (2003) interact distance from Bosnia and Kosovo with indicators for years in which wars were taking place in those locations.

In this chapter, I follow a modified version of this strategy that also borrows from Autor et al. (2013). I take advantage of the German Institute for Employment Research (IAB) Brain-Drain Data, a unique source that contains counts of bilateral emigrant stocks from more than 150 sending countries worldwide residing in OECD member nations (Brücker et al., 2013). I am thus able to replace $I_{o(-g)t}$ with $M_{ot}^{\text{non-US}}$, a measure of emigrants from origin country o living in all OECD member nations *other than* the United States:

$$z_{gt}^{\text{Emigrants}} = \frac{1}{E_{g,1980}} \sum_{o} \pi_{go,1980} \times M_{ot}^{\text{non-US}}$$

[22] That is, distance serves as an appropriate time-invariant, destination-specific interactor for Llull (2017) because it creates enough variation in how much it pushes sending country emigrants to different countries or different broad regions in North America. Distances between U.S. commuting zones and sending countries do not create enough variation for me to replace $\pi_{go,1980}$ here.

$\Delta z_{gt}^{\text{Emigrants}}$ generates plausibly exogenous variation in immigrant locational decisions within the U.S. under the assumption that outflows from origin countries to non-U.S. OECD countries are unrelated to local economic outcomes in the U.S. (Borusyak et al., 2018), or if base-year shares $\pi_{go,1980}$ do not affect local economies during the study period (Goldsmith-Pinkham et al., 2018). Just as Autor et al. (2013) claim that Chinese exports to non-U.S. countries reflect increases in Chinese export productivity rather than product demand in the U.S., I claim that these outflows are much more likely to reflect migration push factors in sending countries rather than labor demand in the U.S.

Because Equation (1.3.1) is identified from within commuting zone-year variation, it requires an instrument that varies at the industry level as well. The standard shift-share approach accommodates this disaggregation by constructing:

$$z_{gkt}^{\text{Standard}} \equiv \frac{1}{E_{gk,1980}} \sum_{o} \pi_{go,1980} \times I_{(-g)okt}$$

Here, the initial shares remain the same, but the aggregate component $I_{(-g)okt}$ is now the number of immigrants from origin country o *working in sector* k in all commuting zones other than g at time t. Note that

$$I_{(-g)okt} \equiv [\rho_{(-g)okt} \times I_{o(-g)t}]$$

where $\rho_{(-g)okt}$ is the proportion of origin o immigrants working in sector k in all commuting zones other than the commuting zone of interest, g. Thus, it still utilizes the network effects provided by $\pi_{go,1980}$ for relevance, but does so separately by sector. While the IAB data does not separate emigrant outflows by sector, the restricted-access demographic data from the Census Bureau does contain detailed information on both the country of origin and industry of workers. Thus, in order to turn $z_{gt}^{\text{Emigrants}}$ into a geography-industry level instrument, I make the following adjustment:

$$z_{gkt}^{\text{Emigrants}} = \frac{1}{E_{gk,1980}} \sum_{o} \pi_{go,1980} \times \rho_{(-r)okt} \times M_{ot}^{\text{non-US}} \qquad (1.3.2)$$

where r is the Census Region that commuting zone g resides in. This strategy takes advantage of the fact that immigrants from particular countries tend to specialize in certain industries due to comparative advantage, separate from demand in a particular local industry. Ultimately, $z_{gkt}^{\text{Emigrants}}$ then predicts the number of immigrants residing in a given commuting zone g based on network-induced locational preference, working in industry k due to country-specific comparative advantage, and pushed into the U.S. by factors stemming from their country of origin.

With our instrument fully detailed, we can now fully describe the utility of each fixed effect in Equation (1.3.1). For the decade ending in year t, α_{gt} removes any effects immigrant inflows have at the commuting zone level as a whole. Under the premise that immigrants do not solely demand goods in the industry in which they work, α_{gt} insulates β from being identified by changes in consumption patterns that can result from immigration (e.g., Hong and McLaren, 2015). This premise is strengthened by the fact that we compare across 40 industry groups within a commuting zone—a level of detail allowed for by the granularity in each Census Bureau data source. Absent α_{gt}, an inflow of immigrants into the "Hospitals" industry group can generate an increase in economic activity in other nontradable industry groups because the new immigrant workers in the "Hospitals" industry group also consume goods and services locally. By including α_{gt} and thus inducing comparison across industry groups within a given commuting zone and decade, β measures the increase in economic activity in the "Hospitals" industry group above and beyond what other industry groups experienced due to this consumer demand effect.

α_{kt} plays an important role in ensuring instrument validity. $\rho_{(-r)okt}$ allocates immigrants from origin country o into industry group k based on national level trends of industry choice for origin country o immigrants (excluding region r). In the absence of α_{kt}, national level shocks to industry k would naturally allocate all workers towards industry group k, regardless of origin. α_{kt} precludes these shocks—which would induce both an increase in economic activity in sector k across all commuting zones and in $\Delta z_{gkt}^{\text{Emigrants}}$—from contaminating β. Instead, with the inclusion of α_{kt} (or further, region-industry-year fixed effects), workers from origin country o must be locating in industry group k above and beyond the national trend, and in regions where they are not affected by labor demand shocks in commuting zone g. Thus, $\Delta z_{gkt}^{\text{Emigrants}}$ is more credibly sourced from immigrants' comparative advantages in certain industries through $\rho_{(-r)okt}$ when α_{kt} is included.

It is also useful to think through an example of how $\Delta z_{gkt}^{\text{Emigrants}}$ precludes reverse causality from identifying β. A simple and relevant example of labor demand pulling immigrants into specific geographies and industries comes from the housing bubble that metastasized between 2000 and 2005, largely in the South and West of the U.S. The housing bubble created a large labor demand shock for construction workers in the South and West Census Regions of the U.S., and induced immigrant workers from Mexico to fill this demand—the kind of inflow an instrumental variable should not use for identification of β. As seen in Panel A of Figure 1.4, Mexican inflows into the construction sector between 2000 and 2005 were more than 10 times larger than from the next closest country. As seen in Panel B of Figure 1.4, general immigrant inflows into the construction sector across commuting zones predominantly took place in housing bubble cities throughout the South and West regions of the country. On the other hand, there is no reason to believe that the U.S. housing bubble would cause large outflows of Mexican emigrants to non-U.S. OECD countries. Furthermore, relative to the national trend, the change in propensity of Mexican immigrants to locate in the construction sector, *outside* of the South and West regions is not unusually strong. Thus, the aggregate component of $\Delta z_{gkt}^{\text{Emigrants}}$ between 2000 and 2005

$$\left[\rho_{(-r)ok,2005} \times M_{o,2005}^{\text{non-US}} - \rho_{(-r)ok,2000} \times M_{o,2000}^{\text{non-US}}\right]$$

should not reflect the labor demand shocks in construction that were occurring in the South and West of the country at that time. These two factors are illustrated in Panel A of Figure 1.5, where Mexico has a much more modest aggregate component for the construction sector between 2000 and 2005. The ultimate result of these corrections can be seen in Panel B, where the instrument-predicted immigrant inflows are far less concentrated in bubble cities.

Results presented below, along with a series of analyses in Section A.2 test several many aforementioned and additional considerations regarding the validity of $\Delta z_{gkt}^{\text{Emigrants}}$ more systematically, with a particular focus on recently-formalized concerns regarding shift-share instrumentation. Table 1.5 and Section A.2.4 address instrument validity using various pre-trends tests. Section A.2.2 compares $\Delta z_{gkt}^{\text{Emigrants}}$ to $\Delta z_{gkt}^{\text{Standard}}$ along with another plausibly exogenous replacement for the "shift" component in a typical migration instrument—lagged birth rates in origin countries. It finds results from this third instrument, $\Delta z_{gkt}^{\text{Births}}$, and $\Delta z_{gkt}^{\text{Emigrants}}$ are nearly identical.[23] Section A.2.3 demonstrates that migrant outflows to non-US OECD countries are a relevant predictor of immigration to the U.S. at the origin country level. Section A.2.5 addresses concerns regarding correlated outcomes across observations with similar "share" components that can undermine inference when using shift-share instrumentation (Adao et al., 2019). It finds no evidence of such a problem in my triple-difference, commuting zone-sector level regressions. Section A.2.6, along with Column 6 of Table 1.5, alleviates concerns about serial correlation in the "shift" component of the instrument (Jaeger et al., 2018). All told, the emigrants-based instrument passes a battery of tests meant to vet its validity for use. $\Delta z_{gkt}^{\text{Emigrants}}$ and $z_{gkt}^{\text{Emigrants}}$ therefore become the instruments of choice for

[23] $\Delta z_{gkt}^{\text{Births}}$ utilizes lagged birth counts in sending countries of individuals who are of prime migration age by the time of our study period as a push factor. Because these are predetermined relative to the analysis, they can be considered a better source of exogenous variation. The reason this chapter does not use $\Delta z_{gkt}^{\text{Births}}$ as its preferred instrument, then, is out of practical rather than validity concerns. The demographic changes represented in $\Delta z_{gkt}^{\text{Births}}$ take years to unfold, making them much more amenable to decade (or longer horizon) level analyses. Meanwhile, the analyses in Section 1.4 require an instrument at a five-year frequency. Thus, Section A.2.2 shows that $\Delta z_{gkt}^{\text{Births}}$ delivers near-identical results to $\Delta z_{gkt}^{\text{Emigrants}}$. With this check in hand, I proceed with the emigrants instrument for the rest of the chapter.

Sections 1.3.4, 1.3.5 and 1.4.

1.3.4 Immigrant Workers and Firm Presence: a More Complete Accounting

1.3.4.1 Educational Content of Immigrant Inflows

Before delving into estimates from Equation (1.3.1), it is important to understand the skill content of the immigrant inflows represented by ΔI_{gkt}, particularly when it is instrumented for by $\Delta z_{gkt}^{\text{Emigrants}}$. As described in Borjas (1999), the more immigrants differ from natives in their skill content, the larger the potential economic surplus that is generated by immigrants in the long run. This insight is built into the model presented in Section 1.5 as well, generating a supply-side channel through which immigration can increase firm presence.

Educational attainment is perhaps the most important way through which a connection between immigration and economic activity can arise in a manner that does not just reflect general population growth, though it is far from the only. The notion that some production is specifically tied to foreign-born workers would not arise if immigrants were identical to natives in all aspects other than their country of birth. Rather, these ties are plausible because many immigrants have characteristics that are different than natives on average—e.g., extreme levels of education (known as the "twin peaks" phenomenon) or increased propensity to start businesses. The average set of native skills thus becomes more scarce when immigrants enter the economy, generating a surplus for natives. This is the key insight in Borjas (1999) that has been built upon in much subsequent work. Here, I consider this idea in the context of firm heterogeneity: some firms particularly rely on these different skills that immigrants bring to the economy, and when immigrants arrive into a local area, firm composition tilts towards these types of firms.

Figure 1.3 provides substantial evidence that immigrant workers differ in their education levels, a proxy for skill, relative to native workers. On average, a smaller proportion of foreign-born workers have more than a high school degree. This gap has increased over time, indicating that immigrant inflows have tended to be relatively concentrated among workers with less education. Table 1.2 confirms this intuition by analyzing the effect of immigrant inflows on the education distribution using Equation (1.3.1) and the change in the low-education share as the outcome. The results also illuminate an important feature regarding the interpretation of the results below: $\Delta z_{gkt}^{\text{Emigrants}}$ does not tend to disproportionately push immigrants of a certain education level to the U.S., relative to the average immigrant inflow. Table A.4 provides additional analysis to support this point.

1.3.4.2 The Effect of Immigration on Firm Presence Within Commuting Zone and Industry Group

Table 1.5 displays first stage, OLS, and IV results from estimating Equation (1.3.1) using a full set of controls and fixed effects. In it, I find a strong, positive effect of immigrant worker presence on firm presence that is larger when corrected for endogeneity. Each immigrant generates roughly 0.05 new firms, on net, when they enter a commuting zone and industry group (Column 3). This estimate survives flexible controls for any confounding factors that may affect firm presence in the commuting zone as a whole through α_{gt} (e.g., general immigrant inflows into the commuting zone or overall population growth) and any confounding factors that may affect firm presence in the industry group nationally through α_{kt} (e.g., secular growth in services sectors).

Previous research has documented that immigrants tend to migrate towards economic opportunity (e.g., Cadena and Kovak, 2016), which may lead to concern about the fact that IV estimates are more positive than their OLS counterparts in Table 1.5. However, comparison with results fully explained in Section 1.3.5 alleviates some of these concerns. Table 1.7 finds that IV estimates of immigrant-induced job creation are smaller than OLS estimates, wholly in line with the notion that immigrants migrate towards labor demand. This result, using the same independent variable and instrument as that used here, also argues against measurement error as a determinant of the relationship between OLS and IV estimates seen in Table 1.5. An alternate explanation is that immigrant entrepreneurs may be attracted to areas with lower net firm entry if these areas feature lower entry costs in the form of decreased potential competition. Another is that areas and industries with fewer, larger firms are more likely to attract immigrants.

Table 1.6 demonstrates the stability and robustness of the estimate in Column 3 of Table 1.5. First, moving from Column 1 to Column 2 indicates the utility of including α_{kt} and α_{gt} relative to a traditional first differences approach (just including α_t). Nearly half of the first difference estimate disappears when we compare across industry groups within a commuting zone, relative to when we simply compare across commuting zone-industry group pairs. These are effects likely explained by consumer demand, either through immigrant inflows that are correlated within geography across industry groups or through general population growth in the commuting zone. Columns 2 through 6 then display a stable estimate, regardless of control and fixed effects sets (note that Column 4 replicates Column 3 of Table 1.5). Of particular note, the estimate in Column 5 survives the inclusion of an endogenous control variable for overall growth in the number of workers, indicating that the effects found in

Table 1.5 and 1.6 are immigrant-specific, and not simply a general labor supply shock.[24] This column also implicitly controls for any non-nativity-specific, within-industry group demand that may generate β, and still finds essentially the same effect. Additionally, the estimates in Column 6 implement the double-instrumentation strategy proposed in Jaeger et al. (2018) to account for serial correlation that can undermine shift-share-based estimate interpretation. The relative strength of the F statistic and continued stability of the contemporaneous effect estimate indicate that Equation (1.3.1) passes this test. Finally, in all columns other than Column 6, I cannot reject the null hypothesis of no pre-trends based on a panel version of the test, described further in Section A.2.4.

Overall, Tables 1.5 and 1.6 generate the first benchmark result in this chapter—a broad-based, positive effect of immigration on firm presence within local markets defined by a commuting zone and industry group. It finds that each immigrant worker (employee or self-employed) leads to the creation of roughly 0.05 firms, on net. Because of the way Equation (1.3.1) is set up, and because there are over 40 industry groups, I contend that this effect primarily originates from the supply side of the product market, with changes in consumer demand largely soaked up by α_{gt}. In addition, the magnitude I estimate is economically significant. Back-of-the-envelope calculations using Table 1.3 help make this point, albeit at the SIC sector level. The mean of ΔI_{gkt} is 0.0474. Supposing these inflows are exogenous and multiplying this figure by $\beta \approx 0.05$ generates the estimate that an average immigrant inflow generates 0.00237 new firms. Using a benchmark mean for the change in firm presence from the change in establishment presence reflected in Table 1.3 (0.0112) implies that immigrant workers, through their effects on production, were responsible for roughly 21 percent of net firm entry within commuting zone and sectors between 1980 and 2010. Meanwhile, immigrants only made up 11 percent of the workforce during this period. Finally, while ΔI_{gkt} includes both employees and self-employed immigrants, back-of-the-envelope calculations also tell us that the effects found here are unlikely to be fully explained by immigrant entrepreneurship. Section A.4.1 shows that each immigrant pushed by $\Delta z_{gkt}^{\text{Emigrants}}$ leads to roughly 0.14 additional self-employed workers. It also shows that each self-employed worker is associated with 0.16 additional establishments. Interacting these two implies an immigrant-entrepreneurship effect of 0.022, less than half of the estimated $\beta \approx 0.05$. Immigrant employees appear to play a co-equal, if not central, role in generating the effects found here, and this motivates their role in the model presented in Section 1.5.

[24]The model presented in Section 1.5 hypothesizes that this is due to the variety of reasons immigrant and native workers may be imperfect substitutes, including educational attainment.

1.3.4.3 Decomposing the Effect: Entry and Exit

Figure 1.6 characterizes the results contained in Column 3 of Table 1.5 by separating the change in the stock of firms into its flow components::

$$\Delta \text{Firms}_{gkt} = \text{Firms}_{gkt}^{\text{Age}<10} - \text{Firms}_{gk,t-10}^{\text{Shut down by time } t} + \text{Residual} \tag{1.3.3}$$

That is, the change in the number of firms in commuting zone g and industry group k between $t-10$ and t can be split into its relative contributions from firms that started between time $t-10$ and t and are operating in gk, firms that were operating in gk at $t-10$ that have shut down by time t, and a residual term that primarily captures net relocations and expansions into a given gk pair from other gk pairs.[25] This decomposition is only enabled through the longitudinal linkages provided in the LBD. Here and throughout the rest of the chapter, firm age is measured as the age of its oldest establishment.

Figure 1.6 plots the results of estimating Equation (1.3.1) using these disaggregated outcomes, each divided by $\text{Workers}_{gk,t-10}$, such that they add up to 0.0527, the primary estimate of interest from Table 1.5. This figure demonstrates that 80 percent of immigrant-induced increase in firm presence comes from the creation of new firms, though there is also a non-trivial prevention of firm death. The residual term plays a negligible role, indicating that the extensive margin firm response to immigrant inflows is largely localized. Of note, the estimate marked "Survived to t+5" in gray indicates that immigrant-induced entrants are not necessarily short-lived. In fact, more than half of the firms in this category survive well past a firm's usual "critical period," until 5 years after the decade end represented by time t.

1.3.4.4 Decomposing the Effect: Firm Size

I next estimate Equation (1.3.1) by employee size bin (within gk), with two outcomes for each bin. The first is simply firm growth in that size bin:

$$\frac{\Delta \text{Firms}_{gkt}^{\text{Size Bin}}}{\text{Firms}_{gk,t-10}^{\text{Size Bin}}}$$

[25] Shut down is defined as all of a firm's *national* establishments having shut down. If all a firm's establishments in a given commuting zone and industry have shut down, but the firm is still in operation nationally, this firm would be counted negatively in the residual term.

The second only reflects the portion of that growth that came from firm entry:

$$\frac{\text{Firms}_{gkt}^{\text{Size Bin, Age}<10}}{\text{Firms}_{gk,t-10}^{\text{Size Bin}}}$$

I switch to growth rates as the outcome because the overall firm size distribution is heavily skewed right. As seen in Figure A.4, when purely accounting for the primary estimate of interest from Table 1.5, the smallest firm size bin dominates.[26] However, this is true for the overall firm size distribution in the economy. Thus, to know if immigrant-induced net firm entry is *particularly* driven by small firms, it is more useful to look at within-bin growth rates.

Figure 1.7 contains some of the most important results in this sub-section. It finds a robust, large effect of immigrant inflows on net firm entry throughout the firm size distribution, and one that is largest at both tails when viewed in within-bin terms. That is, when we account for the fact that most firms are small, immigrant-induced increases in firm presence are not disproportionately seen in the lower end of the firm size distribution. In addition, new firm creation almost entirely drives this effect through firms up to 99 employees. Starting with firms of at least 100 employees, a divergence emerges in which the increase total firm presence is also driven by older firms. These results are consistent with the notion that new firms tend to be smaller, but also illustrate a novel connection between immigrant workers and large firm death prevention.

1.3.4.5 Additional Heterogeneity

Figure 1.8 plots the results of additional heterogeneity analyses across industry groupings and geographic region. The top grouping provides evidence that the effects found thus far are not driven by one particular region. In particular, based on the coefficient in the West Census Region, California is not the driving force behind the results in this section. The next set of analyses designates each industry group (Table 1.4) as tradable or non-tradable, then estimates Equation (1.3.1) separately for each set of industry groups. I generate this designation by aggregating 1980 traded and non-traded employment within each industry group based on the Porter (2003) classification system for 6-digit NAICS codes. Each industry group is then designated as tradable if it had more tradable employment in 1980, and vice versa. The result of this exercise finds a larger effect in nontradable industry groups, but a large effect for tradable industry groups as well. The latter is found in Olney (2013), while the former is not. Finally, I also designate industry groups based on whether they

[26]Note that firms with "0" employees are usually births in year t.

tend to hire higher- or lower-educated workers. Similar to Doms et al. (2010), I do this by assigning industry groups with below the median share (across industry groups) of college equivalent[27] workers in 1980 the "low-education hiring" designation and industry groups with above the median share the "high-education hiring" designation. This exercise shows a slightly higher effect of immigrant workers on firm presence in low-education hiring industry groups.

1.3.4.6 Summary and Discussion

Immigrants appear to generate non-trivial increases in firm presence from the supply side of the product market, as workers. The various heterogeneities that underlie this effect set up the analyses that follow in Sections 1.3.5 and 1.4. Figures 1.6 and 1.7 introduce two important channels through which immigrant absorption can occur. The majority of the effect immigrant workers have on increasing firm presence comes in the form of small-to-medium-sized, new firms. While smaller firms may seem prone to less absorption capability, Decker et al. (2014) use the LBD to document that new firm creation plays a crucial role in job creation nationally, with start-up firms alone accounting for almost 20 percent of gross job creation in a given year. They also find that firms that grow by more than 25 percent year-over-year are disproportionately small and young, and yet account for almost 50 percent of job creation in a given year. Additionally, that there is a robust presence of new, medium-sized firms mitigates concerns that immigrants tend to primarily own and support "subsistence" entrepreneurship (Schoar, 2010). Meanwhile, Figures 1.6 and 1.7 also show a small effect on the prevention of firm exit, but one that is concentrated among larger firms, where the majority of workers are employed. This opens up a channel to the prevention of job loss. Understanding the distinct roles of firm entry and exit prevention in immigrant absorption, and comparing them to intensive margin firm growth responses, is the primary motivation behind Section 1.3.5. Meanwhile, understanding whether the prevention of firm death either prevents productive firm turnover from taking place, or reflects the preservation of more productive firms—which tend to be larger[28]—is the focus of Section 1.4.

That the effect is driven by firm creation and has a larger magnitude in nontradable industries also carries the possibility of new final goods varieties for local consumers, which enhances welfare in models like the one presented in Section 1.5. However, the existence of a large, positive, and precisely estimated effect in tradable industries also validates the interpretation that the effects detected using Equation (1.3.1) are driven by immigrant

[27] 0.5 times the number of workers with "Some College" plus all workers with at least a four-year college degree.

[28] See, e.g., Bernard and Jensen (1999); Leung et al. (2008); Baldwin et al. (2002); van Ark and Monnikhof (1996).

workers on the supply side of the product market.[29] In addition, the existence of a larger effect in industries that primarily utilize lower-educated labor conforms with evidence that lower-educated immigrant and native workers are less substitutable than higher-educated immigrant and native workers (see, e.g., Peri and Sparber, 2009).

1.3.5 Immigrant Absorption

Thus far, this section chronicles a broad, heterogeneous relationship between immigration and net firm entry that stems from both gross firm entry and the prevention of firm exit. Previous literature, meanwhile, has documented the importance of business creation to job creation and overall labor demand (e.g., Decker et al., 2014; Beaudry et al., 2018). In particular, both the entry of small-to-medium sized firms and the prevention of large firm exit open up substantial channels to job creation and the prevention of job loss, respectively. The question of how immigrants are absorbed into labor markets—or, put differently, how the economy creates enough jobs to keep pace with immigrant inflows—may thus hinge on firm entry and firm exit. This sub-section conducts a decomposition exercise to test how important extensive margin firm responses—entry and exit—are to immigrant-induced job creation.

This decomposition is motivated by Dustmann and Glitz (2015), but with a particular focus on the separate margins that go into net firm entry, without restricting the analysis to the tradable sector, and in the context of U.S. local economies. Additionally, I study the effect of immigrant inflows on job creation fully within commuting zone and sector, using Equation (1.3.1) for estimation. Thus, unlike Hong and McLaren (2015), my results are unlikely to stem from immigrant-specific consumer demand. I utilize the following decomposition of employment, enabled by the longitudinal structure of the LBD:

$$\begin{aligned}\Delta\text{Employment}_{gkt} = &- \text{Employment}_{gk,t-10}^{\text{Dead at time } t} + \text{Employment}_{gkt}^{\text{Age}<10} + \Delta\text{Employment}_{gkt}^{\text{Continuers}} \\ &+ \text{Residual}\end{aligned}$$

where the first term represents job loss from firm deaths between $t-10$ and t in commuting zone g and industry group k, the second term represents gross job creation at firms that were born between $t-10$ and t, and the third term represents employment growth at firms that were alive in both $t-10$ and t and in operation in gk. The final term represents employment growth that came from net relocations or expansions into gk from other commuting zones and sectors between $t-10$ and t. As in section 1.3.4.3, each term in this decomposition is divided by the Census-measured workforce in gk at time $t-10$ to retain consistency

[29] Under the assumption that local demand for locally produced tradable goods is not strong.

with independent variable ΔI_{gkt}. Thus, the effect of ΔI_{gkt} on $\Delta\text{Employment}_{gkt}$ measures the number of LBD jobs each immigrant worker generates when they arrive and work in a commuting zone and industry group.

The results of this decomposition are seen in Table 1.7. As a starting point, Column 1 provides estimates of the number of LBD jobs generated by each immigrant. There are two key findings from Column 1. First, and as discussed in Section 1.3.4.2, the OLS estimate in Panel A is substantially larger than the IV estimate in Panel B (more than 40 percent). This accords with the notion that immigrants are attracted to industries and geographies with increasing labor demand and with my claim that $\Delta z_{gkt}^{\text{Emigrants}}$ corrects for some of this endogeneity. Second, the headline estimate from Panel B indicates that each immigrant leads to 0.6 new LBD jobs in the commuting zone and industry group she enters.

Additional context helps interpret this estimate. For a variety of reasons, the number of workers represented in Decennial Census-measured counts like ΔI_{gkt} is likely to be larger than the number of employees represented in LBD-measured counts like $\Delta\text{Employment}_{gkt}$. Most notably, ΔI_{gkt} includes self-employed individuals, but additionally, employees who do not work for establishments covered by the Census Bureau's Business Register. Because the County Business Patterns (CBP) uses the same source data as the LBD, a useful comparison using publicly-available data is one between Decennial Census worker counts from IPUMS-USA and employment from the CBP. Appendix Section A.4.2 shows that each Decennial Census worker is associated with 0.64 CBP jobs, with a coefficient of 0.6 rejected at the 10 percent level. Thus, given that the LBD universe is slightly larger than the CBP (Jarmin and Miranda, 2002), it is reasonable to conclude that Column 1 in Panel B of Table 1.7 indicates near-full, but incomplete, absorption of immigrant workers into a given commuting zone and industry group. There is no evidence for spillover job creation, which is entirely consistent with results in Hong and McLaren (2015), who find that spillover job creation is likely the result of increased immigrant consumer demand. When consumer demand spillovers are controlled for, as they are here, scope for some displacement is in fact to be expected, especially when looking within the fine industry groups used in the analysis and described in Table 1.4. A bevy of previous literature has documented native mobility across occupations in response to immigrant inflows, and Burstein et al. (2017), for example, document substantial native displacement in non-tradable industries.

I next turn to Columns 2–5 in Panel B, which represent the first benchmark result in this chapter. They show clear evidence that localized, extensive margin firm responses play the dominant role in immigrant absorption, with the prevention of firm exit on its own accounting for nearly 50 percent. Overall, point estimates indicate that the extensive margin (including net relocation and expansion) accounts for more than 85 percent of immigrant-induced job creation, leaving only a statistically insignificant 15 percent for the growth of existing firms. Estimates in Columns 2 and 3 accord naturally with results from Figure 1.7. While there are more new, small-to-medium sized firms in the market, death prevention plays an outsized role in immigrant absorption because the firms that are pushed to survive by immigrant worker inflows are large. Nonetheless, that firm creation accounts for around a quarter of immigrant-induced job creation is also consistent with its overall importance to job creation in the U.S. economy.

Comparing Columns 2–5 across Panels A and B also suggests interesting conclusions regarding immigrant locational and industry choices. Immigrants appear attracted to geographies and industries that are experiencing relocations, expansions, within-firm growth, and firm entry. The stronger negative effect on firm death in Panel B relative to Panel A indicates that immigrants may also be attracted to more dynamic local economies, in which productive turnover is taking place.

In total, this sub-section makes important contributions to the literature on immigrant worker absorption into U.S. local economies. Even when comparing across industries within the same commuting zone, local labor markets exhibit substantial absorbtion capacity for immigrant inflows. Strikingly, I cannot reject the null hypothesis that this capacity is entirely driven by firm entry and the prevention of firm exit. Table 1.7 is perhaps the clearest illustration that extensive margin firm responses drive local production's response to immigration. Given the particular importance of death prevention found here, I next turn to understanding the heterogeneous effects immigrants have on firm shut down decisions.

1.4 The Exit Margin

An empirical literature finds that firm turnover can play a critical role in redistributing resources from less to more productive firms, with less productive firms exiting the market (Foster et al., 2008). While Section 1.3.5 demonstrates that the prevention of firm death is overwhelmingly important in absorbing immigrant workers, it could also imply a mechanism through which immigration may stunt productive reallocation—by preventing the exit of low productivity firms. In this section, I test for this mechanism, stratifying a firm-level analysis of survival by correlates of initial firm total factor productivity. Contrary to these concerns, I find that immigrant inflows tend to benefit more productive firms by reducing their shut down probabilities, while culling lower productivity firms from the market. This heterogeneity implies a novel channel from extensive margin firm responses to increased local productivity in response to immigration.

1.4.1 Full Panel: Data and Methods

The analyses in this section utilize two panels of firms from the LBD. The first, termed the "Full Panel," covers all (over 4 million) firms that were in operation in 2000 in the commuting zones and industries covered by the analysis from Section 1.3. It follows these firms in 2005 and 2010, assigning each an indicator variable $\mathbb{1}[\text{Shut Down}]_{ft}$ a value of one if the firm has shut down by year t. Shut down is defined as all establishments of the owning firm having died nationally by time t. The panel is fully balanced, with one observation per firm in each g and k in which it operates for 2000, 2005, and 2010.[30] The restriction to 2000, 2005, and 2010 is necessitated by two constraints: 1) sub-state (e.g., commuting zone), annual estimates of the foriegn-born workforce are only available in 2000 (Decennial Census), and 2005 onwards (ACS); and 2) the instrumental variable $z_{gkt}^{\text{Emigrants}}$ relies on emigrant counts that are available in the IAB Brain Drain dataset only in 2000, 2005, and 2010.

The LBD contains two correlates of productivity, employment and mean labor earnings (payroll divided by employment), which I adjust to generate measures of initial productivity. The adjustment process takes a simple form, in which I log each variable (to ensure a relatively normal distribution of residuals) and remove firm age-by-5-digit NAICS code fixed effects. In the case where productivity is being measured by employment for example, this step is motivated by the fact that productive, young firms may still be small as they settle into the market and by the fact that different detailed industries feature different labor input needs. Firms are then deemed "high productivity," if the residual from this initial regression is positive:

$$\log(\text{Prod. Measure})_{f,2000} = \alpha + \alpha_{k'a} + u_{f,2000} \qquad (1.4.1)$$
$$\mathbb{1}[\text{High Prod.}]_f \equiv \mathbb{1}[\hat{u}_{f,2000} \geq 0]$$

where a indexes firm age, k' indexes a 5-digit NAICS code, and f indexes a firm.

I then estimate the following linear probability models to test how immigrant presence heterogeneously affects firm shut down probabilities within commuting zones and industry

[30] Note that in the "Full Panel," this means the same firm can appear as a "firm" multiple times if it operates in multiple industries or geographies.

groups:

$$\mathbb{1}[\text{Shut Down}]_{ft} = \alpha + \beta^{\text{total}} \left(I_{gkt}\right) + \Gamma X_{gkt} + \alpha_f + \alpha_a + \alpha_{gt} + \alpha_{k't} + \varepsilon_{ft} \tag{1.4.2}$$

$$\mathbb{1}[\text{Shut Down}]_{ft} = \alpha + \beta^{\text{main}} \left(I_{gkt}\right) + \beta^{\text{mod}} \left(I_{gkt} \times \mathbb{1}[\text{High Prod.}]_f\right) + \Gamma X_{gkt} + \alpha_f + \alpha_a + \alpha_{gt} + \alpha_{k't} + \varepsilon_{ft} \tag{1.4.3}$$

with $t \in \{2000, 2005, 2010\}$. I_{gkt} is the immigrant stock of workers in commuting zone g and industry group k at time t, divided by the 2000 worker count in gk. It is instrumented for by z_{gkt}, defined as in Equation (1.3.2), except with a denominator of the 2000 worker count in gk instead of the 1980 worker count in gk.[31,32]

Equation (1.4.2) can simply be thought of as breaking down one decade from Equation (1.3.1) to the firm level—with the crucial inclusions of firm fixed effects α_f and firm age fixed effects α_a.[33] Given our findings in Section 1.3.4.3, we therefore expect $\beta^{\text{total}} < 0$. Estimating Equation (1.4.3) serves the primary purpose for this section, extending our understanding by stratifying the analysis based on initial firm productivity. In particular, β^{main} indicates the extent to which marginal, lower productivity, firms are sustained ($\beta^{\text{main}} < 0$) or pushed out ($\beta^{\text{main}} > 0$) by immigrant worker inflows.

1.4.2 Full Panel: Results

Table 1.8 presents the results of estimating (1.4.2) and (1.4.3) on the "Full Panel." Columns 1-3 first confirm preliminaries and priors: z_{gkt} has a very similar effect on I_{gkt} here as Δz_{gkt} had on ΔI_{gkt} in Section 1.3 (Column 1), and immigrant inflows prevent firm death in this modified specification that hones in on the 2000s (Columns 2 and 3). Once again, extensive margin firm responses are stronger after immigrant exposure is corrected for endogeneity.

Columns 5 and 7 deliver the second benchmark result in this chapter: immigrant inflows into commuting zones and sectors cull lower productivity firms from the market while sustaining and reducing the chances higher productivity firms exit. This is true whether productivity is measured in terms of employment or earnings. Column 5, for example, finds that an average immigrant inflow across a decade (0.04 from Table 1.3) increases the 5-year

[31] The base year is still 1980 for shares $\pi_{og,1980}$.

[32] The primary utility of using emigrants to non-US countries over other exogenous push factors is the use of z_{gkt} in this specification, where it is available at a high (every-five-year) frequency and still strong enough to be a relevant instrument.

[33] For example, α_f subsumes commuting zone-industry group fixed effects α_{gk} and therefore the first difference seen in (1.3.1).

exit probability for low productivity firms by a full percentage point, while reducing the 5-year exit probability for high productivity firms by more than two percentage points.

These results run directly counter the notion that immigrants prevent productive reallocation by sustaining marginal firms. Instead, not only do the most productive firms benefit, but the least productive firms exit—immigrants generate increased productive firm turnover. Section 1.5 posits one potential reason this could take place: some firms can better utilize immigrant labor, but at a cost that makes them positively selected on productivity. Other possibilities include Becker-style discrimination among lower productivity firms and frictions that prevent lower productivity firms from directing search to immigrant workers they may want to hire. Regardless, the empirical results contained in Table 1.8 provide novel evidence that immigrants increase productive reallocation across firms in local U.S. economies. They also provide yet another piece of evidence that immigrants are generating these effects from the supply-side of the product market, under the premise that increased consumer demand alone should sustain rather than cull marginal firms.

1.4.3 SBO Panel: Data and Methods

The remainder of this section employs a second panel that links LBD firms to employee-hiring firms (over 600,000) that were in operation and eunmerated in the 2007 Survey of Business Owners (SBO), following them until 2012—the "SBO Panel." The SBO is another restricted-access Census Bureau survey that allows for two additional advantages at the cost of a shorter and less extensive panel. First, it contains reports of firm revenues at the national level, allowing for additional and more standard measures of firm productivity. These additional productivity measures are generated in the same way as above, using Equation (1.4.1). Second, starting in 2007, the SBO began asking about firm owner nativity. This allows for an enriched understanding of the distributional consequences of immigrant worker inflows for firm owners.

A first set of specifications for this analysis take the form:

$$
\begin{aligned}
\mathbb{1}[\text{Shut Down}]_{ft} &= \alpha + \beta^{\text{total}}\left(I_{gk,t-2}\right) \\
&\quad + \Gamma X_{gk,t-2} + \alpha_f + \alpha_a + \alpha_{gt} + \alpha_{k't} + \varepsilon_{ft} && (1.4.4)\\
\mathbb{1}[\text{Shut Down}]_{ft} &= \alpha + \beta^{\text{main}}\left(I_{gk,t-2}\right) + \beta^{\text{mod}}\left(I_{gk,t-2} \times \mathbb{1}[\text{High Prod.}]_f\right) && (1.4.5)\\
&\quad + \Gamma X_{gk,t-2} + \alpha_f + \alpha_a + \alpha_{gt} + \alpha_{k't} + \varepsilon_{ft}
\end{aligned}
$$

with $t \in \{2007, 2012\}$. The replacement of I_{gkt} with $I_{gk,t-2}$ is necessitated by the availability of $z_{gkt}^{\text{Emigrants}}$ for $t \in \{2005, 2010\}$. Estimating equations (1.4.4) and (1.4.5) serves two purposes: first, they add additional evidence of the productive reallocation result using more conventional productivity measures; and second, they ensure that this modified specification delivers similar results as estimating Equations (1.4.2) and (1.4.3).

This latter reason then validates the use of the following models to further understand the heterogeneity of immigrant-induced firm shut down decisions:

$$\begin{aligned}
\mathbb{1}[\text{Shut Down}]_{ft} = \alpha &+ \beta^{\text{native}} \left(I_{gk,t-2}\right) \\
&+ \beta^{\text{immi}} \left(I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}]_f\right) \\
&+ \Gamma X_{gkt} + \alpha_f + \alpha_a + \alpha_{gt} + \alpha_{k't} + \varepsilon_{ft}
\end{aligned} \tag{1.4.6}$$

$$\begin{aligned}
\mathbb{1}[\text{Shut Down}]_{ft} = \alpha &+ \beta^{\text{native main}} \left(I_{gk,t-2}\right) \\
&+ \beta^{\text{native mod}} \left(I_{gk,t-2} \times \mathbb{1}[\text{High Prod.}]_f\right) \\
&+ \beta^{\text{immi main}} \left(I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}]_f\right) \\
&+ \beta^{\text{immi mod}} \left(I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}]_f \times \mathbb{1}[\text{High Prod.}]_f\right) \\
&+ \Gamma X_{gkt} + \alpha_f + \alpha_a + \alpha_{gt} + \alpha_{k't} + \varepsilon_{ft}
\end{aligned} \tag{1.4.7}$$

Equation (1.4.6) illustrates the differential shut down responses of immigrant- and native-owned firms, on average, an analysis uniquely available by combining the SBO and LBD. Equation (1.4.7) then combines the analyses in Equations (1.4.5) and (1.4.6) by stratifying firm shut down responses on both ownership nativity and productivity.

These specifications test several hypotheses regarding firm ownership and productivity. β^{immi}, for example, is ambiguous in sign: while there may be linkages between immigrant owners and employees that would suggest $\beta^{\text{immi}} > 0$, immigrant firm owners may face more intense competition from other immigrant entrants if immigrant-owned firms are distinct from native-owned firms in terms of employee pool or output within 5-digit NAICS codes. Similar considerations apply to the immigrant-specific coefficients in Equation (1.4.7).

1.4.4 SBO Panel: Results

Column 1 of Table 1.9 shows a still-negative, but imprecise relationship between immigrant presence and firm shut down probability in the SBO Panel. Columns 2–5, however, embody no such ambiguity, and additionally find that revenue-based measures of productivity deliver the same pattern found in Table 1.8: increased immigrant presence sustains higher productivity firms and culls lower productivity firms.

Table 1.10 builds on these results. First, Column 1 indicates that, on average, the entire negative relationship between immigrant worker presence and shut down probability is due to immigrant-owned firms. Given that my specifications flexibly control for commuting zone-wide factors like a larger immigrant customer base, this is highly indicative of linkages between immigrant firm owners and immigrant employees that are advantageous to immigrant-owned firms. Columns 2–5 add a substantial degree of nuance to this result. Among lower productivity firms, those that are owned by immigrants are significantly less likely to shut down in response to increased exposure to immigrant workers than those owned by natives. However, this is a *relative* gain; low-productivity immigrant-owned firms either do not change their shut down behavior or are still more likely to shut down in response to increased exposure to immigrant workers as a whole (adding up coefficients in the first two rows of Panel B). Furthermore, the advantage among high-productivity firms does not appear to be heterogeneous across owner nativity. Thus, among firm owners, a more complete picture emerges: high productivity firm owners are much less likely to exit, regardless of nativity, while native owners of lower-productivity firms are far more likely to exit in response to increased immigrant worker presence. Figure 1.9 summarizes these results, plotting the point estimates of changes in shut down probability for each nativity-productivity pair given a one percent increase in the workforce due to immigration.

1.4.5 Summary and Discussion

The results found in this section regarding the heterogeneous way in which firms' shut down decisions respond to increased immigrant exposure have substantial implications for the overall effect of immigrant workers on local economies. In conjunction with the large increases in firm entry found in Section 1.3, they provide novel evidence that immigrants increase local business dynamism—creative destruction—on the extensive margin. In the context of models that incorporate firm heterogeneity and a non-trivial firm mass, this dynamism has important welfare implications. Next, in Section 1.5, I use one such model to describe how these results can arise in general equilibrium, then show it generates welfare increases from immigration that a canonical model of local production would not.

This increased dynamism is not without distributional consequences, however. Just as the literature on how immigration affects employees has focused on natives at risk of being substituted for in production, it appears that there is a subset of natives who own low productivity firms that are most at risk of having to shut down their business in response to increased immigrant presence. This likely arises because of direct competition with immigrant entrepreneurs and because of advantages immigrant entrepreneurs have in hiring and utilizing immigrant workers that translate to a competitive advantage.

It is also important to note caveats that stem from data limitations that force me to use proxies for firm total factor productivity rather than more direct measures. The measure I employ that has the most in common with the literature is revenues per worker. At best, this is a measure of labor productivity, but at worst simply a measure of the dispersion in use of non-labor inputs by different firms. The key assumption for the validity of this measure, based on my methodology, is that firms within the same 5-digit NAICS code have comparable non-labor input requirements. Even in the case where this is true, some imperfection in the labor market must exist for there to be a direct link between labor and total factor productivity. In models with competitive labor markets—including the one presented below—revenues per worker are independent of total factor productivity. These models, instead, have a direct link between firm size (both employment and revenues) and firm total factor productivity, which is why I prefer size-based measures here. Nonetheless, in practice, firms can differ in size because of total factor productivity or because of idiosyncratic differences in the demand for goods across firms (e.g., due to a foothold in the market). The key assumption for my preferred productivity measures, then, is that 5-digit NAICS code by firm age fixed effects remove enough of the variation in idiosyncratic demand differences across firms such that the remaining variation primarily measures differences in firm total factor productivity.

1.5 Synthesis Through Theory

The previous two sections have delivered key empirical insights into how immigrant workers affect local economies in partial equilibrium (i.e., comparing across sectors within the same commuting zone). This section relies on theory to show how accounting for the kind of firm heterogeneity implied by the partial equilibrium results ultimately impacts general equilibrium analyses of immigration.

1.5.1 Motivation

By definition, models featuring perfect competition are not compatible with firm heterogeneity in the final goods markets, and models featuring both perfect competition and perfect substitutability in the market for foreign-born and domestic labor are not compatible with immigrant-specific effects on firm outcomes. Melding a Melitz (2003) framework with imperfect substitutability in the labor market across immigrant and native workers allows immigrant employees to generate distinct effects on firm outcomes relative to native employees and potentially change the productivity distribution of firms.

However, this combination alone does not capture specific linkages between higher-productivity firms and immigrant employees. Such linkages are both directly implied by the results in Section 1.4 and indirectly by previous work. Mitaritonna et al. (2017), for example, find that manufacturing firms in France employing immigrant workers are more than 11 percent more productive than their counterparts that do not. Kerr and Kerr (2018) find that immigrant-owned firms in the U.S. feature higher sales per employee than native-owned firms in the same state and two-digit sector. This result only gets stronger with the inclusion of finer sector definitions, gender, ethnicity, age, and education of firm owners. Under the premise that immigrant-owned firms are more likely to hire immigrant employees, their results also imply that firms with more immigrant employees are more productive.

Linkages between immigrant employees and higher-productivity firms can arise if firms have to pay higher fixed operational costs in order to access immigrant employee networks and retain immigrant workers for production. These costs can include (but are hardly limited to) hiring translators and liaisons to be able to enter into immigrant job search networks and direct search for immigrants[34], hiring lawyers to work on visa issues, and paying enforcement costs (in expectation) when hiring undocumented immigrants. In the presence of such costs, firms with direct ties to immigrants will be positively selected on productivity because they must be able to afford them.

In order to capture productivity-related reallocative responses to immigration, I thus add an additional heterogeneity to my theoretical framework. In the spirit of Bustos (2011), I introduce two technology types—one of which is more suited to utilize immigrant labor. As motivated in the discussion above, firms must pay an additional fixed operating cost to access this technology, which means they are positively selected on productivity relative to their counterparts using the less-immigrant-heavy technology. Immigrant-heavy firms save more on labor costs when immigrants enter the labor market, pass through their savings to consumers, and thus gain market share by competing it away from firms who do not have special ties to immigrant workers. Their counterparts are forced to exit, a culling of the lowest-productivity firms in the market as a whole. In sum, this model shows how firm entry and exit can drive immigrant-induced endogenous technological change by changing the composition of firms in the economy.

1.5.2 Setup

Individuals are consumer-employees of type $i \in \{I_e, N_e\}$, with I representing foreign-born individuals and N representing native-born individuals and $e \in \{L, S\}$, where L stands for high school degree or *L*ess and S stands for at least *S*ome college. The mass of each labor type in the economy is fixed and employees supply their labor inelastically—the primary comparative static will increase immigrant mass by increasing both I_S and I_L (as an average

[34]See e.g., this Center for American Progress report about Tyson Fresh Meats and its willingness to hire translators, liaisons, and chaplains in order to utilize immigrant labor.

immigrant inflow into the U.S. does). Entrepreneurs are indexed by j with $j \in \{0, 1\}$. Type 1 entrepreneurs choose to produce with a technology that is more immigrant-dependent.

1.5.2.1 Consumer Preferences

Consumer preferences are uniform across consumers. Preferences across firms are of the form presented in Dixit and Stiglitz (1977):

$$\mathcal{U} = \left[F^{\frac{\eta-1}{\mu}} \sum_j \int_{F_j} Q(f)^{\frac{\mu-1}{\mu}} \right]^{\frac{\mu}{\mu-1}}$$

where F_j is the mass of firms owned by entrepreneurs of type j, $F = \sum_j F_j$, $Q(f)$ is the amount demanded by consumers at firm f, which is a single firm in the local economy. $\mu > 1$ is the elasticity of substitution of consumption across firms. When $\eta = 1$, consumers have a taste for variety, as the original Dixit and Stiglitz (1977) model of monopolistic competition dictates. This taste for variety generates external scale effects through which an increasing market size increases welfare. When $\eta = 0$, we shut down this channel and focus on the firm productivity distribution (see, e.g., Egger and Kreickemeier, 2009).

This results in the following demand curves for each firm, which are downward sloping due to product differentiation and substitutability across goods:

$$Q_j(f) = Y F^{\eta-1} P^{\mu-1} p_j(f)^{-\mu} \tag{1.5.1}$$

where $p_j(f)$ is the price charged by firm f of type j and Y is total consumer spending, and the price index P is given by $P^{1-\mu} \equiv F^{\eta-1} \sum_j \int_{F_j} p_j(f)^{1-\mu}$.

1.5.2.2 Firms

Firms have some market power but are non-strategic and take their downward-sloping demand curves as given—a typical monopolistic competition setup. Firm production functions are given by

$$\begin{aligned}
Q_j(z) &= z q_j(z) \\
q_j(z) &= \left[a \left(L_j(z) \right)^{\frac{\sigma_E-1}{\sigma_E}} + \left(S_j(z) \right)^{\frac{\sigma_E-1}{\sigma_E}} \right]^{\frac{\sigma_E}{\sigma_E-1}} \\
L_j(z) &= \left[b_j \left(I_{Lj}(z) \right)^{\frac{\sigma_I-1}{\sigma_I}} + \left(N_{Lj}(z) \right)^{\frac{\sigma_I-1}{\sigma_I}} \right]^{\frac{\sigma_I}{\sigma_I-1}} \\
S_j(z) &= I_{Sj}(z) + N_{Sj}(z)
\end{aligned}$$

where z is a draw of total factor productivity, $q_j(z)$ is a CES aggregator of less-educated labor (L_j) and more-educated labor (S_j), and the lower education labor group is itself a CES aggregator of immigrant and native labor. I consider high education workers to be indistinguishable across nativity.[35] The elasticities σ_E and σ_I govern how substitutable workers of different education and different nativities are, respectively. z is drawn from the same Pareto distribution regardless of entrepreneur type, with shape parameter ϕ and minimum value m. The parameter a governs relative productivity across less- and more-educated workers.

The key difference across firms of type j is the parameter b_j within the low-education aggregator. I assume type $j = 1$ firms depend more on, and better use immigrant labor:

$$\Delta_b \equiv b_1 - b_0 > 0$$

This assumption stands in for a variety of reasons why firms that are more productive are more adept at using immigrants in production. They may be better at allocating immigrants and natives to different tasks, have better access to search networks where there are immigrant job-seekers, or may be less discriminatory toward (suffer less distaste from hiring) immigrant workers.

The cost function is given by

$$\left(\frac{c_j}{z}\right) Q_j(z) + c_j \kappa_j^f$$

where

$$c_j \equiv \left[a^{\sigma_E} \left(c_{Lj}\right)^{1-\sigma_E} + \left(w_S\right)^{1-\sigma_e}\right]^{\frac{1}{1-\sigma_E}}$$

and

$$c_{Lj} \equiv \left[b_j^{\sigma_I} \left(w_{IL}\right)^{1-\sigma_I} + \left(w_{NL}\right)^{1-\sigma_I}\right]^{\frac{1}{1-\sigma_I}}$$

and w_{ie} represents the wage for a worker nativity i and education e and $w_S \equiv w_{NS} = w_{IS}$ because of perfect substitutability among higher-educated workers. κ_j^f is a fixed operating cost that is allowed to vary by entrepreneur type for reasons mentioned above—immigrant-linked firms often appear to pay additional fixed operational costs in order to utilize immigrant labor. Thus, in order to access the immigrant-specific production boost represented by $b_1 > b_0$, they must pay a proportional cost every period, τ, such that $\kappa_1^f = \tau \kappa_0^f$.

The cost function leads to a familiar pricing rule in models of monopolistic competition and Dixit and Stiglitz (1977) preferences:

[35] There is more robust evidence for imperfect substitutability among low-education workers. See, e.g., Peri and Sparber (2009).

$$p_j(z) = \left(\frac{\mu}{\mu - 1}\right)\left(\frac{c_j}{z}\right) \tag{1.5.2}$$

That is, the firm still charges a constant markup over its marginal cost, but the firm's marginal cost reflects the two different types of labor it aggregates. Firms compete through prices, and so firms that are able to pass on declines in c_j to consumers through p are able to gain in market share.

1.5.3 Equilibrium Definition

Entrepreneurs only stay in the market if they are profitable. This defines a cutoff productivity for type 0 firms:

$$\pi_0(z_0^*) \equiv 0 \tag{1.5.3}$$

A second cutoff exists, at which marginal producers are indifferent between the immigrant-heavy (type 1) and immigrant-light (type 0) production mode:

$$\pi_0(z_1^*) \equiv \pi_1(z_1^*) \tag{1.5.4}$$

Entrepreneurs with productivities below z_0^* exit the market, entrepreneurs with productivities in $[z_0^*, z_1^*]$ produce with technology 0, and entrepreneurs with productivities above z_1^* produce with technology 1. Entrepreneurs do not know their z prior to entry, and must pay an entry cost. The next equilibrium condition is free entry:

$$\mathbb{E}[\pi(z)] = \mathbb{E}[\pi(z)|z > z_0^*]\mathbb{P}[z > z_0^*] = c_0\kappa^e \tag{1.5.5}$$

where κ^e is a sunk (entry) cost entrepreneurs pay to take productivity draws, denominated in units of output. When profits are high enough, entrepreneurs enter until they no longer expect to recover their entry costs.[36] The price level, P is given by

$$P \equiv n_e\left[\int_{z_0{}^*}^{z_1{}^*} p_0(z)^{1-\mu}g(z)dz + \int_{z_1{}^*}^{\infty} p_1(z)^{1-\mu}g(z)dz\right] \tag{1.5.6}$$

where n_e is the endogenous mass of entrepreneurs who take productivity draws. Consumer spending Y is set equal to labor payments, and the final equilibrium conditions occur in the labor market, setting labor supply and labor demand equal for low-education immigrant and

[36]Note that the assumption that entry costs scale with c_0 (instead of c_1 or a combination) is mostly made for analytical convenience. However, a simple, plausible justification is that producers do not invest in the costs to access immigrant labor until after entry activities have been completed and they find out they have a draw of z above z_1^*. Thus, the entry activities are paid for using type 0 technology.

native workers. High-education worker wages are set to be the numeraire, $w_S \equiv 1$.

1.5.4 Equilibrium

The key items of interest revolve around z_0^*. First, I define

$$R_z \equiv \frac{z_1^*}{z_0^*} = \left[(c_0)^{\mu} \left(\frac{c_1\tau - c_0}{(c_1)^{1-\mu} - (c_0)^{1-\mu}} \right) \right]^{\frac{1}{\mu-1}} \tag{1.5.7}$$

and

$$\theta \equiv 1 + R_z^{-\phi} \left(\frac{c_1\tau - c_0}{c_0} \right) \tag{1.5.8}$$

Solving Equations (1.5.3) through (1.5.6) then yield

$$z_0^* = m \left[\left(\frac{\kappa_0^f}{\kappa^e} \right) \left(\frac{\mu - 1}{\phi - (\mu - 1)} \right) \theta \right]^{\frac{1}{\phi}} \tag{1.5.9}$$

$$F = Y \left(\frac{1}{c_0 \kappa_0^f} \right) \left(\frac{\phi - (\mu - 1)}{\phi\mu} \right) \tag{1.5.10}$$

The key variable in this solution is θ, which sets Equation 1.5.9 apart from standard productivity cutoff expressions derived in similar models (e.g., Melitz, 2003). It introduces the notion that entry and exit decisions for marginal type 0 firms depend on type 1 firms, through their ability to steal away market share when their costs go down. If c_1 goes down by more than c_0 in response to a shock, θ rises, which causes z_0^* to rise as well. The rise in z_0^* forces marginal type 0 firms to exit the market.

This mechanism drives the results below because of how it relates to the labor market. Section A.5.2 derives equilibrium in the labor market, with the upshot that demand curves slope downward. Thus, when a low-education tilted inflow of immigrants occurs, the wages of less-educated immigrants deteriorate the most of any group. In turn, c_1 falls by more than c_0 because $b_1 > b_0$.

1.5.4.1 Value Added of the Model: P

Price index P is inversely proportional to welfare. With z_0^* in hand, we can show[37]

$$P^{1-\mu} = [\text{Const.}](c_0)^{1-\mu} F^{\eta} (z_0^*)^{\mu-1} \theta$$

[37] $\text{Const.}] = \left(\frac{\mu}{\mu-1}\right)^{1-\mu} \left(\frac{\phi}{\phi-(\mu-1)}\right)$

where F stands for firm mass. We then have

$$-\frac{d\log(P)}{dI} = \underbrace{-\frac{d\log(c_0)}{dI}}_{\text{Analogous to rep. firm model}} + \underbrace{\left(\frac{\eta}{\mu-1}\right)\frac{d\log(F)}{dI}}_{\text{Increased variety through more firms}} + \underbrace{\frac{d\log(z_0^*)}{dI}}_{\text{Culling of marginal firms}} + \underbrace{\left(\frac{1}{\mu-1}\right)\frac{d\log(\theta)}{dI}}_{\text{Technology switching}} \tag{1.5.11}$$

This expression clarifies the value added of this modeling framework. First, in a canonical, representative firm model of production in which all firms have access to type 0 technology, the welfare impact of immigrants on native workers through prices would be defined only by the first term of Equation (1.5.11). This expression contains three additional avenues that this canonical framework misses.

The next two terms are driven by extensive margin changes: firm mass and the productivity level at the shut down margin. Sections 1.3 and 1.4 each delivered evidence that these reduced form parameters are positive in sign in partial equilibrium. Setting $\Delta_b = 0$, τ =1, and $c_0 = c_1$ implies a model that simply melds monopolistic competition in the output market with imperfect substitutability in the labor market. In such a model, Equations (1.5.7) through (1.5.9) show that $\theta = 1$ and that z_0^* is therefore constant. Thus, such a model would open up an avenue from firm mass to welfare through product variety—as long as consumers demand variety ($\eta = 1$)—but would not affect the firm productivity distribution. A model with $\Delta_b > 0$ and $\tau > 1$ opens the door to additional increases in welfare through a rising θ—which increases welfare both independently and through z_0^*. Simulation results below separate these channels.

1.5.5 Simulations

1.5.5.1 Calibration

Table 1.11 shows key model calibrations, set to match the U.S. economy in 2000, which had an immigrant share of 0.12 (55 percent of which has at most a high school degree) and 0.07 establishments per employee. Two calibrations are particularly difficult without employer-employee linked data. The first is the difference between b_1 and b_0, Δ_b. I will show results that vary this difference in order to demonstrate how big it has to be for θ to rise and for there to be productive reallocation across firms. For a given Δ_b, b_0 is pinned down by the immigrant-native wage gap among low-education workers.

The second difficult parameter is τ, which controls the cost firms pay to obtain Δ_b, and determines how selected on productivity firms that pay it end up being. I also vary

τ in simulations rather than set an arbitrary calibration, but this process revealed that $\tau < 1.5$ leads to unstable simulations and $\tau \geq 1.6$ leads to simulations that are essentially indistinguishable, when all other parameters are held fixed. I thus set τ to 2 for the figures presented below (see Table 1.12).

Under these calibrations, I run an experiment that increases the immigrant stock in the workforce in a specific way that matches the empirical content of the immigrant inflows in this chapter. Each additional immigrant represents 0.65 additional immigrants with no more than a high school degree and 0.35 immigrants with at least some college (see Table A.4). As with a general immigrant inflow to the U.S. over the last 30 years, this experiment tilts the labor force towards lower educational attainment workers.

1.5.5.2 Results

Figure 1.10 shows the results from an example simulation with $\Delta_b = 0.3$ and $\tau = 2$. The top left figure shows that labor costs fall more for type 1 firms. This mechanism filters through to the rest of the results: because type 1 firms see lower marginal costs, the cutoff for switching to type 1 technology moves down. The same is not true for type 0 firms, even through immigrant entry does lower their production costs. This is because type 1 firms are able to price compete away market share, leading to a higher productivity bar for type 0 firms to be able to stay in the market. The result is more entry overall (of both types, but particularly of type 1 firms), exit by marginal type 0 firms, and an increase in native welfare. An immigration shock equivalent to a one percent increase in the population generates a 0.49 percent increase in native welfare. We can then use (1.5.11) to decompose this effect results into its component parts. Specifically,

$$\begin{aligned}
\mathcal{W}_I &\equiv \frac{d\log(\text{Real Native Income})}{dI} \\
&= \frac{d\log(w_{NU}N_U + w_S N_S)}{dI} - \frac{d\log(P)}{dI} \\
&= \underbrace{\frac{d\log(w_{NU}N_U + w_S N_S)}{dI} - \frac{d\log(c_0)}{dI}}_{\text{Standard}} + \underbrace{\left(\frac{\eta}{\mu-1}\right)\frac{d\log(F)}{dI}}_{\text{Variety}} + \underbrace{\frac{d\log(z_0^*)}{dI}}_{\text{Culling}} + \underbrace{\left(\frac{1}{\mu-1}\right)\frac{d\log(\theta)}{dI}}_{\text{Switching}}
\end{aligned}$$

where N_L and N_S are the fixed stock of native workers with at most a high school degree and more than a high school degree, respectively. This expression simply adds changes to native nominal income, $(w_{NL}N_L + w_S N_S)$, to Equation (1.5.11).

The results of model simulations over a range of Δ_b are captured in Figure 1.11.[38] Three important findings emerge: first, the "standard" portions of the immigrants surplus that accrue to natives through wage and price changes, $\frac{d\log(w_{NL}N_L+w_S N_S)}{dI} - \frac{d\log(c_0)}{dI}$ (in blue), are a

[38] Note that when $\Delta_b = 0$, τ is set to 1. Otherwise, $\tau = 2$ as usual.

relatively small component of the immigrant surplus when we account for firm heterogeneity—less than 20 percent in the simulations conducted here. Gains from variety through increased firm presence, $\left(\frac{1}{\mu-1}\right)\frac{d\log(F)}{dI}$ (in gray), are the largest component of the immigrant surplus—insofar as consumers value variety. See Table 1.13 for more details on the decomposition.

However, productive reallocation across firms, $\frac{d\log(z_0^*)}{dI}+\left(\frac{1}{\mu-1}\right)\frac{d\log(\theta)}{dI}$ (in red), particularly through the culling of marginal firms, $\frac{d\log(z_0^*)}{dI}$ (in darker red), plays a significant role as well—at least 30 percent of the immigrant surplus comes from productive reallocation. Returning to a primary motivation in this chapter, these general equilibrium reallocation effects stem from the supply side of the product market—they occur due to immigrant characteristics as employees (making labor costs cheaper) and because they induce and increase entrepreneur mass n_e, which ultimately raises z_0^* through increased competition. Even if we are to believe that gains through variety are over-stated in models of monopolistic competition, taking firm heterogeneity into account is critical to understanding how immigration affects the economy. Specifically, the bifurcation in productivity that is generated by allowing firms to pay a cost to better-utilize immigrant workers more than doubles the native welfare gain that is associated with immigration.

1.5.6 Discussion

These exercises, placed in the context of the empirical results presented above, demonstrate why accounting for heterogeneous firm responses can lead to estimates of immigrant-generated welfare that are substantially larger than those that come from canonical models of labor demand. By allowing for dispersion of productivity across firms, models with the Melitz (2003) structure open a channel to welfare through the productivity distribution. In particular, a rising productivity cutoff increases welfare by lowering prices—a *first-order* welfare gain analogous to shifting out the production possibilities frontier. Yet another first-order welfare gain—not modeled here—could arise if competition among firms raised the elasticity of substitution across products (see, e.g., Blanchard and Giavazzi, 2003). Welfare gains that arise from changes to wages and firm input costs, by contrast, are small because the labor market is characterized by perfect competition. This is also the case with accounting exercises conducted when both the labor market and the product market are characterized by perfect competition (see, e.g., Borjas, 1999 and Ottaviano and Peri, 2008).

1.6 Conclusion

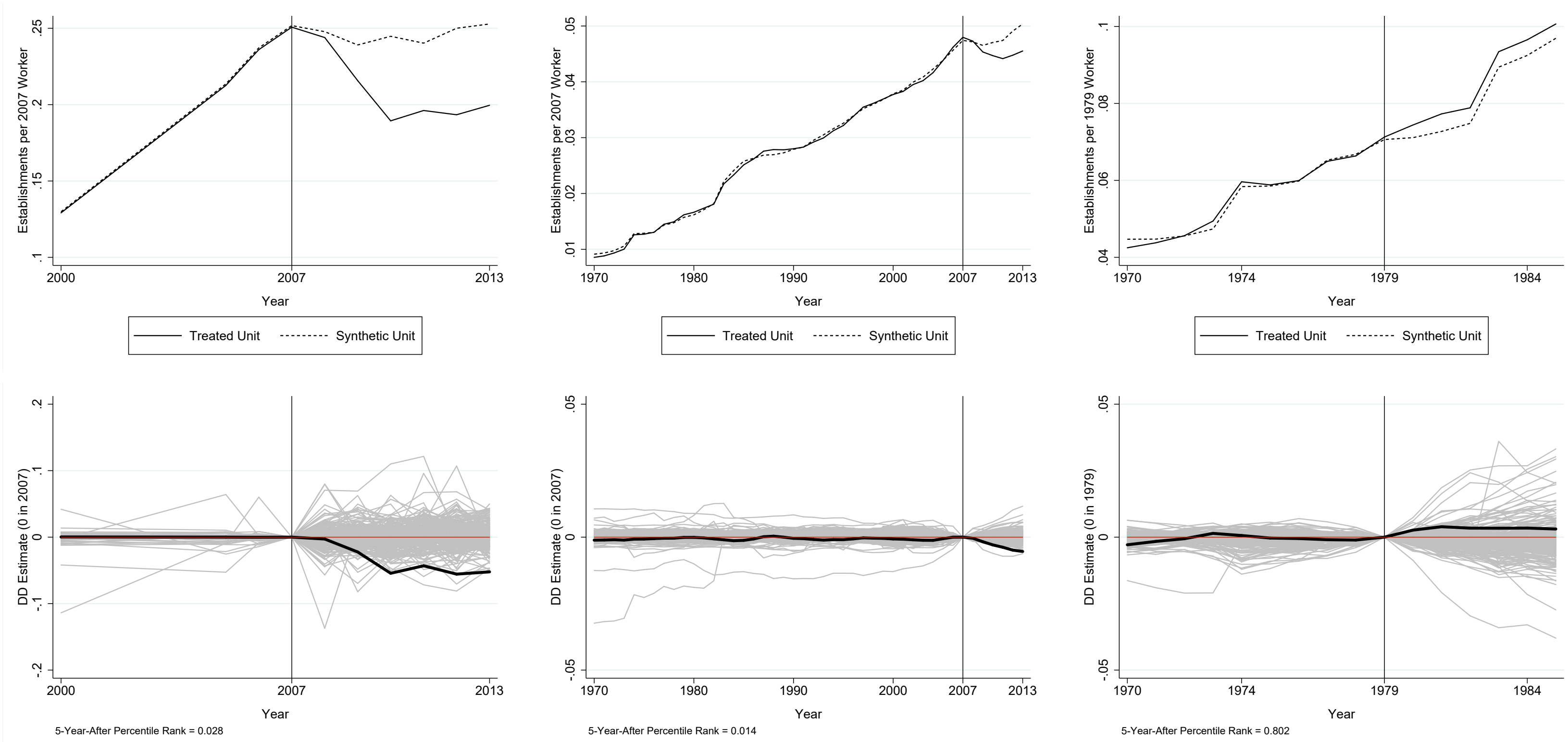
This chapter documents several empirical relationships that reveal the critical role firm entry and exit play in absorbing immigrants into and mediating the effects immigrants have on local economies. A plausibly causal positive relationship between immigrants and business presence drives small-to-medium sized firm creation and prevents exit by older, large firms within commuting zone-industry pairings over three decades in which immigration was a defining feature of demographic change in the U.S. These extensive margin firm responses were responsible for the majority of job creation that absorbed immigrants into local labor markets. Contrary to what we would expect if these effects were driven solely by consumer demand or uniformly lower labor costs across firms, immigrant inflows appear to cull lower productivity firms from a local economy, increasing average local productivity but generating disparate consequences for firm owners of different nativity.

At its broadest point of view, this chapter considers the importance of placing immigration in the context of firm heterogeneity. This requires us to move away from models of perfect competition and towards models that feature market failures—such as the model of monopolistic competition presented in Section 1.5. In these models, increased competition among firms move economies closer to a first-best world, eliminating dead-weight loss in the process. When viewed in this context, the empirical results found in this chapter generate direct channels to first-order welfare gains. They also speak to a burgeoning literature on non-wage immigrant absorption, implying that 1) within-industry, immigrant-absorbing endogenous technology adoption as an immigrant absorption mechanism and 2) connections between immigration and total factor productivity may be tightly linked to firm entry and turnover in the U.S.

Future work will seek to delve into and contextualize these links. Employer-employee linked data, for example, can help test the model's assertion that more productive firms within U.S. industries tend to both hire more immigrant workers and benefit more from immigrant inflows. Additionally, cross-geographic comparisons that stratify immigrant absorption outcomes—including wage changes, native displacement, and productivity—by variables that reflect the ease of starting and shutting down a business can help validate this chapter's suggestion that flexibility on the extensive margin is a key determinant of the relative success in U.S. immigrant labor market assimilation.

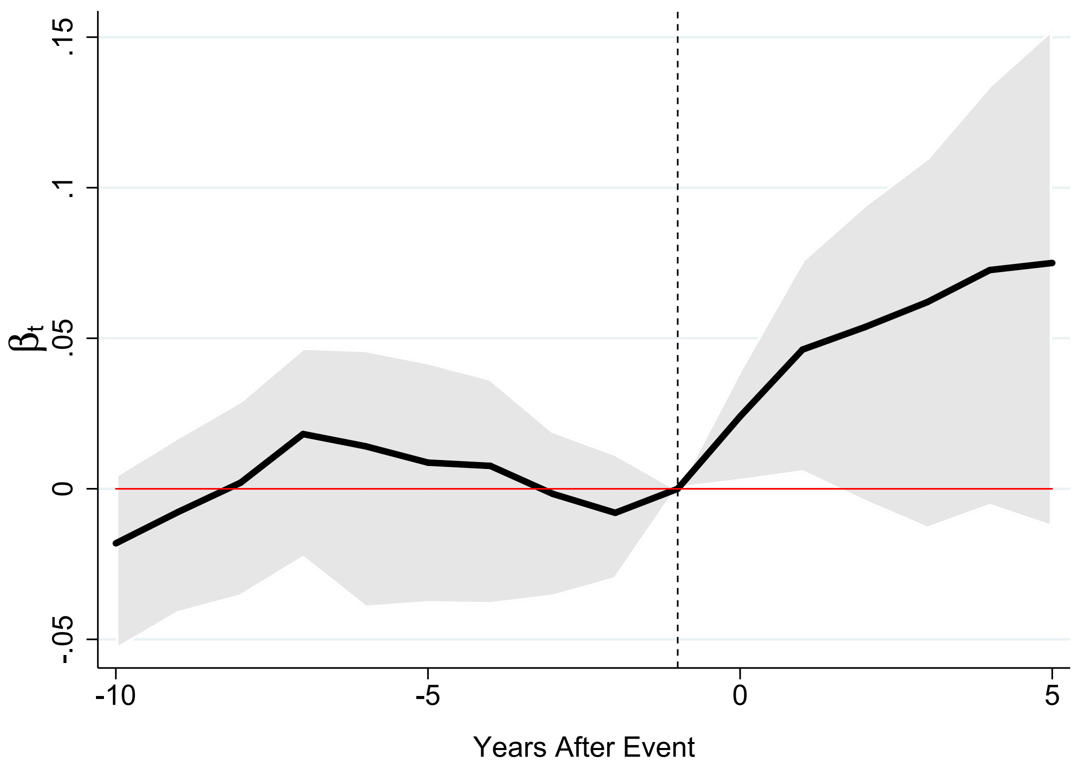
Figure 1.1: Individual Synthetic Control Results—Measures per Initial ($t^* - 1$) Worker

Phoenix MSA: Immigrants **Phoenix MSA: Establishments** **Miami MSA: Establishments**



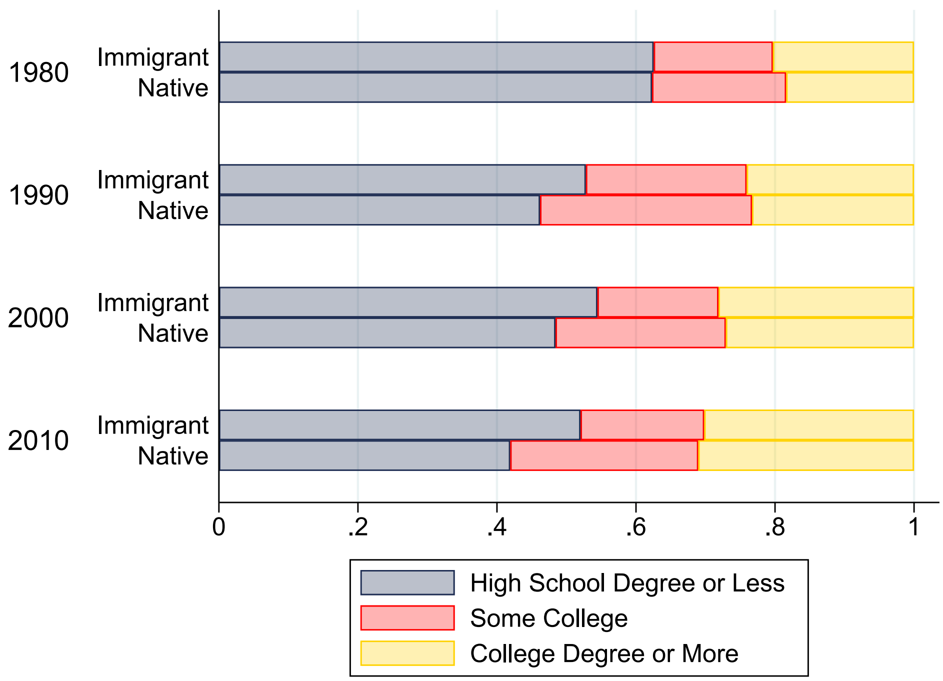
Source: Author's calculations from IPUMS-USA and County Business Patterns.

Figure 1.2: Pooled SCM Results, Establishments per Immigrant (β_t)



Source: Author's calculations from IPUMS-USA and County Business Patterns.
Notes: 95% confidence intervals represented by gray area around point estimates indicated by solid black line. These confidence intervals only apply to the "reduced form"—"first stage" only used as a scaling factor. Combined estimates reflect the cases of Phoenix-Mesa-Scottsdsale, AZ with 2008 as year 0 and Miami, FL with 1980 as year 0.

Figure 1.3: Educational Attainment of U.S. Workers—Shares by Nativity

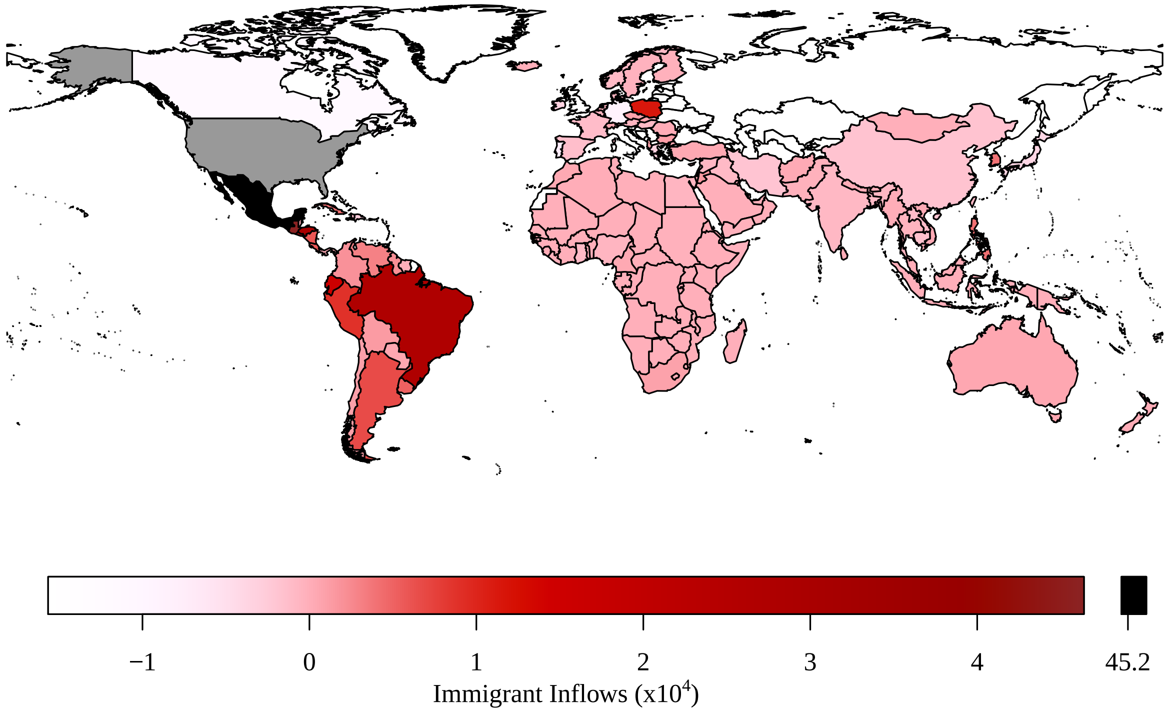


Notes: Data obtained from IPUMS-USA.

Figure 1.4: Endogenous Immigrant Inflows into the Construction Industry, 2000–2005

Panel A: Net Immigration by Country, k=Construction

$$[I_{ok,2005} - I_{ok,2000}]$$



Panel B: Immigrant Inflows into k=Construction

$$[I_{gk,2005} - I_{gk,2000}]$$

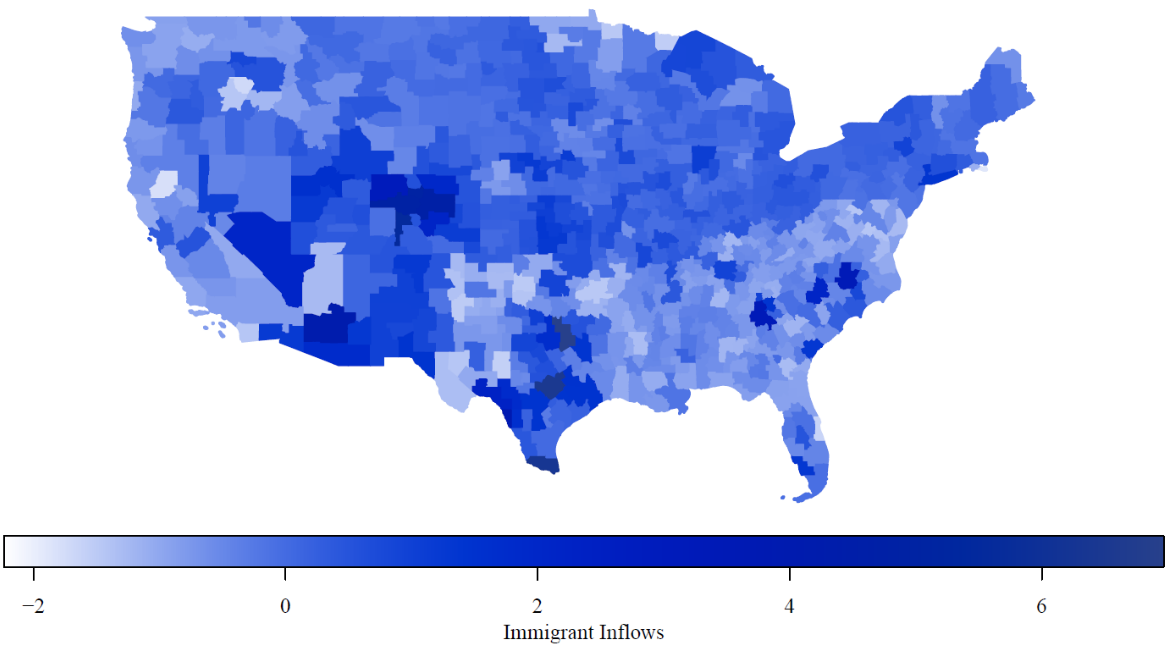
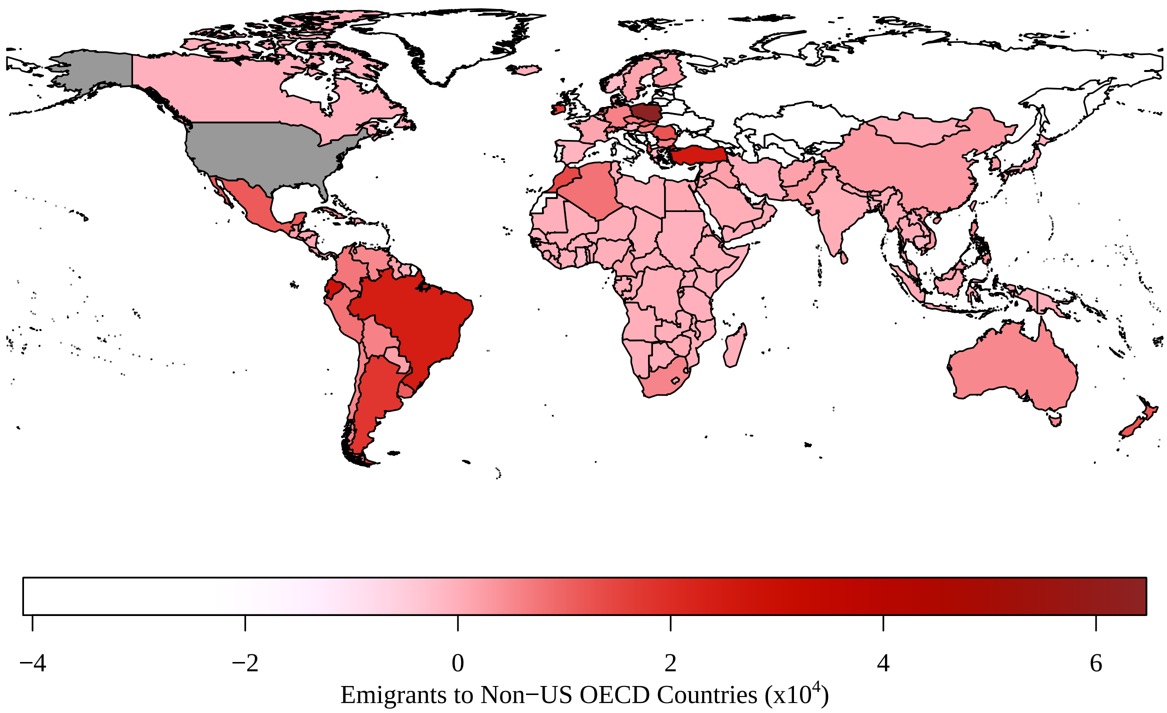


Figure 1.5: Exogenous Immigrant Inflows into the Construction Industry, 2000–2005

Panel A: Emigration Shock, k=Construction

$$\left[(\rho_{ok,2005} \times M^{\text{non-US}}_{o,2005}) - (\rho_{ok,2000} \times M^{\text{non-US}}_{o,2005})\right]$$



Panel B: Instrument-Predicted Inflows into k=Construction

$$\left[z_{gk,2005} - z_{gk,2000}\right]$$

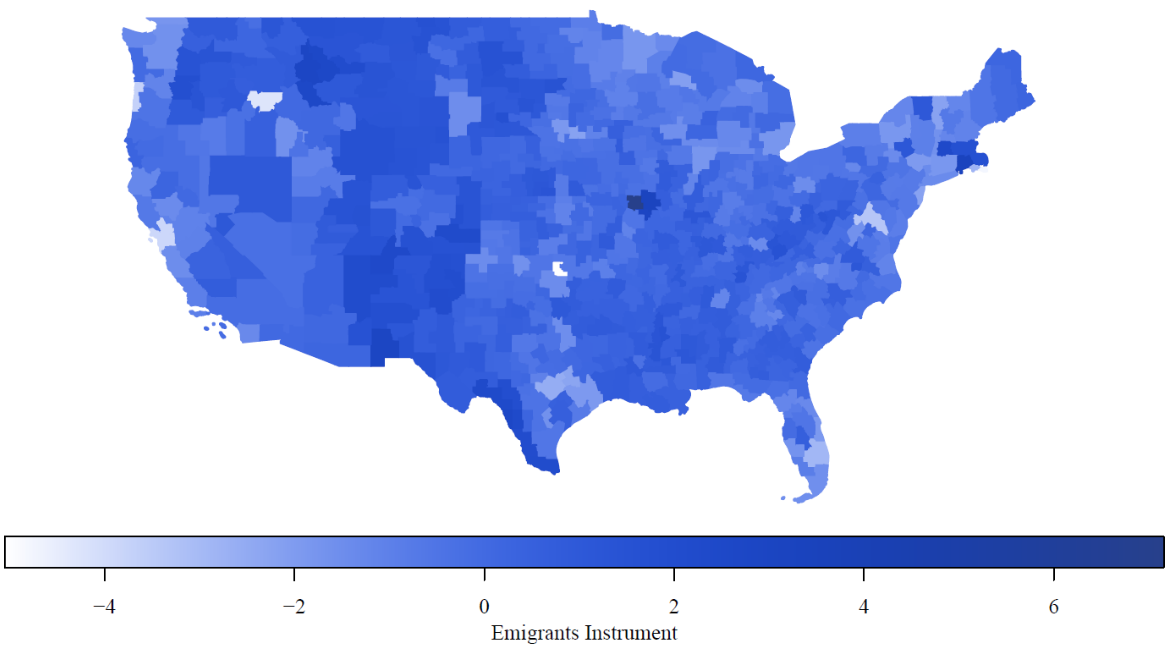
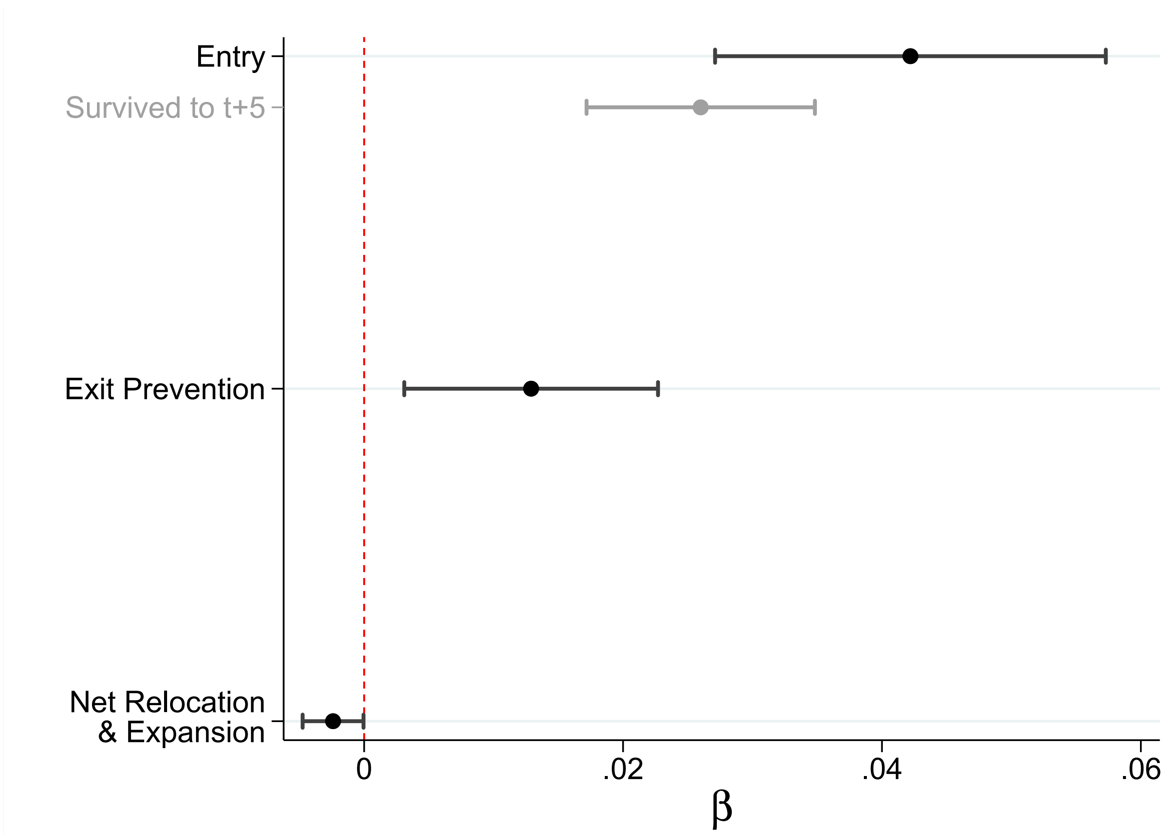
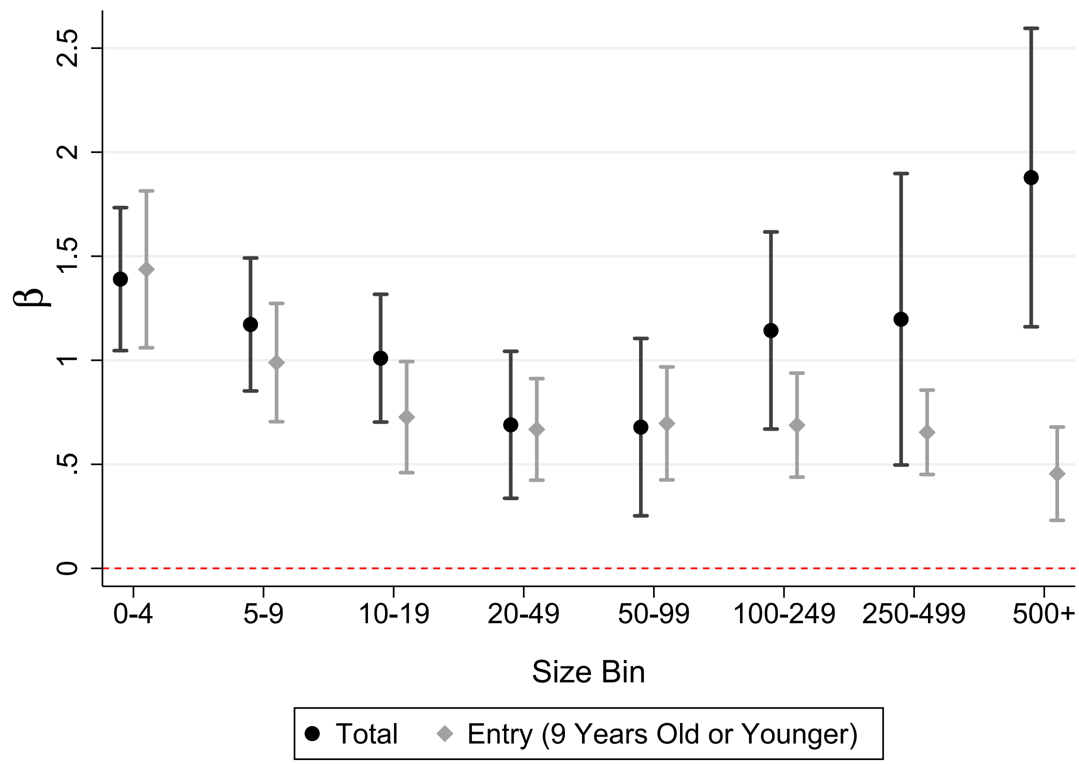


Figure 1.6: The Effect of Immigration on Firm Presence—Flow Decomposition (IV Results)



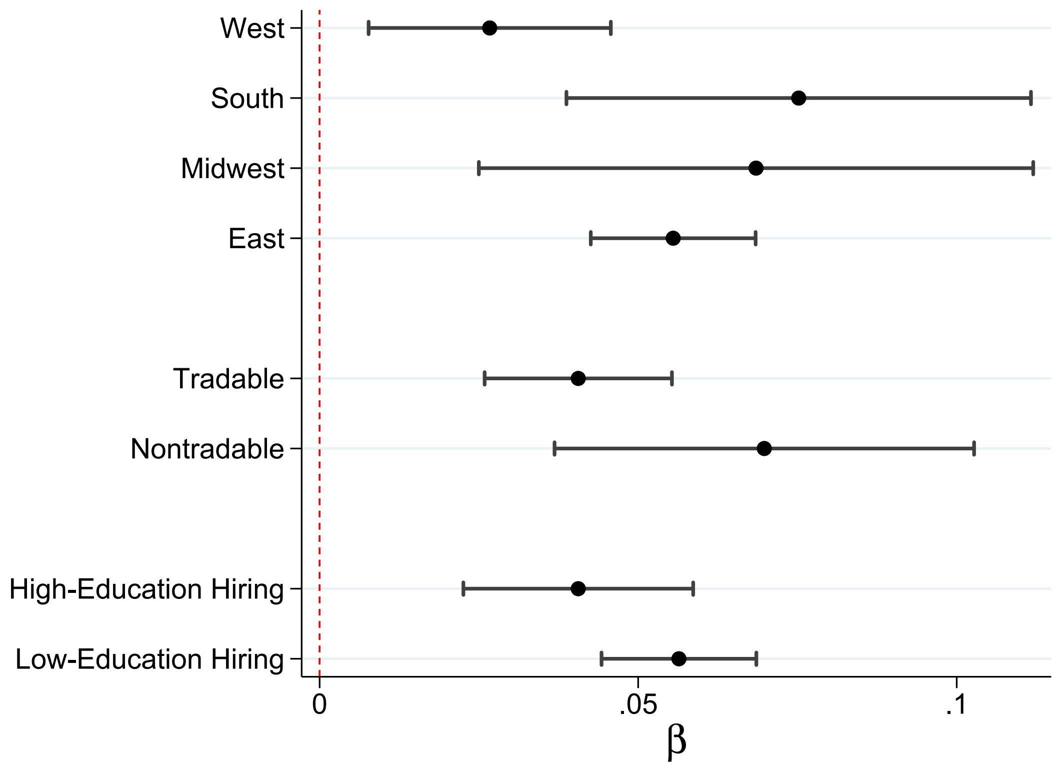
Notes: See Equation (1.3.1) for specification. See Section 1.3.4.3 for details on outcome variables. Data accessed and analyzed in Michigan Census Research Data Center.

Figure 1.7: The Effect of Immigration on Firm Presence—Within Size Bin (IV Results)



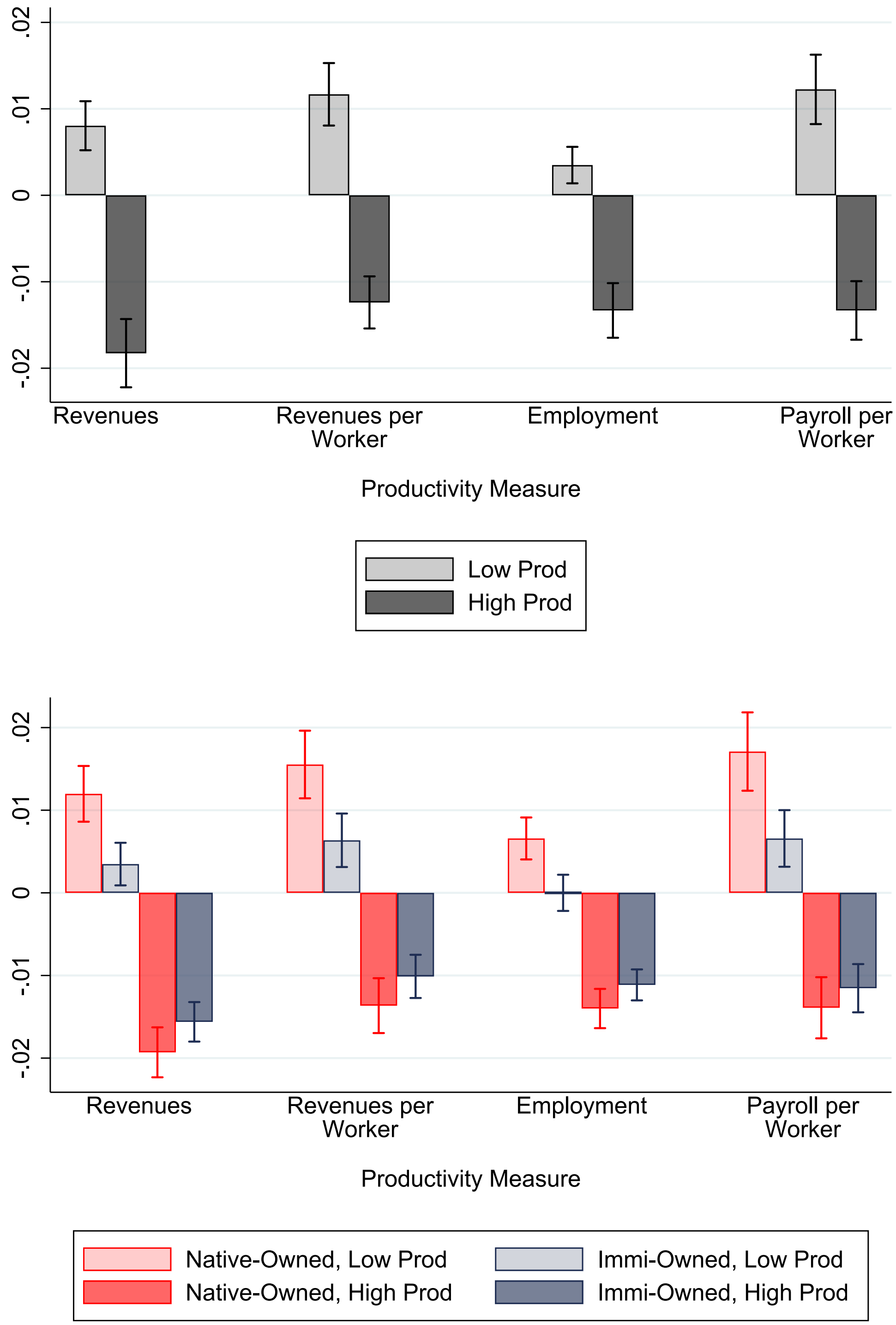
Notes: See Equation (1.3.1) for specification. See Section 1.3.4.4 for details on outcome variables. Data accessed and analyzed in Michigan Census Research Data Center.

Figure 1.8: The Effect of Immigration on Firm Presence—Heterogeneity (IV Results)



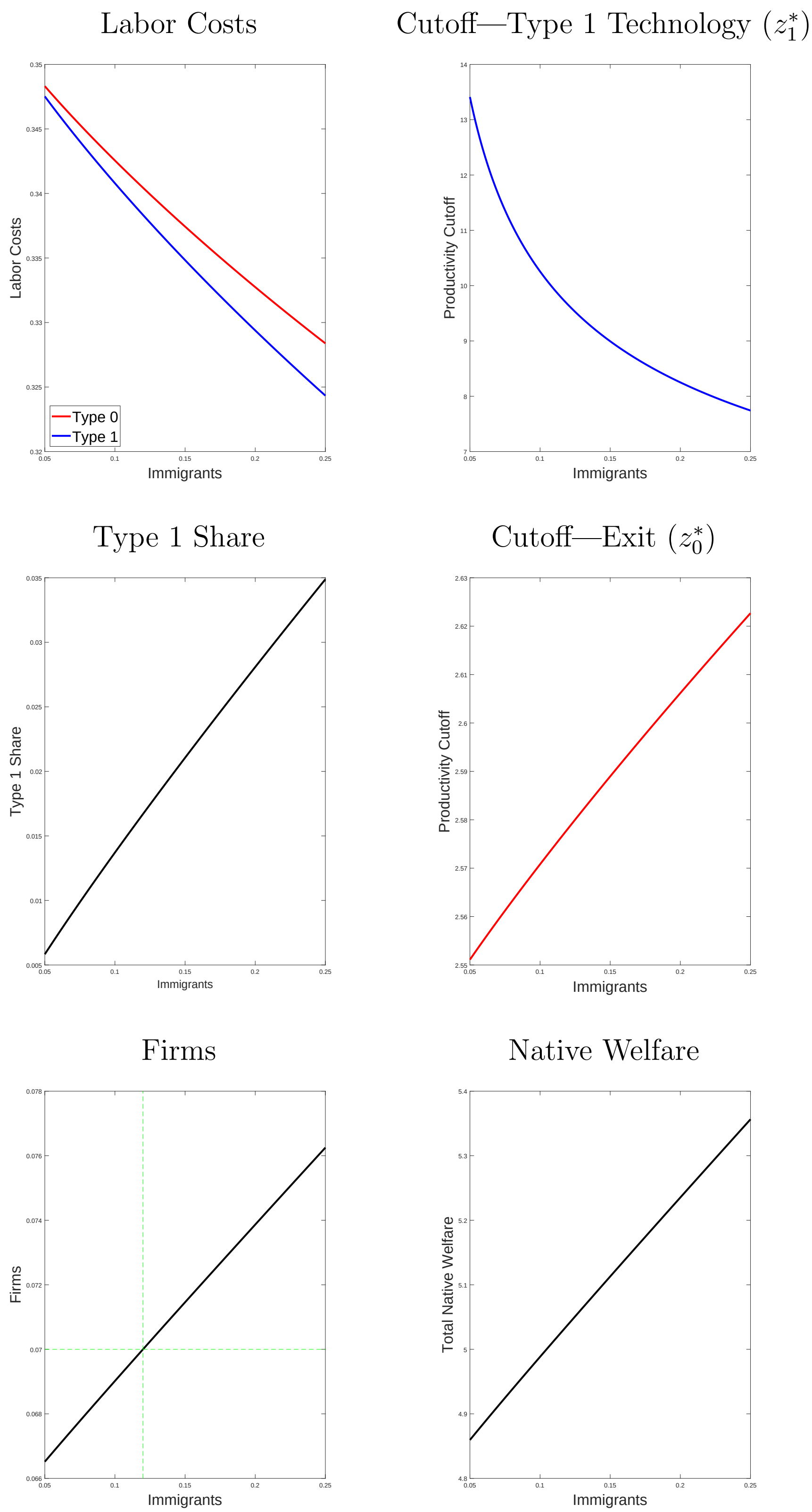
Notes: See Equation (1.3.1) for specification. Data accessed and analyzed in Michigan Census Research Data Center.

Figure 1.9: Change in Shut-Down Probability from a 1% Immigration Shock (IV Results)



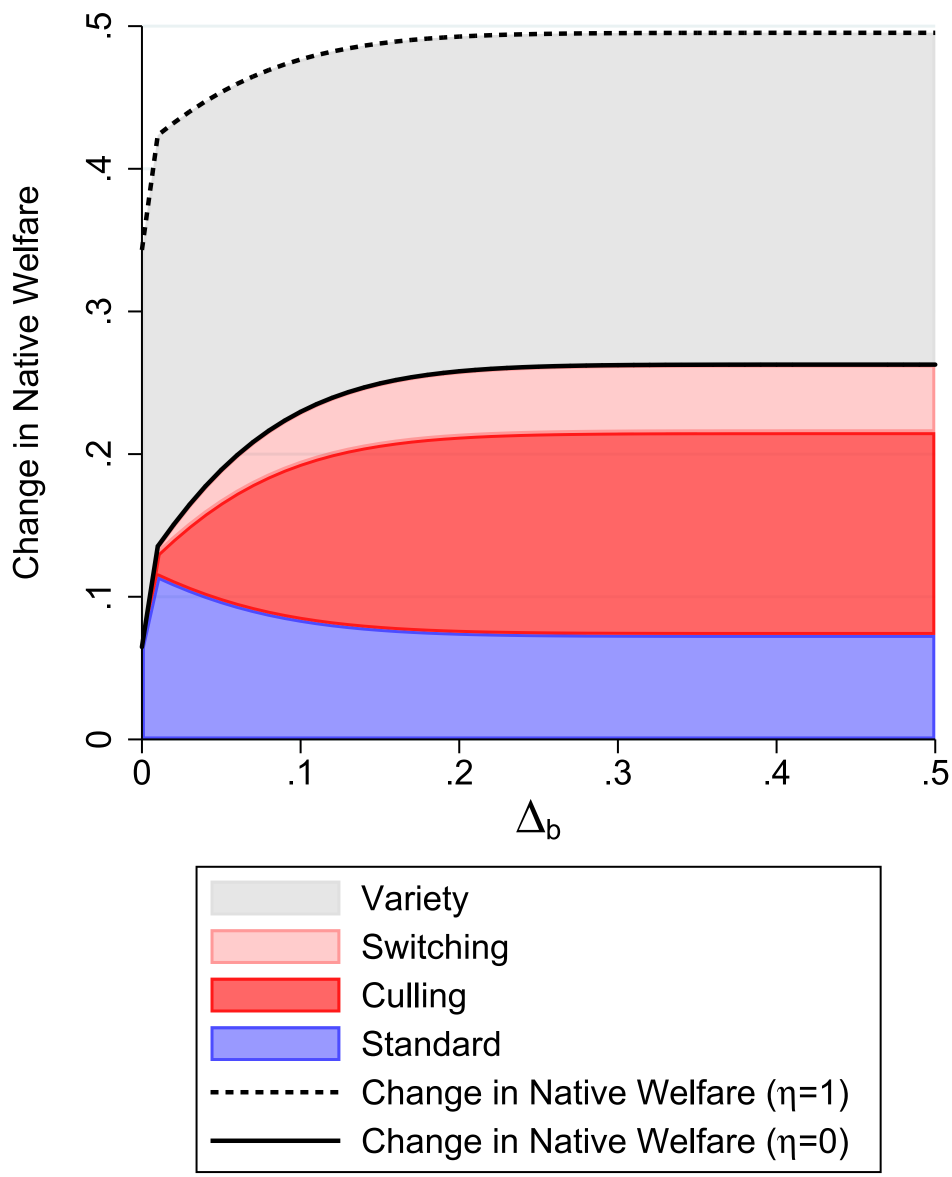
Notes: See Equations (1.4.5) and (1.4.7) for specification and Table 1.10 for underlying coefficients. Data accessed and analyzed in Michigan Census Research Data Center.

Figure 1.10: Example Simulation with $\Delta_b = 0.2$ and $\tau = 2$



Notes: Green dashed lines indicated data moments that were targeted in calibration.

Figure 1.11: Simulations—Percent Change in Native Welfare from a 1% Immigration Shock



Notes: When $\Delta_b = 0$, $\tau = 1$; otherwise, $\tau = 2$.

Table 1.1: "First Stage" Scaling Factors δ^{1S}

Case	Treatment Start (t^*)	Variation Source	Δ_I^{DD}	Empirical p-value
Phoenix-Mesa-Scotsdale, AZ	2008	Arizona LAWA	-0.052	0.056
Miami-Fort Lauderdale-West Palm Beach, FL	1980	Mariel Boatlift	0.066	—

Source: First row—author's calculations from IPUMS-USA and County Business Patterns. Second row—Card (1990) p. 248 and Table 1.

Table 1.2: Educational Content of Immigrant Inflows (Publicly Available Data, k = SIC Sector)

	Outcome: Change in High Education Share of Labor Force			
	(1)	(2)	(3)	(4)
ΔI_{gkt}: Immigrant Inflows per Initial Worker	-0.0913***	-0.0888***	-0.0706***	-0.1020***
	(0.0161)	(0.0303)	(0.0156)	(0.0327)
High Education Definition	H.S. Degree+	H.S. Degree+	College Equiv.	College Equiv.
Instrument	None—OLS	Emigrants	None—OLS	Emigrants
Within R^2	0.016	0.016	0.015	0.012
α_{gt}, α_{kt}	✓	✓	✓	✓
Region × SIC Sector × Year FE	✓	✓	✓	✓
1980 Controls × Year FE	✓	✓	✓	✓
Observations	15,162	15,162	15,162	15,162

Notes: See Equation (1.3.1) for specification. College equivalent refers to 0.5 times workers with Some College plus all workers with a College Degree or More—see, e.g., Doms et al. (2010). All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. Data obtained from IPUMS-USA.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.3: Summary Statistics (Publicly Available Data, k = SIC Sector)

	Workers per Establishment	Immigrants per Worker	Δ Establishments per Initial Worker		ΔI_{gkt}: Immigrant Inflows per Initial Worker	
SIC Sector	Mean	Mean	Mean	Std Dev	Mean	Std Dev
Construction	10.68	0.1227	0.0172	0.0285	0.0926	0.1113
Finance, Insurance, Real Estate	13.30	0.0974	0.0180	0.0186	0.0340	0.0360
Manufacturing	64.25	0.1310	-0.0001	0.0037	0.0176	0.0404
Retail Trade	14.07	0.1194	0.0078	0.0129	0.0528	0.0554
Wholesale Trade	14.04	0.1117	0.0015	0.0217	0.0282	0.0502
Services	15.66	0.1024	0.0177	0.0128	0.0556	0.0499
Transportation & Utilities	28.31	0.0915	0.0034	0.0080	0.0389	0.0475
Total	25.53	0.1132	0.0112	0.0160	0.0474	0.0571

Notes: Data obtained from IPUMS-USA and County Business Patterns. Weighted by 1980 workers in SIC-commuting zone pair.

Table 1.4: Industry Groups

Grouping	1990 Census Codes	2007 NAICS Codes
Construction	60	23
Management of companies	710	55
Utilities	422, 450, 451, 470–472	22, 486, 562
Manufacturing – Food	100, 101, 102, 110 , 111, 112, 120, 121, 122, 130, 610	311–312
Manufacturing – Clothing	132, 140, 142, 150, 151, 152, 220, 221, 222	313–316
Manufacturing – Wood & Furniture +	160–162, 231, 232, 241, 242, 250–252, 261, 262	321, 322, 327, 337
Manufacturing – Plastics +	180–182, 190–192, 200, 201, 210–212	324–326
Manufacturing – Metals & Machinery	270–272, 280–282, 290–292, 300, 301, 310–312, 320, 321, 331, 332, 380	331–333
Manufacturing – Electrical & Household	322, 340–342, 350, 371, 372, 381, 390, 391	334, 335, 339
Manufacturing – Transportation	351, 352, 360–362, 370	336
Wholesale Trade – Durable	500, 501, 502, 510–512, 521, 530–532	423
Wholesale Trade – Nondurable	540–542, 550–552, 560–562	424
Retail Trade – Vehicles	612, 620, 622	441
Retail Trade – Household Durables	580–582, 631–633	442–444
Retail Trade – Food & Gas	601, 602, 611, 621, 650	445, 447
Retail Trade – Misc.	590, 640, 642, 651, 652, 661, 662, 681, 682	446, 451, 453
Retail Trade – Apparel	623, 630, 660	448
Retail Trade – Dept. & Variety Stores	591, 592, 600	452
Retail Trade – Fuel, Catalog, Vending	663, 670–672	454
Transportation	400, 420, 421	481–483
Trucking	410	484, 492
Bus & Taxi	401, 402	485
Warehousing & Storage	411	493
Non-Telephone Communication	440, 852	515, 519
Telecomm & Data Processing	441, 442, 732	517, 518
Savings Institutions	700–702	521, 522
Insurance	711	524
Real Estate	712	531
Professional Services	12, 721, 741, 841, 882, 890–893	541, 711
Admin. & Support Services	20, 432, 722, 731, 740	561
Educational Services	842, 850, 851, 860	611
Health Services excl. Hospitals	812, 820–822, 830, 840	621
Hospitals	831	622
Nursing & Residential Care Facilities	832, 870	623
Social Services	861–863	624
Entertainment Services	742, 800–802, 810, 872	512, 532, 712, 713
Lodging	762, 770	721
Eating & Drinking Places	641	722
Repair Services	750–752, 760, 782, 790	811
Personal Services	771, 772, 780, 781, 791	812
Unions & Religious Organizations	873, 880, 881	813

Table 1.5: The Effect of Immigration on Firm Presence

	Outcome: ΔI_{gkt}	Outcome: $\Delta \text{Firms}_{gkt}$	
	(1)	(2)	(3)
$\Delta z_{gkt}^{\text{Emigrants}}$: Emigrants Instrument	0.2410***		
	(0.0204)		
ΔI_{gkt}: Immigrant Inflows per Initial Worker		0.0322***	0.0527***
		(0.0030)	(0.0055)
Specification Type	1st Stage	OLS	Emigrants IV
Within R^2	0.1211	0.0260	0.0348
α_{gt}, α_{kt}	✓	✓	✓
Region × Industry Group × Year FE	✓	✓	✓
1980 Controls × Year FE	✓	✓	✓
Observations	59,000	59,000	59,000

Notes: See Equation (1.3.1) for specification. Outcome variable $\Delta \text{Firms}_{gkt}$ is divided by initial workforce in gk to retain consistent scaling with ΔI_{gkt}. All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-industry group level. Data accessed and analyzed in Michigan Census Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.6: The Effect of Immigration on Firm Presence—Stability and Robustness of IV Estimates

	$\Delta \text{Firms}_{gkt}$: Net Firm Entry per Initial Worker					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔI_{gkt}: Immigrant Inflows per Initial Worker	0.0835***	0.0448***	0.0481***	0.0527***	0.0464***	0.0569***
	(0.0070)	(0.0057)	(0.0059)	(0.0055)	(0.0059)	(0.0122)
$\Delta I_{gk,t-10}$: Lagged Immigrant Inflows per Initial Worker						-0.0157
						(0.0146)
Instrument(s)	$\Delta z_{gkt}^{\text{Emigrants}}$	$\Delta z_{gkt}^{\text{Emigrants}}$	$\Delta z_{gkt}^{\text{Emigrants}}$	$\Delta z_{gkt}^{\text{Emigrants}}$	$\Delta z_{gkt}^{\text{Emigrants}}$	$\begin{pmatrix} \Delta z_{gkt}^{\text{Emigrants}} \\ \Delta z_{gk,t-10}^{\text{Emigrants}} \end{pmatrix}$
Panel Pre-Trends Test p-value	0.3244	0.7193	0.7620	0.2697	0.3750	0.0820
Kleibergen-Paap Wald 1st Stage F Statistic	261.8	81.80	88.74	140.1	129.7	29.30
Within R^2	0.0482	0.0305	0.0506	0.0348	0.0674	0.0315
α_{gt}, α_{kt}		✓	✓	✓	✓	✓
1980 Controls $\times$ Year FE			✓	✓	✓	✓
Region $\times$ Industry Group $\times$ Year FE				✓	✓	✓
Control for $\frac{\Delta \text{Workers}_t}{\text{Workers}_{t-10}}$					✓	
Observations	59,000	59,000	59,000	59,000	59,000	40,000

Notes: See Equation (1.3.1) for specification. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-industry group level. Data accessed and analyzed in Michigan Census Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.7: Decomposition of Immigrant-Induced Job Creation

			Jobs Created at Firm Type (per Initial Worker):							
	Total			Deaths		Entrants		Continuers		Net Relocations & Expansions
	(1)	=	−	(2)	+	(3)	+	(4)	+	(5)
Panel A: OLS										
ΔI_{gkt}: Immigrant Inflows per Initial Worker	0.8558***			-0.1106***		0.2906***		0.1850***		0.2695***
	(0.0700)			(0.0352)		(0.0347)		(0.0344)		(0.0413)
Percent of Total	100			12.92		34.00		21.62		31.49
Panel B: Emigrants IV										
ΔI_{gkt}: Immigrant Inflows per Initial Worker	0.6014***			-0.2573***		0.1572***		0.0857		0.1012
	(0.1186)			(0.0828)		(0.0575)		(0.0721)		(0.0800)
Percent of Total	100			42.78		26.14		14.25		16.83

Notes: See Equation (1.3.1) for specification. See Section 1.3.5 for details on outcome variables. All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-industry group level. Data accessed and analyzed in Michigan Census Research Data Center.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.8: Immigrant Presence and Firm Exit, Stratified by Productivity (Full Panel)

	Outcome: I_{gkt}	Outcome: $\mathbb{1}[\text{Shut Down}]_{ft}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
z_{gkt}: Emigrants Instrument	0.2732***						
	(0.0265)						
I_{gkt}: Immigrants per Initial Worker		-0.0336***	-0.0709***	0.1911***	0.2604***	0.3327***	0.5273***
		(0.0041)	(0.0269)	(0.0131)	(0.0339)	(0.0213)	(0.0521)
$I_{gkt} \times \mathbb{1}$[High Prod.]				-0.5001***	-0.8494***	-0.7073***	-1.067***
				(0.0315)	(0.0642)	(0.0431)	(0.0797)
Effect for High Prod. Firms					-0.5890***		-0.5393***
					(0.0515)		(0.0526)
Specification Type	1st Stage	OLS	Emigrants IV	OLS	Emigrants IV	OLS	Emigrants IV
Productivity Measure (logged)	—	—	—	Employment	Employment	Payroll per Employee	Payroll per Employee
α_{gt}, α_{kt}	✓	✓	✓	✓	✓	✓	✓
Region × Industry Group × Year FE	✓	✓	✓	✓	✓	✓	✓
2000 Controls × Year FE	✓	✓	✓	✓	✓	✓	✓
Within R^2	0.0616	0.0001	<0.0001	0.003	0.0014	0.0059	0.0044
Observations	15,000,000	15,000,000	15,000,000	15,000,000	15,000,000	15,000,000	15,000,000

Notes: See Equations (1.4.2) and (1.4.3) for specification and (1.4.1) for productivity definitions. All specifications include control variables for 2000 log employment, 2000 establishments per worker, 2000 self employment share, 2000 college share, and 2000 under-40 share in the commuting zone-sector interacted with year fixed effects. Standard errors clustered at the commuting zone-industry group level. Data accessed and analyzed in Michigan Census Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.9: Immigrant Presence and Firm Exit, Stratified by Productivity (SBO Panel)

	Outcome: $\mathbb{1}[\text{Shut Down}]_{ft}$				
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
$I_{gk,t-2}$: Immigrants per Initial Worker	-0.0269*	0.1433***	0.1854***	0.0620**	0.1806***
	(0.0161)	(0.0314)	(0.0373)	(0.0247)	(0.0384)
$I_{gk,t-2} \times \mathbb{1}[\text{High Prod.}]$		-0.5617***	-0.4687***	-0.3214***	-0.5049***
		(0.0861)	(0.0730)	(0.0641)	(0.0826)
Panel B: Emigrants IV					
$I_{gk,t-2}$: Immigrants per Initial Worker	-0.1080	0.8044***	1.168***	0.3494***	1.225***
	(0.0763)	(0.1446)	(0.1843)	(0.1077)	(0.2048)
$I_{gk,t-2} \times \mathbb{1}[\text{High Prod.}]$		-2.630***	-2.407***	-1.681***	-2.557***
		(0.2876)	(0.2882)	(0.2023)	(0.3289)
Productivity Measure (logged)	—	Employment	Payroll per Employee	Revenues	Revenues per Employee
α_{gt}, α_{kt}	✓	✓	✓	✓	✓
Region × Industry Group × Year FE	✓	✓	✓	✓	✓
2000 Controls × Year FE	✓	✓	✓	✓	✓
Observations	611,000	611,000	611,000	611,000	611,000

Notes: See Equations (1.4.4) and (1.4.5) for specification and (1.4.1) for productivity definitions. All specifications include control variables for 2000 log employment, 2000 establishments per worker, 2000 self employment share, 2000 college share, and 2000 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by SBO survey weights. Standard errors clustered at the commuting zone-industry group level. Data accessed and analyzed in Michigan Census Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.10: Immigrant Presence and Firm Exit—The Role of Immigrant Entrepreneurship

	Outcome: $\mathbb{1}[\text{Shut Down}]_{ft}$				
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
$I_{gk,t-2}$: Immigrants per Initial Worker	0.0097	0.1816***	0.2130***	0.1108***	0.2136***
	(0.0175)	(0.0326)	(0.0351)	(0.0271)	(0.0406)
$I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}]$	-0.1422***	-0.1456***	-0.1254*	-0.1753***	-0.1277**
	(0.0358)	(0.0479)	(0.0642)	(0.0443)	(0.0509)
$I_{gk,t-2} \times \mathbb{1}[\text{High Prod.}]$		-0.5530***	-0.4821***	-0.3424***	-0.4900***
		(0.0831)	(0.0675)	(0.0651)	(0.0857)
$I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}] \times \mathbb{1}[\text{High Prod.}]$		-0.0470	0.0737	0.0377	-0.0611
		(0.1085)	(0.1029)	(0.0906)	(0.0802)
Panel B: Emigrants IV					
$I_{gk,t-2}$: Immigrants per Initial Worker	0.0223	1.198***	1.553***	0.6582***	1.710***
	(0.0846)	(0.1721)	(0.2088)	(0.1296)	(0.2419)
$I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}]$	-0.2790***	-0.8495***	-0.9173***	-0.6581***	-1.052***
	(0.0908)	(0.1356)	(0.1469)	(0.1255)	(0.1776)
$I_{gk,t-2} \times \mathbb{1}[\text{High Prod.}]$		-3.128***	-2.919***	-2.059***	-3.101***
		(0.3426)	(0.3280)	(0.2585)	(0.3781)
$I_{gk,t-2} \times \mathbb{1}[\text{Immi Owned}] \times \mathbb{1}[\text{High Prod.}]$		1.219***	1.272***	0.9445***	1.289***
		(0.2347)	(0.2296)	(0.2330)	(0.2590)
Productivity Measure (logged)	—	Employment	Payroll per Employee	Revenues	Revenues per Employee
α_{gt}, α_{kt}	✓	✓	✓	✓	✓
Region × Industry Group × Year FE	✓	✓	✓	✓	✓
2000 Controls × Year FE	✓	✓	✓	✓	✓
Observations	611,000	611,000	611,000	611,000	611,000

Notes: See Equations (1.4.6) and (1.4.7) for specification and (1.4.1) for productivity definitions. All specifications include control variables for 2000 log employment, 2000 establishments per worker, 2000 self employment share, 2000 college share, and 2000 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by SBO survey weights. Standard errors clustered at the commuting zone-industry group level. Data accessed and analyzed in Michigan Census Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table 1.11: Key Calibrations

Parameter	Value	Target Moments	Source
Panel A: Individually Calibrated			
σ_E	1.5	—	Ottaviano and Peri (2012)
σ_I	10	—	Ottaviano and Peri (2012)
μ	4	Average U.S. Markup = 32%	Christopoulou and Vermeulen (2012)
ϕ	3.1	$\phi > \mu - 1$	—
κ^e	3	—	—
m	1	—	—
Panel B: Jointly Calibrated			
a	0.64	$\frac{w_{NU}}{w_S} = 0.52$	2000 Census
b_0	$[0.15, 0.55]$	$\frac{w_{IU}}{w_S} = 0.4$	2000 Census
κ_0^f	0.25	$F = 0.05$	2000 Business Dynamics Statistics

Table 1.12: Percent Change in Native Welfare for a 1% Increase in Workforce Due to Immigration

		Δ_b				
		0.1	0.2	0.3	0.4	0.5
	1.6	0.479	0.495	0.497	0.498	0.498
	1.7	0.479	0.495	0.497	0.498	0.498
τ	1.8	0.479	0.495	0.497	0.498	0.498
	1.9	0.479	0.495	0.498	0.498	0.498
	2	0.479	0.495	0.498	0.498	0.498

Table 1.13: Components of Immigrant Surplus

	Δ_b				
	0.1	0.2	0.3	0.4	0.5
Standard: $\frac{d\log(w_{NU}N_U + w_S N_S)}{dI} - \frac{d\log(c_0)}{dI}$	0.176	0.152	0.149	0.148	0.148
Firm Mass ($\eta = 1$): $\left(\frac{\eta}{\mu - 1}\right)\frac{d\log(F)}{dI}$	0.518	0.476	0.470	0.470	0.470
Culling: $\frac{d\log(z_0^*)}{dI}$	0.230	0.279	0.286	0.286	0.286
Switching: $\left(\frac{1}{\mu - 1}\right)\frac{d\log(\theta)}{dI}$	0.077	0.093	0.095	0.095	0.095
Total:	1	1	1	1	1

CHAPTER II

Bad Times, Bad Jobs? How Recessions Affect Early Career Trajectories

2.1 Introduction

The state of the business cycle at labor market entry has substantial and persistent effects on the earnings and career trajectories of young workers.[1] Because of the strong element of chance associated with labor market entry during a recession, the popular press has labeled cohorts with the misfortune of being exposed to years of earnings losses as "unlucky."[2]

Initial wage losses for unlucky cohorts stem from the fact that availability of high-wage jobs is strongly procyclical (see, e.g., Okun, 1973 and McLaughlin and Bils, 2001); however, the translation of short-term fluctuations in wages into long-term scarring effects is a product of several different factors. First, search frictions can hinder the movement of workers between employers thereby extending the duration of recession-induced losses. Second, when these frictions rise with tenure, they can generate long-term human capital mismatch if recession entrants do not find work in jobs, occupations, and industries for which they have already specialized (see, e.g., Oreopoulos et al., 2012). Third, employers may be slow to learn about the true quality of recession entrants relative to expansion entrants because of greater initial mismatch. Finally, the fact that much of the labor market is characterized by long-term wage-setting rather than a spot market slows down the convergence between recession entrants and expansion entrants (see, e.g, Beaudry and DiNardo, 1991).

In this chapter, we make two advances relative to the literature. First, we provide a precise estimate of the importance of employer-specific and non-employer-specific factors

[1] See, e.g., Kahn (2010) and Altonji et al. (2016) for evidence from the United States and Oreopoulos et al. (2012) for evidence from Canada.

[2] See, e.g., during a recession: Catherine Rampell, "Many With New College Degree Find the Job Market Humbling," *New York Times*, May 8, 2011. In contrast, during an expansion: Ben Casselman, "This Year's College Grads Are The Luckiest In A Decade," *FiveThirtyEight*, May 6, 2016.

in explaining the long-term impact of recessions on the wages of labor market entrants. While prior studies have found that employer and occupational characteristics play a role in explaining recession-induced penalties (see, e.g., Oyer, 2006, Oyer, 2008, Oreopoulos et al., 2012, and Rinz, 2019), the precise, quantitative role employers play in generating scars for recessionary entrants remains unknown. Our approach relies on the two-way fixed effect wage decomposition developed in Abowd et al. (1999) (AKM), which we use to partition recession-induced wage losses into components that are employer-specific and non-employer-specific. This exercise provides a more concrete comparison of between- versus within-employer explanations of the long run consequences of entering the labor market during a recession relative to existing literature.

Second, and perhaps more importantly, we provide the first evidence regarding how non-pecuniary compensation changes for unlucky cohorts relative to lucky cohorts. We estimate the value of working for each employer in our dataset by implementing a revealed preference-based estimator of job utility developed in Sorkin (2018). Within this framework, non-pay amenities are measured as variation in employer-specific pay holding employer value fixed, whereas rents accruing to workers are measured as variation in employer-specific pay that is explained by employer value.[3] From a welfare perspective, this exercise facilitates a more holistic accounting of recession-induced scarring because utility losses depend not only on losses in pay but also on changes to non-pay amenities (see, e.g., Rosen, 1987).

Our analysis uses linked employer-employee administrative data from Germany and studies the trajectories of new graduates of vocational training programs. Over two-thirds of the German workforce holds a vocational training degree, making our results relevant for a large proportion of the labor market—including a variety of skill types, occupations, and industries. Using these data, we establish three central findings. First, we show that the broad story of recession-induced scarring exists even in an economy with strong active labor market programs for employment and re-training. The typical recession in Germany lowers wages for new entrants by 4.9 percent cumulated over a 10 year horizon.[4] Second, we show that 1.9 percentage points (40 percent) of the total loss is explained by workers matching to lower paying firms. The remaining 3 percentage points (60 percent) of the total loss comes from non-employer-specific factors, including human capital mismatch, slow market-wide employer learning, and infrequent wage renegotiation. Our results therefore indicate that

[3]Under the revealed preference approach, job utility combines all unobserved amenities including such factors as job security, hours flexibility, commuting convenience, etc.

[4]The magnitude of recession-induced losses for unlucky cohorts in Germany is approximately equal to that found in Canadian data. Oreopoulos et al. (2012) find that a typical recession results in a 5 percent loss of earnings cumulated over 10 years. Estimates for recession-induced losses on unlucky cohorts in the United States are substantially larger. For each percentage point increase in the unemployment rate, Kahn (2010) finds a 6-7 percent loss in wages that decays to 2.5 percent after 15 years.

the majority of recession-induced losses in pay come from factors that are not specific to employers but rather a product of the labor market in general.

Finally, we show that 1.5 percentage points of the 1.9 percentage point employer-specific-pay penalty is compensated for by non-pay amenities whereas only 0.4 percentage points reflect losses from rent sharing. Thus, fully three-fourths of the employer-specific pay penalty is explained by higher non-pecuniary compensation. After netting out relative gains in non-pay amenities, the cost of recessions for young German labor market entrants drops by 30 percent from 4.9 percent of wages cumulated over 10 years to 3.4 percent. These estimates indicate that the welfare costs of labor market entry during a recession are overstated by a non-trivial margin if evaluated using only pecuniary losses.[5]

Using rich data on workers and establishments, we then assess the mechanisms that drive disparate career trajectories in the face of differing entry conditions. We show that low-pay, high-amenity employers are more likely to make job offers during recessions. Young workers in unlucky cohorts are beholden to the offer distribution they face upon entry, and their set of options are tilted towards towards employers that offer amenities in lieu of pay. This finding is consistent with prior research that links compensating differentials to unemployment risk (see, e.g., Abowd and Ashenfelter, 1981). It is also consistent with the observation that sectors which feature cyclically stable labor demand, such as education or healthcare, are often associated with higher job satisfaction than sectors which feature cyclically sensitive labor demand, such as construction or finance.

In addition, we show that recessionary entrants are much less likely to work at the same employer, in the same occupation, or in the same industry for which they trained. Our findings therefore suggest that much of the wage penalty faced by unlucky cohorts owes to losses in employer-, occupation-, and industry-specific human capital, analogous to the experience of workers who face involuntary job loss.[6] By illustrating the importance of changes in pay and amenities over the business cycle, our results also provide new evidence for the view that wage gains associated with industry switching during expansions are partly driven by lower non-pay amenities.[7]

The rest of the chapter is organized as follows: Section 2.2 provides a simple theoretical

[5]It is important to add that our conclusion on welfare implications relates only to employer-specific utility and may not capture other welfare relevant consequences of young workers' exposure to adverse aggregate conditions. For instance, Maclean (2013) and Schwandt and von Wachter (2016) find evidence of worse health and increased mortality among unlucky cohorts in the United States.

[6]See, e.g., Neal (1995) emphasizing the importance of industry-specific human capital in displacement scarring. See von Wachter and Bender (2006) emphasizing the importance of firm-specific human capital especially in the context of German vocational trainees.

[7]McLaughlin and Bils (2001) conjecture that such a phenomenon is possible but do not verify it empirically.

framework that formalizes the role of labor market risk in generating compensating differentials and explains how cyclical shocks alter the relative composition of wages and amenities available in the labor market. Section 2.3 describes linked employer-employee data from the Institute for Employment Research (IAB) of the German Federal Employment Agency, and how we utilized this data to construct our variables of interest. Section 2.4 describes our empirical strategy and presents our main results along with robustness checks. Section 2.6 concludes.

2.2 Theoretical Framework

This section outlines a simple model of compensating wage differentials generated by variation in unemployment risk across firms . The model formalizes one important channel through which workers who enter the labor market during recessions obtain higher non-pay amenities relative to those who enter during expansions.

2.2.1 Setup

Consider an economy with homogenous risk-averse workers and two types of firms: type R firms and type S firms. The production function for each firm type $j \in \{R, S\}$ in period $t$is

$$y_{jt} = z_{jt}F(h,k), \;\; F_h > 0, F_k > 0, F_{hh} < 0$$

where z_{jt} is an firm-specific productivity draw and F is a production function whose argument h represents hours employed in production and whose argument k represents capital. $F_{hh} < 0$ simply operationalizes a diminishing marginal product of labor. From an employment perspective, type R firms are deemed to be risky whereas type S firms are deemed to be safe in the sense that:

$$\begin{bmatrix} z_{Rt} \\ z_{St} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{z} \\ \bar{z} \end{bmatrix}, \begin{bmatrix} \sigma_R^2 & \rho_{RS} \\ \rho_{RS} & \sigma_S^2 \end{bmatrix} \right) \tag{2.2.1}$$

$$\rho_{RS} > 0 \tag{2.2.2}$$

$$\sigma_R^2 > \sigma_S^2 \tag{2.2.3}$$

In equation (2.2.1), the mean productivity level $\bar{z}$ represents the steady state. Although the shocks are positively correlated, type R firms exhibit greater variance in productivity than type S firms. A cyclical shock is defined to occur when both firm types obtain productivity draws that are jointly above or jointly below the steady state level. In recessions, $z_{jt} < \bar{z}$ and in expansions, $z_{jt} > \bar{z}$ for each j. For any realization of firm-specific productivity shocks and for fixed capital stock $\bar{k}$, equilibrium labor demand, h, can be generically written using the firm's first order condition for profit maximization as

$$F_h(h^*_{jt}, \bar{k}) = \frac{w}{z_{jt}}, \tag{2.2.4}$$

where w is the equilibrium wage rate and h^*_j is the optimal hours demand for the firm. $F_{hh} < 0$ guarantees that $\frac{\partial h^*_{jt}}{\partial z_{jt}} > 0$.

2.2.2 Wage determination

We assume that the labor market does not function as a spot market. Instead, as in Azariadis (1975) and Abowd and Ashenfelter (1981), workers and firms agree to long-term implicit contracts where wages are fixed but hours are variable. The key element of these long-term implicit contracts is that they insure risk averse workers against future labor market shocks that are propagated through firms' productivity draws.[8] The wage associated with each contract can be ascertained as follows. First, define the indirect utility for a worker who receives a wage w at firm j as $V(w, h_j)$, with utility an increasing, concave function of earnings, $U(w \cdot h_{jt})$. At a common wage rate, w, the expected indirect utility associated with a type R contract will be lower than the expected utility associated with a type S contract. This conclusion follows from the fact that z_{Rt} is drawn from a distribution that second order stochastically dominates z_{St} and that higher variability in productivity translates directly into higher variability of hours demand through (2.2.4). Consequently, risk aversion on the part of workers implies that

$$\mathbb{E}[V(w, h_{Rt})] < \mathbb{E}[V(w, h_{St})].$$

Figure 2.1 illustrates this gap in expected indirect utility over a range of potential hours realizations around the steady state level $\bar{h}$. In a competitive labor market, workers will demand a wage premium to compensate for the additional risk associated with employment at type R firms. Define the risk premium as the compensating wage differential $\Delta \equiv w_R - w_S > 0$ where w_R is the wage rate for a type R firm and w_S is the wage rate for a type S firm. The equilibrium wage rates are those which ensure that the expected indirect utility associated

[8] See Rosen (1985) and references therein on other models of risk premia in long-term implicit labor market contracts.

with employment at each firm type is equalized:

$$\mathbb{E}\left[V(w_S+\Delta,h_{Rt})\right]=\mathbb{E}\left[V(w_S,h_{St})\right].$$

2.2.3 Effect of cyclical shocks

Cyclical shocks re-order the expected relative productivity of each firm type. In expansions the following condition holds due to (2.2.1) and (2.2.3)[9]:

$$\mathbb{E}[z_{Rt}|z_{Rt}>\bar{z}]>\mathbb{E}[z_{St}|z_{St}>\bar{z}].$$

That is, risky firms are expected to be more productive than safe firms. Conversely, in recessions[10]:

$$\mathbb{E}[z_{Rt}|z_{Rt}<\bar{z}]<\mathbb{E}[z_{St}|z_{St}<\bar{z}].$$

That is, safe firms are expected to be more productive than risky firms. Combining the productivity shifts over the cycle with the labor demand equation (2.2.4) generates the following implications. First, expansions generate an increase in hours demanded by risky firms relative to safe firms. Thus,

$$\mathbb{E}[h_{Rt}|z_{Rt}>\bar{z}]>\mathbb{E}[h_{St}|z_{St}>\bar{z}]. \tag{2.2.5}$$

Second, recessions generate an increase in hours demanded by safe firms relative to risky firms. Thus,

$$\mathbb{E}[h_{Rt}|z_{Rt}<\bar{z}]<\mathbb{E}[h_{St}|z_{St}<\bar{z}]. \tag{2.2.6}$$

Finally, combining (2.2.5) and (2.2.6) shows us how the state of the economy affects hours demanded by risky *relative to* safe firms[11]:

$$\underbrace{\left|\mathbb{E}[h_{Rt}|z_{Rt}<\bar{z}]-\mathbb{E}[h_{Rt}|z_{Rt}>\bar{z}]\right|}_{\text{Drop in hours demanded by } R \text{ during recession}} > \underbrace{\left|\mathbb{E}[h_{St}|z_{St}<\bar{z}]-\mathbb{E}[h_{St}|z_{St}>\bar{z}]\right|}_{\text{Drop in hours demanded by } S \text{ during recession}}.$$

This simple framework illustrates two key implications for the observed wages of labor

[9]Specifically, $\mathbb{E}[z_{jt}|z_{jt}>\bar{z}]=\bar{z}+\sigma_j\sqrt{\frac{2}{\pi}}$ because z_{jt} is distributed normally for each j.

[10]Specifically, $\mathbb{E}[z_{jt}|z_{jt}>\bar{z}]=\bar{z}-\sigma_{jt}\sqrt{\frac{2}{\pi}}$.

[11]The same intuition holds in extensions with multiple firm types and firms with different average productivities (and therefore different levels of hours demand during steady state).

market entrants. First, the potential for cyclical shocks generates compensating wage differentials that are designed to indemnify workers against risk. Second, the state of the economy materially influences the share of hours demanded by risky firms relative to safe firms. This generates a composition effect in which recessions see relatively more hours demanded by safe firms compared to risky firms. Thus, we should expect labor market entrants to be relatively more likely to receive offers from low-wage/low-risk firms during a recession. We work toward an empirical test of this hypothesis in the following sections.

2.3 Data and Construction of Key Variables

Our analyses are based off the Sample of Integrated Labor Market Biographies (SIAB) (vom Berge et al., 2013). These data comprise a longitudinal two percent random sample of all individuals in Germany that ever worked, claimed unemployment insurance benefits, or sought job seeking assistance. In total, the sample describes the labor market histories of just over 1.6 million workers starting in 1975 and ending in 2010.[12]

The SIAB is a linked establishment-worker dataset that is organized in terms of spells. For employed individuals, each spell enumerates a match between a given worker and a given establishment. For non-employed individuals, a spell can enumerate a period of unemployment benefit receipt or a period of participation in an active labor market program. Spell lengths are measured at daily precision. In addition to demographics, data on individuals include average daily wages, educational qualifications, occupation, and state of residence. Establishment information is available at annual frequencies and includes industry codes, location, size, and median wage rates.

Because vocational training earnings are subject to Social Security contributions, spells of young workers who are apprentices in Germany's vocational training system are fully enumerated in the SIAB. With precise information on the occupation, wage rate, and start and end dates of training, the SIAB provides us with an unusually detailed set of information on workers both before and after they graduate from vocational training. In addition to exploiting the linked worker-establishment nature of the data, we rely on the timing of labor market entry made possible by observing workers before and after their training is complete.

Before delving further into how we estimate the impact of business cycle shocks on the career trajectories of young German workers, we first discuss how we use the SIAB to estimate variables that are critical inputs in our analyses. These variables include establishment-specific

[12]The data exclude employment in the civil service as well as self-employment. Marginal part time employment or so called "mini-jobs" are tracked in the SIAB starting in 1999. Mini-jobs are low-wage jobs with a monthly income threshold of 450 euro. Participation in active labor market programs is tracked starting from 2000.

measures of wage premia, utility, compensating differentials, rents, and occupation-specific unemployment rates at the state level.

2.3.1 Wage Decomposition

Following AKM, Card et al. (2013) (CHK) use the Integrated Employment Biographies (IEB)—which is the universe of data from which the SIAB sample is drawn—to estimate person and establishment fixed effect components of daily average wages. For worker i employed at establishment j in year t, the decomposition can be summarized as

$$\log(\text{wage}_{it}) = \alpha_i + \psi_j \mathbf{1}\{i \text{ works at } j \text{ in } t\} + \boldsymbol{x}'_{it}\beta + r_{it} \tag{2.3.1}$$

Person fixed effects, α_i, incorporate individual specific skills that are rewarded equally across employers. The establishment fixed effect, ψ_j, is a proportional premium that is paid by establishment j to all its employees. $\boldsymbol{x}_{it}$ is a vector of unrestricted year dummies as well as quadratic and cubic terms in age fully interacted with educational attainment. These controls account for aggregate and life-cycle determinants of wages. Consistency of the parameter estimates requires that the error term, r_{it}, is uncorrelated with α_i, ψ_j and the $\boldsymbol{x}_{it}$. CHK provide a detailed discussion about the validity of the identifying assumptions.

The CHK person and establishment fixed effects are identified only within the connected set of establishments—i.e., the set of establishments that either hire from or lose workers to other establishments in the set. CHK find that the largest connected set encompasses 95 percent of establishments in the IEB. Because the SIAB is a two percent random sample of worker histories drawn from IEB, the set of connected employers is relatively small. Fortunately, the original CHK person and establishment fixed effects are provided as supplements to the SIAB dataset, precluding the need for analysts to re estimate (2.3.1) within a sparsely connected dataset.

2.3.2 Estimating Establishment Values

Much of the literature that builds on the AKM decomposition treats the establishment fixed effect as a measure of economic rents shared by workers. However, a long tradition in economics has posited that employer-specific components of pay (ψ_j) can vary not only because of factors such as rent sharing or efficiency wages, but also because of amenities that are priced in the labor market as compensating differentials.[13]

[13] See, e.g, Rosen (1974) and Rosen (1987) for theory, and Lucas (1977), Freeman (1978), and Brown (1980) for empirical evidence.

Building on this tradition, Sorkin (2018) proposes a novel methodology to estimate employer-specific utility, exploiting the voluntary movements of workers between employers in order to infer the relative utility of each employer, which is also referred to as employer value. Implementing this revealed preference argument requires three key assumptions. First, all workers have the same ex-ante preferences over jobs. Second, all jobs within an employer are deemed to be identical from the standpoint of non-pay characteristics. Finally, all workers—both employed and non-employed—search randomly from the same offer distribution.

Taking these assumptions to linked employer-employee data, Sorkin develops an estimator that aggregates the choices of workers into unique establishment values, the utility an employee derives from working at a particular establishment. Intuitively, the estimator rewards employers for making more hires from other high-quality employers and penalizes them for voluntary departures. Akin to the connectedness requirement in AKM and CHK, values are only calculable within the *strongly connected* set of employers. Strong connectivity is defined as a set of employers who both gain *and* lose workers to other employers in the set.

Because the establishment values are estimated using a revealed preference argument, a crucial step in this procedure is to separate voluntary from involuntary movements of workers. This distinction is required both for employer-to-employer movements (EE moves) and movements from employment to non-employment (EN moves).[14] Sorkin's methodology relies on the notion that workers who separate from establishments that are shrinking are more likely to be involuntarily displaced, whereas workers separating from growing establishments are more likely to be voluntary departures. Thus, comparing the rate of worker exit from shrinking and growing establishments provides a benchmark for the probability that a given move from a shrinking establishment is involuntary.[15] We mimic this procedure in our establishment value estimation.

Unlike Sorkin (2018), whose analyses are based on the universe of U.S. workers covered by unemployment insurance in the Longitudinal Household Employer Dynamics (LEHD) data, we rely on the set of establishments that employ workers in our two percent longitudinal sample. Differences across these data settings necessitate additional modifications to Sorkin's methodology. First, while he limits strong connectivity to firms linked only by EE flows,

[14]Note that unemployment and labor force non-participation are taken to be the same for the purposes of this estimation procedure.

[15]The rate of worker separation from growing establishments represents "expected" turnover that is fueled by worker quits. Then, the rate of worker separation from shrinking establishments that is above and beyond this "expected" rate disciplines the probability that a given worker move from a shrinking establishment is involuntary. The methodology thus assumes that all moves from growing establishments are voluntary. The likelihood of voluntary exit is separately estimated for EE and EN moves.

we expand the set to include employers linked both by EE flows as well as transitions of workers through non-employment. Because of the ubiquity of movements in and out of non-employment, this modification expands the scope of strong connectivity and makes the computation of values feasible even within a two percent random sample like the SIAB.[16] This modification is fully consistent with the estimating equations in Sorkin's model which allow non-employment to obtain its own relative value.[17] Second, when coding EE and EN transitions, Sorkin reduces the quarterly LEHD data where workers potentially have multiple employers in a given year, into a data set where each worker is associated with a single employer—also known as the annual dominant employer—in a given year. In the SIAB, job spell lengths are measured at daily precision, allowing us to take advantage of a more complete set of job-to-job transitions. When a worker has two or more overlapping job spells in a given year (i.e. works for multiple employers at the same time), we select the spell associated with the highest total earnings. We treat periods between jobs that are longer than 90 days as non-employment spells regardless of whether an individual received unemployment benefits or sought job seeking assistance. Third and finally, unlike Sorkin, we do not impose any earnings or age restrictions on workers in our sample in order to maximize the number of establishments that we can include in the strongly connected set.

2.3.3 Decomposing Establishment Fixed Effects into Rents and Compensating Differentials

Within the framework of the utility posting job-search model he proposes, Sorkin assumes that the value of being employed at employer j (V_j) can be written as an additively separable function of the employer-specific component of pay and an employer-specific non-pay amenity:

$$V_j = \omega(\psi_j + a_j). \tag{2.3.2}$$

where ω is utility per log euro, ψ_j is the employer-specific component of pay, and a_j is the employer-specific non-pay amenity. Using establishment value estimates and CHK establishment fixed effect estimates, we re-arrange equation (2.3.2) to estimate

$$\psi_j = \pi V_j + \epsilon_j \tag{2.3.3}$$

[16]We rely on the MATLAB package `MatlabBGL` provided by David Gleich to find the largest strongly connected set in the SIAB.

[17]Put differently, workers who make EN and NE transitions help to identify both the estimate of non-employment value as well as estimates of employer values.

and then obtain the residual terms $\hat{\epsilon}_j = \psi_j - \hat{\pi}V_j$. Because the residuals are orthogonal to V_j by construction, they capture components of a_j that generate variation in pay holding utility constant. As such, they correspond to the non-pecuniary amenities defined as compensating differentials in Rosen (1987). The fitted value, $\hat{\pi}V_j$, is an estimate of rents accruing to workers because it captures variation in pay that is correlated with utility.[18]

Figure 2.2 shows the relationship between studentized values and establishment fixed effects in the SIAB replicating a Figure 5 from Sorkin (2018), but with German data. The blue circles plot average establishment value and average fixed effects within each ventile of the establishment value distribution. The red lines show one standard deviation bands of the establishment fixed effect distribution within each value ventile. The upward slope of the line of best fit shows that workers typically value employers that pay more, thereby indicating the importance of rents. Nevertheless, there is a wide spread of establishment-specific pay variation conditional on a given level of establishment utility, thereby indicating the presence of compensating differentials.[19]

Figure 2.3 plots average values and establishment fixed effects across firms within broad industry categories. The line depicting the overall relationship between V_j and ψ_j across firms is the same as that in Figure 2.2, but the scatter plot across industries gives us further information about which sectors are associated with high pay and/or high amenities. As expected, sectors associated with higher pay and low amenities (above the line) include Mining, Construction, and Finance. Meanwhile, establishments in the Health & Social Work sector impart roughly the same value to their employees as Manufacturing establishments despite paying substantially lower, on average. This indicates that Health & Social work establishments offer more non-pay amenities to their workers, consistent with our prior beliefs about relative work satisfaction in these sectors. In total, we view Figure 2.3 as important qualitative validation that the values we have estimated have economic meaning.

[18] The CHK establishment fixed effects are estimated separately by gender, giving us two observations per establishment. Due to sample size limitations, we pool together worker flows of both genders when estimating establishment values. Thus, we additionally include a dummy variable for gender on the right-hand side of equation (2.3.3) to remove average differences in establishment-specific pay between men and women.

[19] Our estimates of establishment value are based off a 2 percent sample of workers and therefore embody measurement error. As a consequence, the slope of the line of best fit shown in Figure 2.2 is attenuated. In Appendix B.2, we use a split-sample instrumental variable (IV) approach to evaluate the quantitative impact of measurement error in our analyses. We find that OLS-based estimates of equation (2.3.3) are indeed attenuated relative to the IV estimates. However, correcting for this bias has no economically substantive impact on our key findings. Given that $\hat{\epsilon}_j$ (compensating differentials) and πV_j (rents) are ultimately used as outcome variables, this is to be expected as long as the measurement error is classical.

2.3.4 Occupation-Specific Unemployment Rates

In order to capture variation in business cycle conditions relevant to the young workers we study, we take advantage of the SIAB's extensive data and large sample size to estimate two customized unemployment rates. Our preferred measure is the state- and occupation-specific unemployment rate. This measure provides an effective representation of labor demand, especially for workers entering the labor market in a given state with training in a given occupation. In addition, we calculate the occupation-specific national unemployment rate. This unemployment rate provides a broader measure of labor demand that allows us to test whether our results are driven by endogenous migratory responses.

Each unemployment rate is calculated using a similar methodology. We first assign every worker in our dataset a status of employed, unemployed, or out of the labor force on the 15th of each month using the SIAB's labor market status variable.[20] We assign employed individuals the occupation of their current job and unemployed workers their last known occupation. For employed workers who are currently in multiple jobs with multiple occupations, we assign them the occupation of the job that paid higher daily wages. Using the state of residence variable included in the SIAB, we then aggregate these observations into monthly, occupation-specific unemployment rates. Finally, we aggregate monthly unemployment rates into yearly unemployment rates by averaging across months and weighting monthly rates by the underlying number of observations associated with each month.

The primary value of using the SIAB to construct unemployment rates is to better characterize the state of the business cycle specific to young workers as they enter the labor market. Given that the data come directly from administrative sources, they also have the advantage of capturing the the unemployment rate as well as, or better than, survey-based unemployment rates. Nonetheless, to demonstrate that unemployment rates calculated directly from the SIAB match publicly-available unemployment rate measures for Germany, we present a comparison between our estimates of the national unemployment rate from the SIAB and two measures provided by the Organization for Economic Co-Operation and Development (OECD) in Figure 2.4. The first OECD unemployment measure is survey-based while the second is the registered unemployment rate, calculated from administrative data by counting the number of individuals who register with the government as unemployed in order to receive unemployment benefits, auxiliary benefits like community assistance and health assistance, or to signal the need for assistance with job search. The administrative data-based OECD measure more closely aligns with the SIAB-based estimates in levels

[20]The SIAB provides detailed information about worker status that we aggregate to three simple categories.

because both use registration for unemployment benefits rather than survey self-reports to ascertain unemployment.[21]

The trends of these three measures track each other closely, which is the relevant check for our panel data-based analysis. Furthermore, at the sub-national level, a within-state regression of survey-based OECD unemployment rates on unemployment rates estimated from the SIAB produces a coefficient of 0.978 (t =16.39).[22] In sum, while the OECD does not report occupation-specific unemployment rates for direct comparison, we are confident that our measure identifies relevant changes to business cycle conditions. Meanwhile, constructing the unemployment rate with administrative data allows us to define an individual's labor market based on both occupation and state, which more accurately represents the conditions they face when entering the labor market.

2.3.5 Sample Construction

An important characteristic of the SIAB is that it distinguishes between employment spells associated with vocational training from those that are not. Since this distinction is crucial for the implementation of our identification strategy, our most important sample restriction is to only consider workers whose first spell in the SIAB is in vocational training. We make four additional restrictions to hone in on the career paths of young trainees during the first ten years of their labor market experience. First, we limit our sample to those who complete their training before turning 30. Second, in order to keep comparisons consistent over time, we restrict our sample to April 1999 and later, after which the SIAB began enumerating mini-job work spells.[23][24] Third, we drop observations in occupation-state-years from which unemployment rates were computed with 20 or fewer observations. Finally, in order to avoid shifting composition based on outcome variables, we restrict the analysis

[21]The difference in levels between the two OECD sources arises for two reasons. First, non-employed survey respondents may indicate that they are job-seekers even though they are not registered as unemployed (if, for example, they are not eligible for unemployment benefits and thus see no advantage in registering). Second, survey respondents who have registered to receive unemployment benefits may indicate that they are not seeking a job when responding to the survey (if, for example, they register primarily to get auxiliary benefits such as health or community assistance). The latter group outweighs the former in our study period and likely also accounts for the fact that unemployment rates estimated from the SIAB are higher than the survey-based measure. In addition, unemployment rates obtained from the SIAB are higher in levels because these data exclude civil servants and self-employed workers who likely exhibit lower rates of joblessness than the rest of the labor force.

[22]The specification is $U_{st}^{\text{OECD survey-based}} = \alpha + \rho U_{st}^{\text{SIAB}} + \theta_s + \varepsilon_{st}$ where s is a state. Registered unemployment rates are not available at the state level from the OECD.

[23]Data drawn before and after this period are difficult to compare because of the change in enumeration.

[24]CHK estimate Equation (2.3.1) using different time windows in the IEB. We rely on estimates from the 2002-2009 time window and extrapolate these backward to 1999 and forward to 2010 to cover our sample window.

dataset to those individuals who are employed at establishments where establishment fixed effects and values are available, when they are employed. As noted in Sections 2.3.1 and 2.3.2, this restriction amounts to analyzing workers who are employed by establishments in the strongly connected set.

In order to align the timing of the analysis, we assign each worker a dominant employer for each year. The dominant employer is defined as the employer that pays a given worker the most in earnings for a calendar year. We obtain earnings for a given job spell by multiplying daily wages by spell length, then sum within employer to determine the total earnings from each employer for a given worker within a year. The employer with the highest earnings is given "dominant" status, and it is this employer's value, daily wage, and other associated characteristics that are used as outcomes below.[25]

2.4 Identification

Our identifying assumptions exploit key features of Germany's apprenticeship training system. Before proceeding to the empirical analysis, we first provide some institutional background on how the system works and why it lends itself to our estimation strategy.

Apprenticeship training in Germany is also known as dual vocational training because it combines workplace and classroom training in a roughly 60-40 split. The typical young worker begins her vocational training after secondary schooling by starting an employer-sponsored apprenticeship in one of approximately 350 officially recognized occupations. Employers, unions, and government agencies jointly regulate the course content and program length associated with training in each of the occupations to meet quality standards. Trainee wages are set by collective bargaining agreements which vary both by state and by occupation (Kuppe et al., 2013). Apprenticeships are the most common form of higher education in Germany, with over two-thirds of the workforce holding a vocational training degree.

Two aspects of the German apprenticeship system are particularly important for our analyses. First, occupational segmenting of the German labor market is a natural consequence of a system that is designed to promote occupation-specific skills among young workers. This feature makes between-occupation heterogeneity a more important dimension of youth labor market sorting in Germany than the United States, for example. Second, the duration of training programs are regulated, with the typical course taking about 3 years and culminating in a qualifying examination. The pre-set training duration makes it less likely that young workers can selectively enter the labor market when cyclical conditions are favorable. Even

[25]In cases where there is more than one spell with the dominant employer, the daily wage is a weighted average across spells where the weight is the number of days.

so, we use the mode within detailed training occupation to measure expected, rather than observed, training duration. This allows for a reduced form strategy in which we assign individuals a date of labor market entry based on expected training duration rather than their true labor market entry date, which could be subject to limited manipulation. Year of entry is thus defined by the training start date plus the modal training duration time in a given occupation.

Given this institutional setting, our identifying assumption is that occupation-specific unemployment rates prevailing at the time of expected labor market entry are unrelated to the timing of labor market entry. Given that we assign the timing of labor market entry based on the modal training duration within an occupation, this assumption rests on the notion that individuals do not change their training occupation or the timing of their training *start* based on labor market conditions that manifest (usually 3) years later. This broad assumption encompasses two important points: first, workers with particular unobserved characteristics cannot manipulate their initial labor market conditions through selective entry; and second, that the introduction of a particularly poor cohort in terms of unobserved variables is not responsible for adverse aggregate conditions.

Under this framework, we estimate the effect of initial aggregate conditions on labor market outcomes by exploiting variation in the occupation-state-cohort (osc) specific unemployment rate U_{osc} using the following specification:

$$y_{it} = \beta_e U_{osc} + \Gamma \mathbf{X}_i + \theta_{s(i)} + \theta_{o(i)} + \theta_{c(i)} + \theta_{e(i)} + \theta_t + \varepsilon_{it} \tag{2.4.1}$$

In equation (2.4.1), i represents an individual, $\theta_{s(i)}$ is a vector of state of training fixed effects, $\theta_{o(i)}$ is a vector of training occupation fixed effects, $\theta_{c(i)}$ is a vector of year-of-expected-entry (training start + modal training duration in the detailed occupation) fixed effects, $\theta_{e(i)}$ is a vector of potential experience (year minus year of expected entry) fixed effects, θ_t is a vector of year fixed effects, and $\mathbf{X}_i$ is a vector that contains a constant term, a dummy for the individual's gender, a dummy for whether the individual is a German citizen, and a vector of fixed effects for the individual's age at start of training.[26] We use β_e to trace out the average effect of initial labor market conditions U_{osc} on outcomes y_{it} for the first ten years of young workers' careers, where career start is defined by predicted training end. We cluster standard errors at the occupation-state-cohort level.

Table 2.1 provides initial evidence for the validity of our identifying assumptions by showing that workers are similar in terms of age, training wages, nationality, gender, and successful completion of apprenticeships whether they enter the labor market when the

[26]As in Oreopoulos et al. (2012), we identify $\theta_{c(i)}$, $\theta_{e(i)}$, and θ_t by dropping an extra year fixed effect.

unemployment rate is below or above the median of a given occupation-state cell. To the extent that they are economically meaningful, any differences in these observed characteristics are eliminated because we include them in the conditioning set $\mathbf{X}_i$.

To provide further evidence for our identification strategy, we formally test whether young workers are able to speed up or delay entry based on aggregate labor market conditions by re-estimating Equation (2.4.1) using training duration as the outcome:

$$[\text{Training Duration}]_i = \beta U_{osc} + \Gamma \mathbf{X}_i + \theta_{s(i)} + \theta_{o(i)} + \theta_{c(i)} + \theta_{e(i)} + \varepsilon_{it} \tag{2.4.2}$$

We do this under two scenarios: in the first, $c(i)$ is an individual's actual year of entry, and in the second, $c(i)$ is expected year of entry based on modal training duration within occupation—our preferred measure. Table 2.3 shows the results from this validation exercise. Two key points emerge. First, there appears to be some ability for workers to manipulate training duration based on the business cycle, but on average, this ability is very small. When entry is defined by the last day of an individual's training, a one standard deviation increase in U_{osc} generates 12 day increase in training duration, on average. In contrast, when entry is defined by training start date plus the modal training completion time in an occupation, there is no relationship between U_{osc} and training duration. We take this as evidence that our reduced form strategy corrects a small endogeneity bias relative to assigning individuals their true date of training completion.

2.5 Effect of Cyclical Shocks on Pecuniary and Non-Pecuniary Outcomes

2.5.1 Primary Results

The four panels of Figure 2.5 show the effect of a 1 percentage point increase in U_{osc} at expected labor market entry on daily wages, establishment fixed effects, rents, and compensating differentials over 10 years of potential experience. The top left panel illustrates that young workers face an initial wage loss of about 0.6 percent, a gap that steadily narrows over the next 10 years. The wage losses incurred in these first 10 years are economically meaningful: our estimates imply that a one standard deviation increase in U_{osc} at the time of labor market entry induces a 4.9 percent present discounted value (PDV) loss in daily wages cumulated over the next decade.[27] Assuming away differences in earnings arising from the number of days worked, our PDV daily wage loss estimate from Germany is similar to

[27] To conduct this calculation, we use the mean daily wage at each potential experience year, $\bar{w}_e$, the standard deviation of U_{osc}, $\sigma_U = 7.34$, a discount rate $r = 0.05$, and the β_e coefficients in the following

the 5 percent earnings loss accrued over 10 years for the average Canadian recession graduate estimated in Oreopoulos et al. (2012).[28]

Establishment fixed effects, shown in the top right panel, drop by about 0.2 percent on impact, and slowly rise. Taken on its own, the establishment fixed effect result hides important dynamics that would be invisible were they not decomposed into rent and compensating differential components. The effect of adverse entry conditions on these sub-components of wages are shown in the lower two panels. The initial gap in rents is small and closes steadily over time, with no statistically significant that it still exists after 10 years. In contrast, non-pay amenities (or negative compensating differentials) rise by about 0.15 percent on impact and then slowly converge by year 10. These results imply that unlucky workers start their careers in lower-paying jobs that feature higher non-pay amenities compared to their lucky counterparts. Over time, these workers are successful in transitioning to higher paying establishments and closing the gap in rents. However, they continue to work at establishments that provide a relatively higher share of non-pay amenities until year 10.

2.5.2 Career Paths of "Lucky" and "Unlucky" Cohorts

Figure 2.6 illustrates the career paths of young workers in absolute rather than relative terms. Establishment value (V) is shown on the horizontal axis and establishment fixed effects (ψ) are shown on the vertical axis. The lines show the first 10 years of career trajectories in (V, ψ) space for two different initial conditions, obtained by plotting the predicted values of V and ψ from Equation (2.4.1) for each level of potential experience using counterfactual values of U_{osc}. The "Low U" group faces initial unemployment rates at the 10th percentile of within occupation-state unemployment rates, while the "High U" group faces initial unemployment rates at the 90th percentile. Lastly, the gray dashed lined plots a fitted line estimated from Equation (2.3.3) showing the average relationship between establishment fixed effects and establishment values for employers in the SIAB.

Both "Low U" and "High U" groups move to the northwest, gaining in value and in employer specific pay on essentially the same line—an approximation of the early-career job ladder in (V, ψ) space. Workers who face low initial unemployment rates begin their careers at higher paying, higher value firms. However, a part of this pay gap comes from lower non-pay amenities (this is evidenced by the fact that the "Low U" career path is further

expression: $100 \times \left(1 - \frac{\sum_{e=0}^{10}\left[\bar{w}_e(1+\sigma_U\beta_e)/(1+r)^e\right]}{\sum_{e=0}^{10}\left[\bar{w}_e/(1+r)^e\right]}\right)$.

[28] Oreopoulos et al. (2012) construct their estimate assuming that recessions induce a 5 percentage point change in the regional unemployment rate. During our study period (1999-2010), the regional unemployment rate in Germany exhibited a standard deviation of 4.9 percentage points.

above the dashed line than the "High U" career path in any given year). Both types of workers retain the higher-pay/lower-amenity trade off as they gain experience, but "High U" workers start from a lower level. In general, the career path has a higher slope than the average relationship between establishment fixed effects and establishment values indicating that individuals trade non-pay amenities for pecuniary compensation as they move up the job ladder.

2.5.3 Decomposition of Recession-Induced Losses

We next decompose recession-induced losses into four major categories in Table 2.4.[29] Appendix B.1 provides details on how we calculate the various estimates presented in the table. As mentioned above, the total pecuniary loss for cohorts who enter the labor market during a recession is estimated as a 4.9 percent reduction in the present value of wages, cumulated over a decade.

The subsequent rows of the table illustrate various novel decompositions that we are able to calculate by applying unique data and empirical techniques. The first row shows that 1.9 percentage points, or 40 percent, of the total loss is attributable purely to the fact that unlucky cohorts match with lower paying employers. These employer-specific losses are further split into rents and amenities in the second and third rows of the table. The 1.9 percentage point reduction in employer-specific pay arises from a 0.4 percentage point reduction in rents and a 1.5 percentage point gain in amenities. Thus, fully three-quarters of employer-specific pay reductions are compensated for by relative gains in amenities. The final row of the table shows that 3 percentage points, or 60 percent, of the total loss is not explained by employer-specific factors. This 60 percent incorporates losses due to factors such as human capital mismatch, changes in outside offers, slow market-wide learning, and infrequent wage re-negotiation.

Taken together, the results shown in Figure 2.5, Figure 2.6, and Table 2.4 present new perspective about the effects of cyclical conditions on the early career outcomes of young workers. The welfare cost of recession entry, when viewed only in pecuniary terms, is a 4.9 percent reduction in the present value of wages. However, when we account for the fact that 1.5 percentage points of that loss is compensated for by relative gains in non-pay amenities, the overall impact of typical recession is effectively a 3.4 percent reduction in the present value of wages. Consequently, using pecuniary measures alone overstates the overall loss by about 30 percent. In other words, a purely pecuniary comparison between the two cohorts presents an incomplete picture about the welfare losses imposed by initial labor market

[29] These estimates reflect a one standard deviation increase in unemployment rates at entry which is intended to simulate a recession.

conditions.

2.5.4 Mechanism 1: Heterogeneity in Job Creation Over the Cycle

Having illustrated the major implications of cyclical shocks on young workers' career trajectories, we now turn to investigate the mechanisms that drive the results we discussed above. The first of these mechanisms is based on the model we presented in Section 2.2which explicitly incorporates a compensating differential for employment stability (lack of unemployment hazard) into the competitive equilibrium wage rate. In this framework, some establishments are less cyclical in their hiring and firing decisions, and can therefore offer lower equilibrium wages. The overall composition of hiring establishments would tilt toward these stable establishments during a contraction, decreasing the average establishment fixed effect measured for contractionary entrants and increasing the average measured compensating differential. This divergence across firms could occur for a variety of reasons, including both cross- and within-sector differences in the cyclicality of productivity.

Here, we investigate whether firms with higher measured non-pecuniary amenities are less cyclical in their hiring by relating establishment growth to the business cycle, stratified by an establishment's measured time-invariant amenity value, a_j. The regression model generally takes the form

$$g_{jst} = \alpha + \gamma_U(U_{st}) + \gamma_a(a_j^{CD}) + \gamma_{int}(U_{st} \times a_j^{CD}) + \alpha_s + \alpha_t + \varepsilon_{jst} \tag{2.5.1}$$

where g_{jst} represents establishment growth between year t and $t+1$, U_{st} is the state-year unemployment rate, and a_j^{CD} is a studentized version of $-\hat{\epsilon}_j$ from Equation (2.3.3) — our measure of non-pecuniary amenities that generate variation in pay holding utility constant. γ_{int} is our key coefficient of interest, measuring the extent to which establishments with higher non-pecuniary amenities are more ($\gamma_{int} > 0$) or less ($\gamma_{int} < 0$) cyclical in their hiring.

The results from estimating Equation (2.5.1) can be seen in Table 2.5. They offer evidence that establishments that provide higher non-pay amenities are less cyclical in their net employment growth. This suggests a direct link between the amenities we measure and an establishment's hiring decisions: those establishments that are more likely to hire during a recession naturally offer more job security, and this job security is itself an amenity. Workers are naturally more likely to receive job offers from more stable establishments (and industries) when they enter the labor market during an economic contraction, which naturally increases their non-pecuniary compensation and decreases their wage rate.

2.5.5 Mechanism 2: Displacement-Induced Human Capital Mismatch

We further use the rich structure of the SIAB to uncover patterns of worker mobility that underpin the results from Section 2.5. The four panels of Figure 2.7 summarize these mechanisms by showing the impact of a one percentage point increase in U_{osc} at labor market entry on the probability of remaining in one's training industry and the probability of remaining in one's training occupation. As with the previous results, we show these impacts over a 10 year horizon.

Figure 2.7 indicates that adverse initial conditions have the effect of displacing newly trained workers from their training industries and training occupations. Because training employers are important contributors to wage growth in the German context, the displacement we see in these figures suggests that there are strong parallels between labor market entry during adverse cyclical conditions and the unemployment scar that follows involuntary job loss (e.g. Jacobson et al., 1993, Davis and von Wachter, 2011, and Krolikowski, 2017). Furthermore, mismatch in industry- and occupation- specific human capital engendered by this displacement likely explains some of the wage penalties, which also has analogs in the literature that proposes mechanisms for earnings losses following job loss (e.g. Jarosch, 2015 and Krolikowski, 2017).

However, focusing purely on pecuniary losses as much of the prior literature has, would have led one to conclude that the cumulative effect of these mobility scars is overwhelmingly negative. In contrast, we find that affected cohorts recover completely in terms of rents and values even though they are more likely to work outside of the industries and occupations in which they trained. These displacements also shed light on the nature of cyclical upgrading in the career paths traversed by lucky and unlucky cohorts. For example, McLaughlin and Bils (2001) note that the wage gains obtained by workers who enter high-wage industries during expansions could, in fact, be attributable to a compensating differential channel.[30] While they do not seek to empirically verify this possibility, the patterns we present here provide evidence both of its existence and of its magnitude. Taken together, the evidence on outcomes and mechanisms suggests that the overall costs of industry and occupational displacement may not be as damaging as one might have been led to conclude in the absence of richer measures of compensation.

2.5.6 Robustness Checks

In this section we investigate the robustness of our identification strategy to a variety of potential threats. We consider the effect of potential simultaneity that arises from estimating

[30]McLaughlin and Bils (2001) focus on a Roy-model and a queuing model to explain cyclical upgrading.

values as well as regression coefficients with the same underlying sample of worker mobility. Next, we examine the extent to which connectedness restrictions required for the estimation of establishment fixed effects and values generate selection bias in our sample. We find that our results remain robust to these concerns.

2.5.6.1 Simultaneity of Value and Regression Model Estimation

A potential concern about our empirical strategy comes from the fact that identification of the β_e coefficients in Equation (2.4.1) exploits some of the same worker moves that identify establishment values as described in Section 2.3.2. If rents and compensating differentials, dependent variables in Equation (2.4.1), are estimated using the same underlying source of variation as the β_e coefficients, then our estimates could exhibit simultaneity bias. In particular, the establishments that employ the workers in our sample could have higher values precisely because the workers in our sample chose to work there.

In order to purge our econometric analysis of such a bias, we re-estimate the establishment values by dropping all individuals who were trainees as of April 1999 or later from the set of worker moves. The results of this exercise are presented in Figure 2.8. The upper panel shows results based on values estimated from the restricted sample, whereas the lower panel duplicates the results from Figure 2.5 based on values estimated from the full sample. It appears that there is some scope for upward bias in estimated coefficients from Equation (2.4.1) on value-related outcomes. However, this bias does not alter the basic patterns of our results: workers who face worse initial aggregate conditions recover in terms of rents but not in terms of compensating differentials eight years later.

2.5.6.2 Sample Selection Bias

As discussed in Section 2.3.1 and 2.3.2, our analysis is restricted to the set of establishments for which establishment fixed effects and establishment values are estimable. The first restriction requires connectedness and the second requires strong connectedness. While these restrictions allow us to estimate Equation (2.4.1) using a variety of outcomes for a fixed sample, we want to ensure that our results are not influenced by endogenous sample selection. To do so, we take advantage of the fact that daily wages are available for all individual-establishment observations in our data, regardless of whether the establishments are in either of these connected sets.

Figure 2.9 contrasts our Figure 2.5 results for log wages from the restricted sample in the left panel to an unrestricted sample of all young workers in our data in the right panel.[31]

[31]These estimates are obtained using actual year of entry and U_{osc} (our preferred specification from Section

Comparing the two panels, we see that adverse initial labor market conditions have a very similar effect on log wages across the restricted and unrestricted samples, starting from a 0.6 percent dip for each percentage point increase in U_{osc} and recovering to around 0.4 percent by the eighth year of an individual's post-training career. Thus, while our analysis is confined to the 115,673 individuals represented by the left panel, this test provides good evidence that their labor market experience reflects the broader early career paths of young German workers.

2.6 Discussion and Conclusion

This chapter provides a new perspective on the costs imposed by cyclical shocks on young workers. Using rich employer-employee linked data from Germany we replicate the major finding of existing research showing that adverse cyclical conditions at labor market entry generate persistent wage losses on affected workers. Implementing a revealed preference based method of measuring compensating differentials across a large sample of establishments, we find that low wages for unlucky cohorts are explained by reduced rents and by offsetting compensating differentials, with the latter accounting for the majority. We offer suggestive evidence that employers hiring workers during downturns provide more long-term job security than those who hire workers during expansions, which explains part of the offsetting compensating differential that unlucky cohorts obtain. Our findings indicate that focusing on earnings losses alone overstates the long-term welfare losses of labor market entry during recessionary periods.

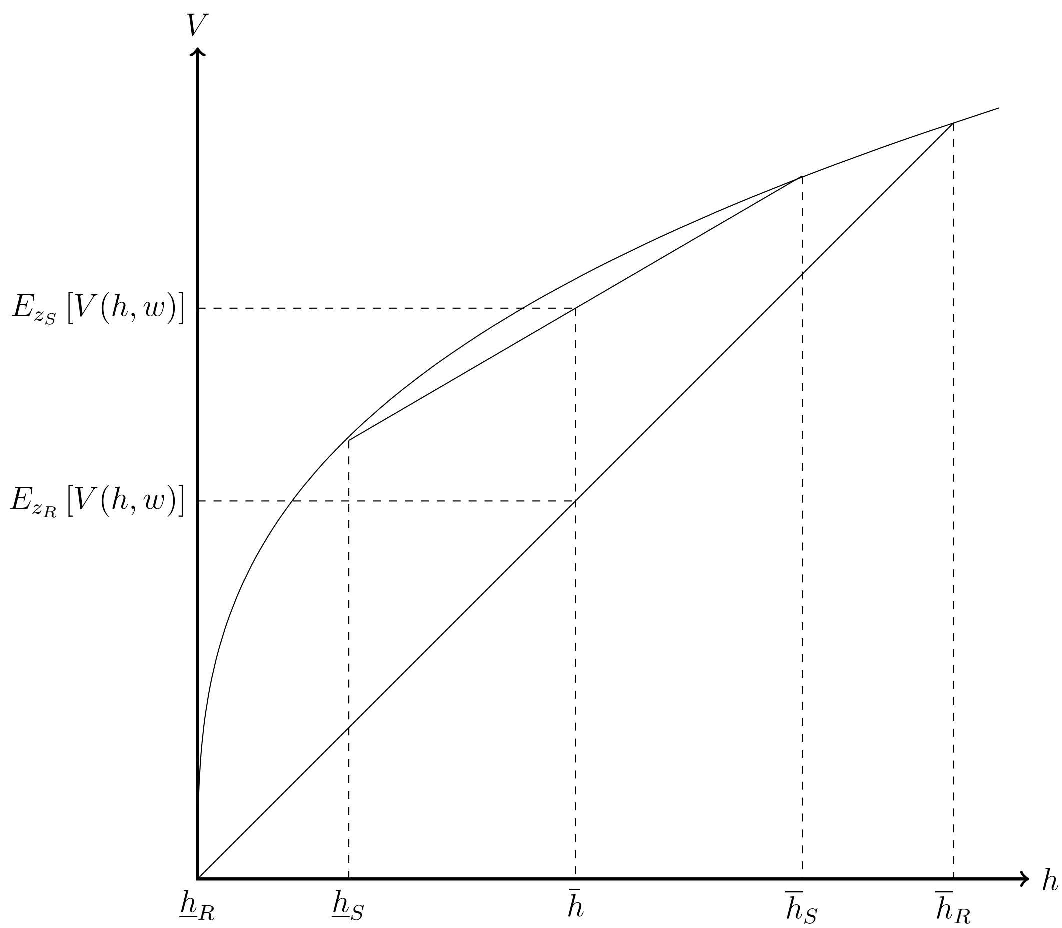
We find that cyclical shocks in Germany displace workers away from the occupations and industries in which they gained specialized apprenticeship training. As such, the wage losses that unlucky workers face is explained by the mismatch in human capital similar to that experienced by workers who suffer involuntary job loss. Nevertheless, our results indicate that these unlucky workers continue to climb the utility ladder of job quality by accruing returns in the amenity rather than rent component of establishment-specific pay. These findings shed new light on the role that compensating differentials play in cyclical upgrading and downgrading.

The sample of newly trained workers that we study includes a wide variety of skill types, broadening the relevance of our results beyond college graduates. Nevertheless, our conclusions are specific to the German labor market which differs substantially in terms of employment protections, unemployment benefits, healthcare provision, and re-training

2.5), but the comparative results are qualitatively similar within any of the robustness checks discussed in this section.

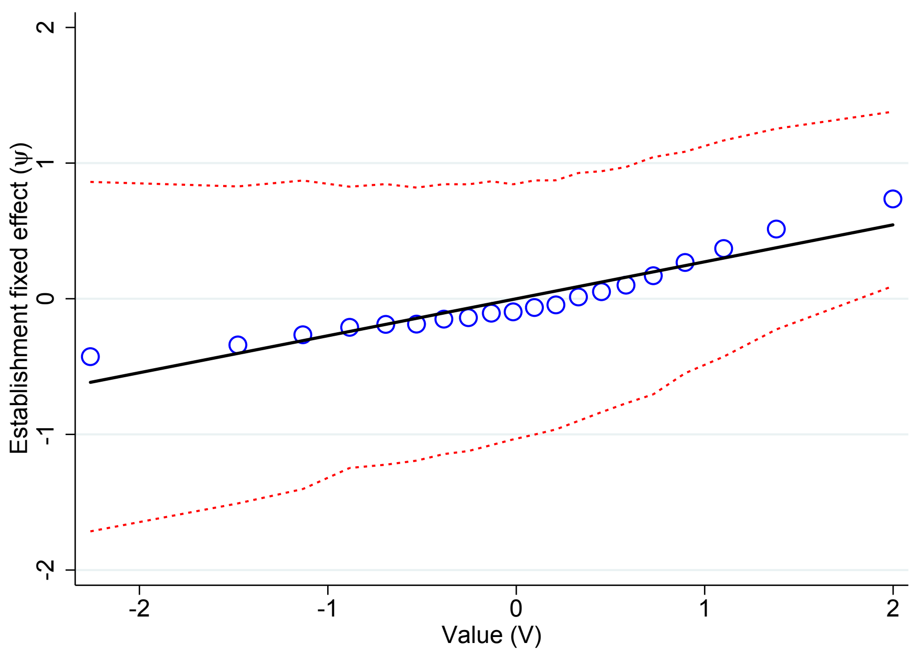
programs relative to the United States and Canada which are the two other countries in which recession-induced wage losses have been studied. Furthermore, the utility consequences that we focus on are isolated to job specific components and do not speak to factors such as health or consumption.

Figure 2.1: Compensating Wage Variation Due to Hours Volatility



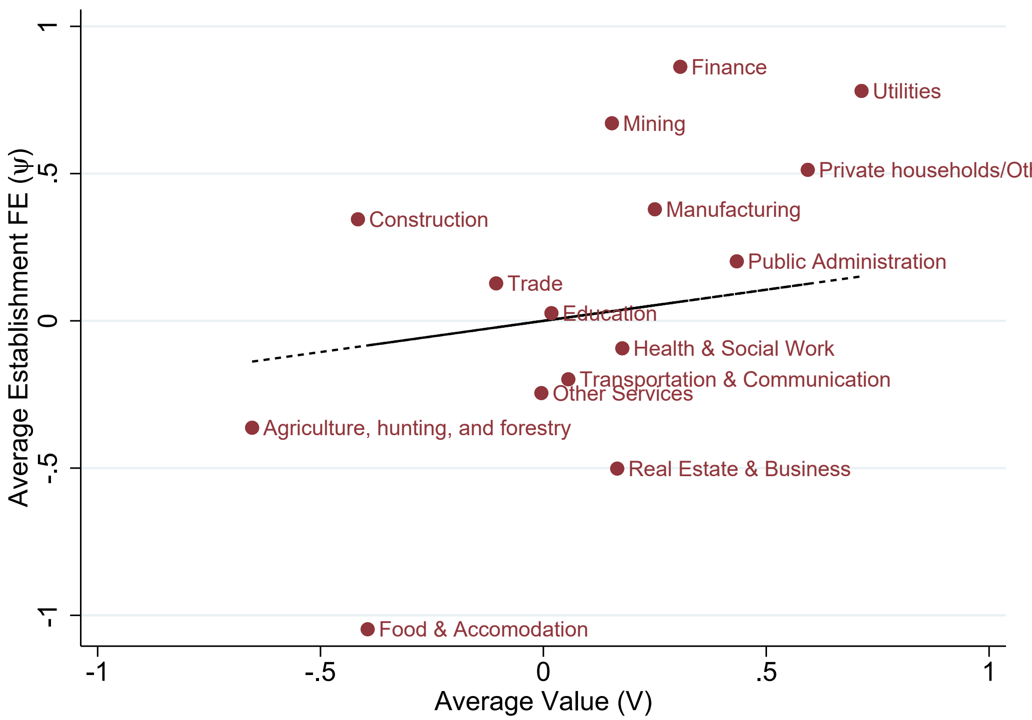
Notes: Thresholds illustrated on the horizontal axis represent the bounds of potential hours demand at risky (R) and safe (S) firms. $\bar{h}$ is the steady state level of hours demand associated with productivity $\bar{z}$. Underbars represent average values of h conditional on $z < \bar{z}$. Overbars represent average values of h conditional on $z > \bar{z}$.

Figure 2.2: Relationship Between Establishment Values and Compensation



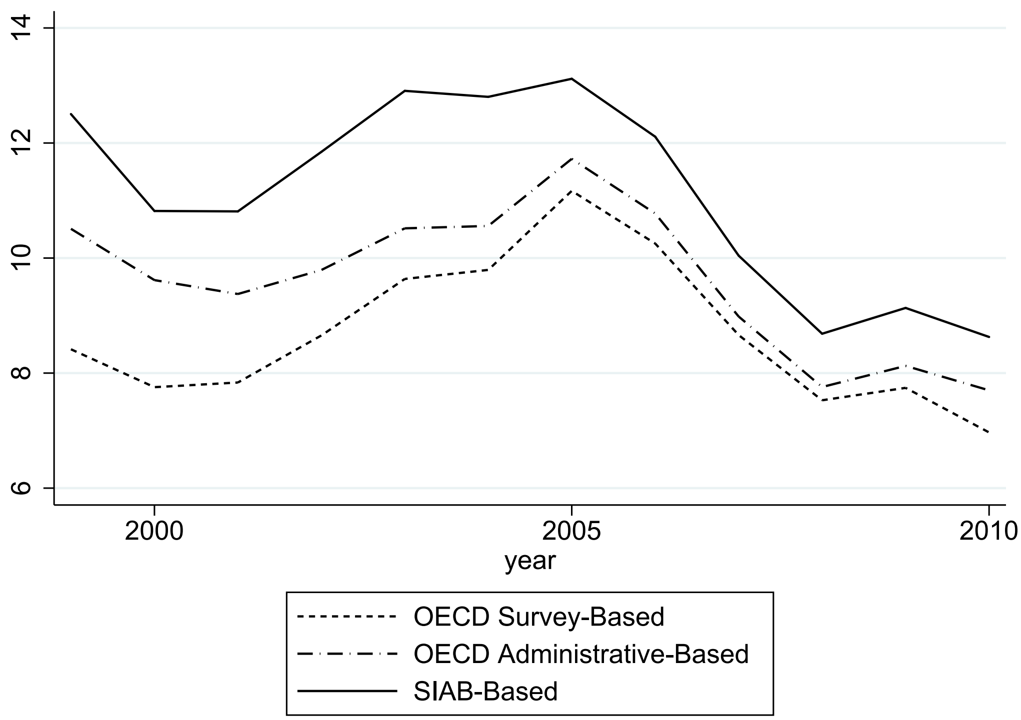
Notes: The value and establishment fixed effect estimates shown in this graph are studentized. The blue circles show the average establishment fixed effect and average establishment value within each ventile of the establishment value distribution. The red lines show one standard deviation bands of the establishment fixed effect distribution.

Figure 2.3: Relationship Between Establishment Values and Compensation Across Industries



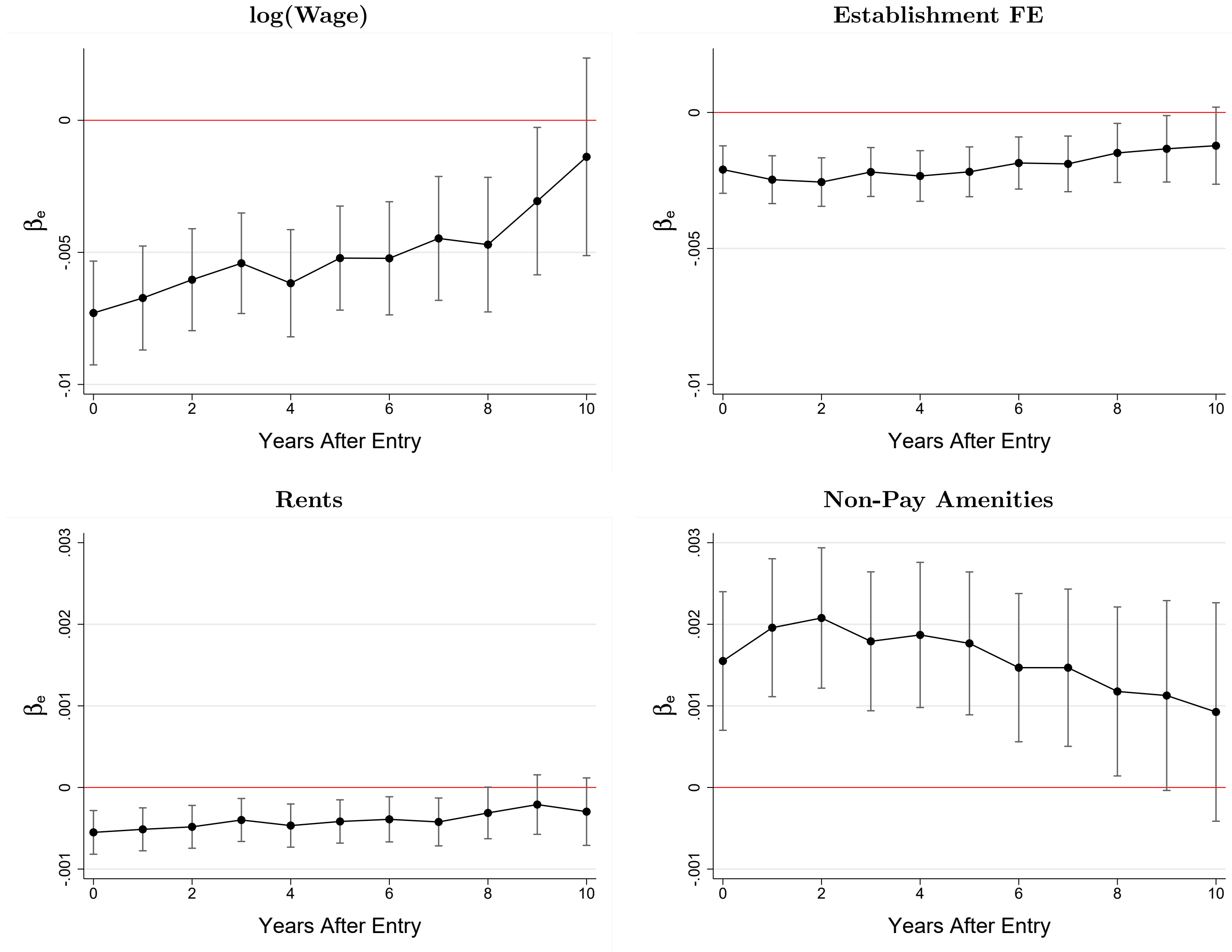
Notes: The value and establishment fixed effect estimates shown in this graph are studentized.

Figure 2.4: OECD versus SIAB National Unemployment Rates



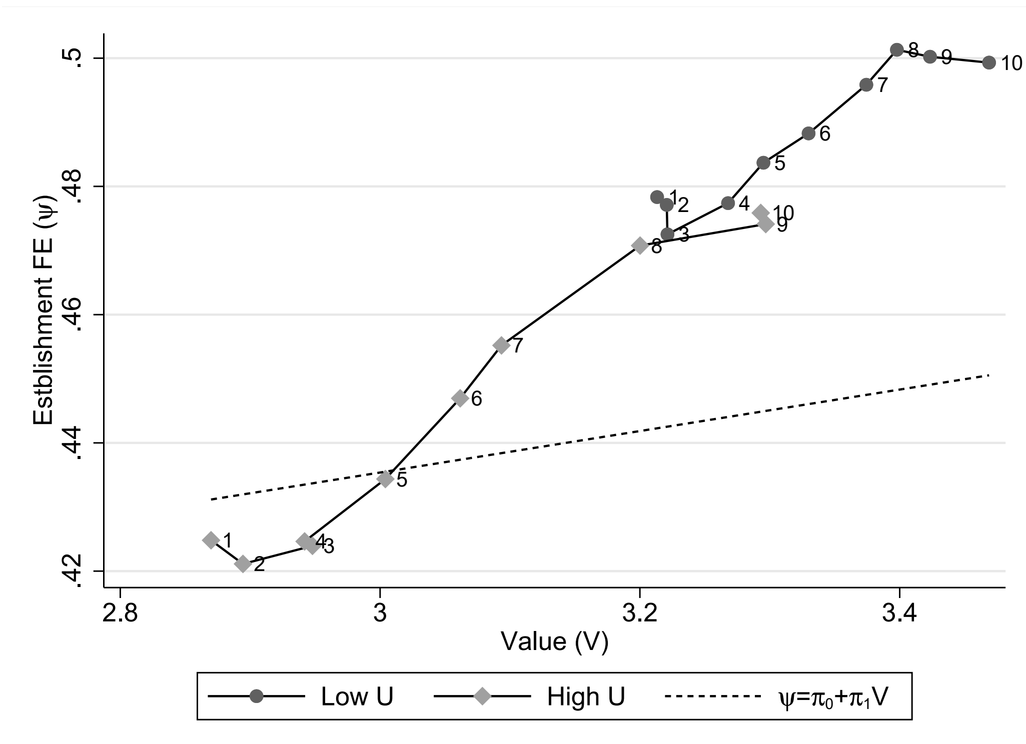
Notes: OECD unemployment rate obtained from OECD Stat. See Section 2.3.4 for details of of how unemployment rates are estimated using the SIAB.

Figure 2.5: The Effect of Entry Conditions (U_{osc}) on Early Career Outcomes



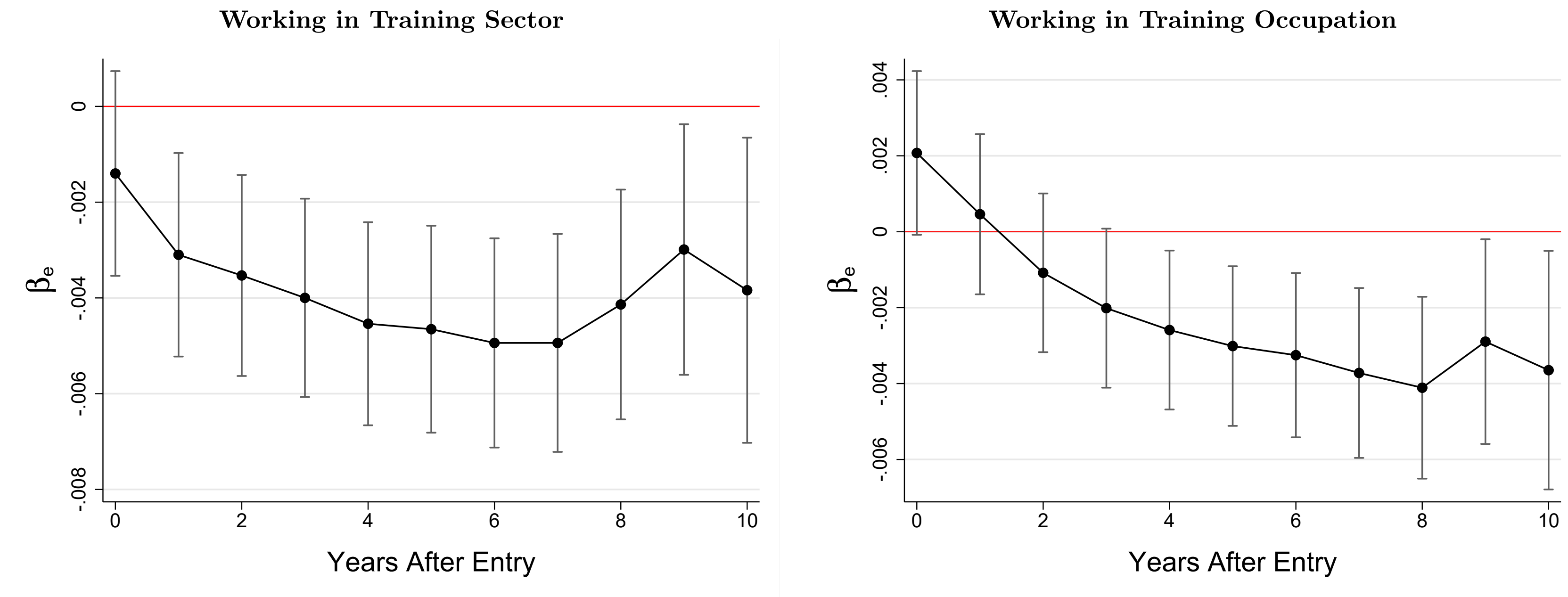
Notes: See Equation (2.4.1). All estimated coefficients are in log wage units. Sample size for each specification is 115,673. 95% confidence intervals represented by bars, with standard errors clustered at the state-cohort-occupation level. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Figure 2.6: Implied Career Paths of Young Workers



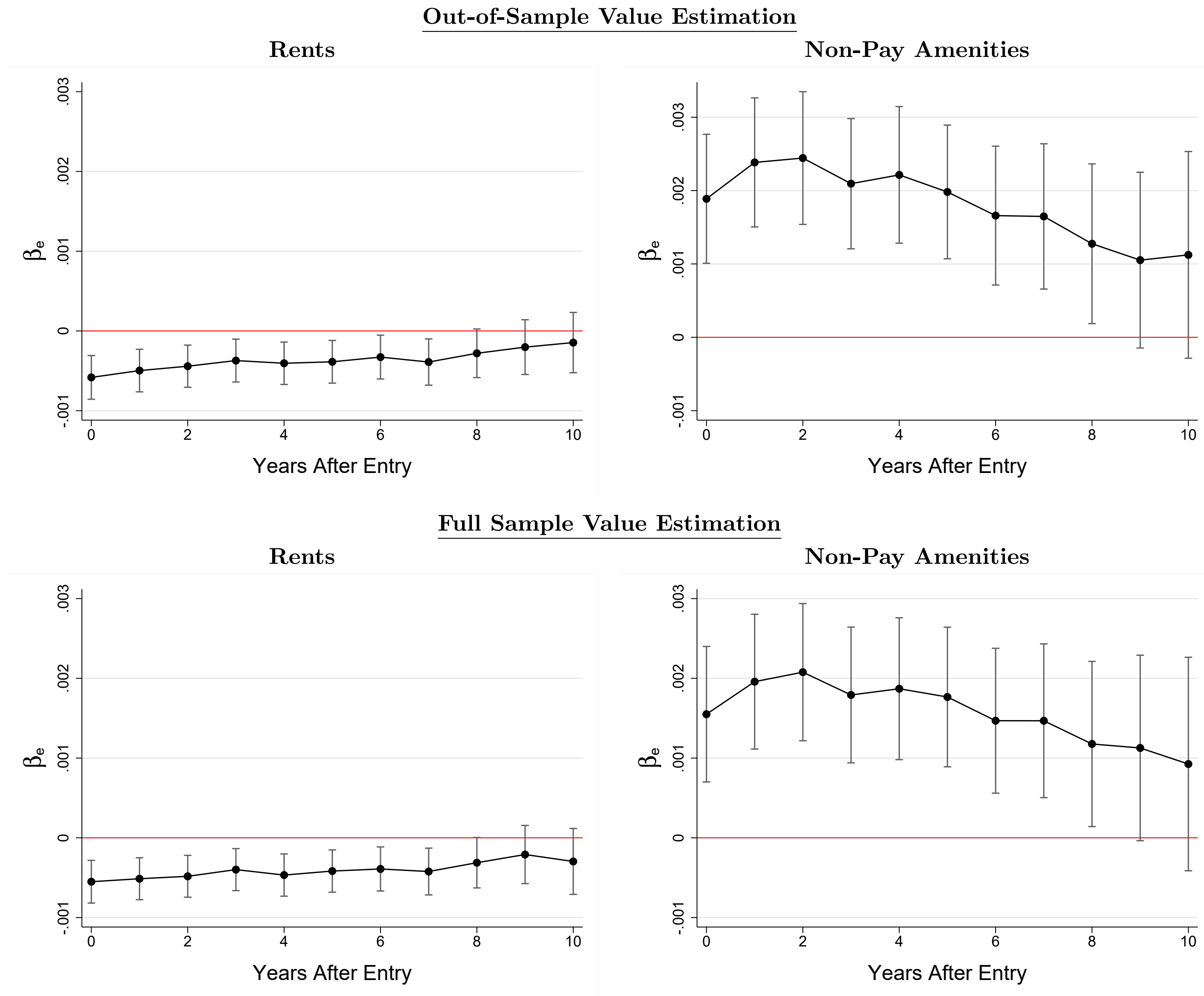
Notes: See Section 2.5 for details.

Figure 2.7: The Effect of Entry Conditions (U_{osc}) on Early Career Mobility



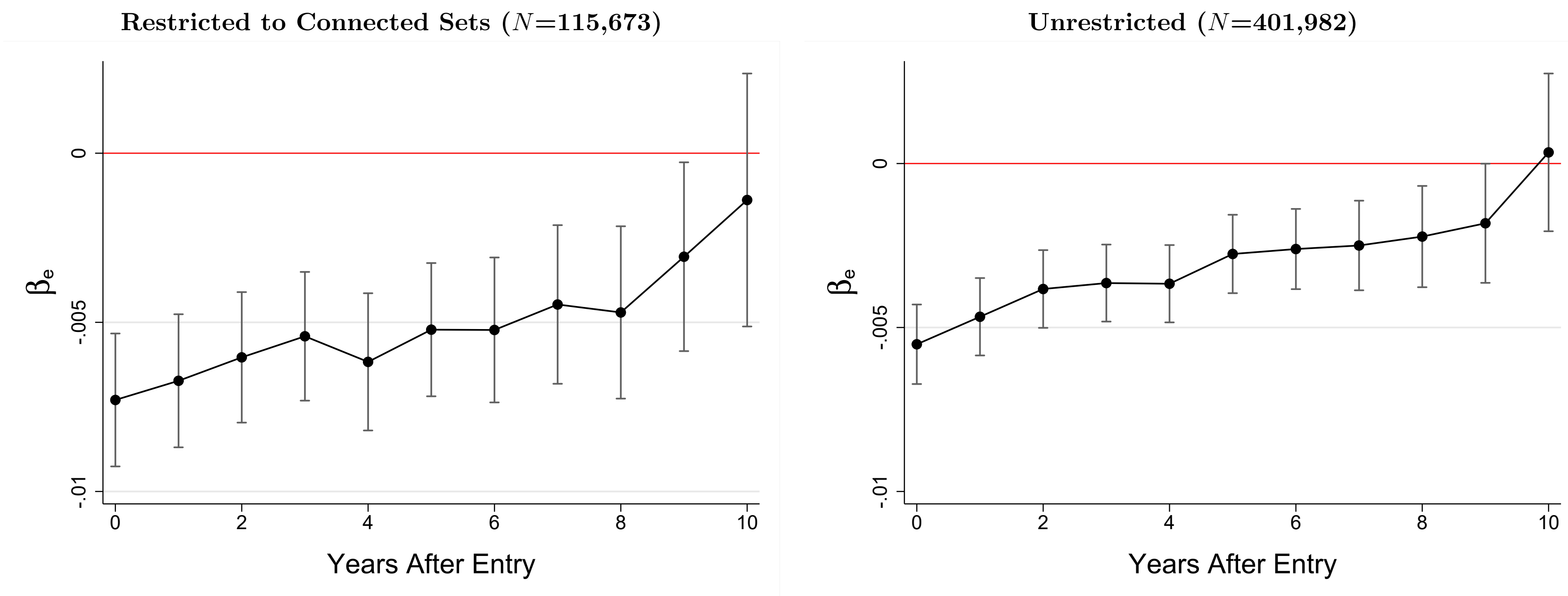
Notes: See Equation (2.4.1). All estimated coefficients are in log wage units. Sample size for each specification is 115,673. 95% confidence intervals represented by bars, with standard errors clustered at the state-cohort-occupation level. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Figure 2.8: Simultaneity Bias Robustness



Notes: Sample size for "Out-of-Sample Value Estimation" figures is 119,988. Sample size for "Full Sample Value Estimation" figures is 125,363. 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Figure 2.9: The Effect of Entry Conditions (U_{osc}) on Log Wages—Restricted versus Unrestricted Sample



Notes: 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German indicator variable, and a female indicator variable.

Table 2.1: Mean Individual Characteristics by Aggregate Entry Conditions

	$U_{osc} < U_{os}^{MED}$	$U_{osc} \geq U_{os}^{MED}$
Age at start of training	17.4	17.3
Median wage at training firm	87.9	86.2
German national (=1)	0.94	0.95
Female (=1)	0.41	0.43
Obtained vocational training certification (=1)	0.79	0.78

Notes: U_{os}^{MED} refers to the median occupation-state-specific unemployment rate during the study sample, 1999-2010. Median wage at training firm refers to median wage at individual's training firm in the year of training completion.

Table 2.2: Identifying Variation—Unemployment Rates Faced by Young Workers at Time of Entry

				Percentile				
Independent Variable	Description	Mean	St. Dev.	10	25	50	75	90
U_{osc}	State-Occupation Unemployment Rate	10.88	7.34	4.05	5.56	8.85	14.20	21.05
U_{oc}	National Occupation Unemployment Rate	9.94	3.80	6.18	7.67	9.57	11.24	12.86

Notes: Occupation assigned to individuals based on the occupation of their training apprenticeship. Unemployment rates calculated as described in Section 2.3.4.

Table 2.3: Identification Strategy—Validation Tests

	Training Duration (yrs)	Training Duration (yrs)
U_{osc}	0.0044**	-0.0015
	(0.0021)	(0.0030)
Year of Entry Definition	Last day of training	Predicted last day of training based on occ. mode

Notes: See Equation (2.4.2). * p<0.1 ** p<0.05 *** p<0.01.

Table 2.4: Decomposition of Recession-Induced Losses

	Δ PDV wage (10 years)
Losses in employer-specific pay	1.9%
Due to losses in rents	0.4%
Due to relative gains in non-pay amenities	1.5%
Other losses	3.0%
Total wage loss	4.9%

Table 2.5: Establishment Growth over the Business Cycle

	(1) Establishment Growth	(2) Establishment Growth	(3) Log Establishment Size
U_{st}: State-Year Unemployment Rate	-0.0050***	-0.0051**	-0.0079***
	(0.0012)	(0.0024)	(0.0018)
a_j^{CD}: Non-Pay Amenities	0.0019	—	-0.0035
	(0.0054)	—	(0.0069)
$U_{st} \times a_j^{CD}$	0.0017***	0.0088**	0.0015**
	(0.0005)	(0.0011)	(0.0007)
Establishment FE	No	Yes	No
Lagged Log Establishment Size	No	No	Yes

Notes: All models include state and year fixed effects. a_j^{CD} is studentized.
* p¡0.1 ** p¡0.05 *** p¡0.01.

CHAPTER III

Taken by Storm: Hurricanes, Migrant Networks, and U.S. Immigration

Moving from one's country of origin is among the most consequential decisions a person can make. Substantial numbers of people migrate internationally: estimates of migration over the five-year periods between 1990-95 and 2006-10 range from 34 to 41 million international migrants, or roughly 0.6 percent of world population (Abel and Sander, 2014). Substantially larger numbers—more than 600 million adults—express a desire to move permanently to another country (Pelham and Torres, 2008). Labor migration to the developed world leads to large income gains for migrants (McKenzie et al., 2010),which benefit not only the migrants themselves but also those remaining behind in origin countries. Remittances sent by migrants to their home countries amounted to $432 billion in 2015, far exceeding official development assistance (World Bank Group 2016), with substantial benefits for recipient households.[1]

It is also important to understand the economic impacts of natural disasters, and how those affected cope in their aftermath. Disasters cause extensive human losses and economic damages worldwide. Hurricanes are among the most damaging, accounting for roughly 40 percent of deaths and 38 percent of monetary damages caused by all natural disasters from 1995-2015 (CRED 2016). Migration in response to a natural disaster can help affected populations escape worsened living conditions in their home areas (Piguet et al., 2011). Understanding the effects of weather-related disasters becomes additionally important due to climate change. Anthropogenic warming of the climate has been linked to increased frequency and intensity of Atlantic hurricanes from the 1970s to the present (Walsh et al., 2016). Climate models predict increases in the frequency of the most intense hurricanes, and the intensity of accompanying rainfall, as the planet continues to warm (Kossin et al., 2017). A better understanding of the impacts of hurricanes on migration, and the extent to which

[1]Studies include Yang and Martinez (2006), Yang (2006), Yang (2008b), Gibson et al. (2014), Ambler et al. (2015), Clemens and Tiongson (2017), and Theoharides (2018).

such impacts are heterogeneous across storm-affected areas, can be an important input into estimates of the economic and human impacts of climate change.

In this context, a number of interrelated questions are of interest. Do negative shocks in migrant-origin countries encourage or inhibit international migration? What roles do prior migrant networks play in facilitating the outmigration response to origin-country shocks? If the effect on migration is positive, does it occur via legal or illegal (undocumented) channels? To what extent is any resulting migration temporary or permanent? None of these questions have obvious answers from a theoretical standpoint, and past empirical findings are either nonexistent or point in different directions across studies.

We estimate the impact of hurricanes on international migration to the U.S. from 159 origin areas over a quarter-century, examine how this response is moderated by the existence of prior migrant networks, and assess the extent to which migration responses operate via legal or undocumented (illegal) channels. Theoretically, the impact of hurricanes on migration, and the extent to which the effect of hurricanes is heterogeneous with respect to prior migrant networks, is unclear. Consider individuals choosing whether to stay in home locations or to bear a fixed cost and migrate to a more attractive destination. Home-location negative shocks have an ambiguous effect on migration: while they raise the return to migration, they also can raise the fixed cost of migration (or make it more difficult to finance migration fixed costs). In addition, the extent to which prior migrant networks stimulate additional migration in response to home-country shocks is unclear. On the one hand, prior migrant networks can reduce the fixed cost of migrating, making migration more responsive to negative shocks at home. On the other hand, insurance provided by prior migrants (remittances sent in the wake of shocks) can reduce the desire to migrate.

Our empirical work aims to resolve these theoretical ambiguities in an important international migration context. Our outcome of interest is annual U.S. immigration rates from 1980-2004 for each observable origin location, as constructed from U.S. Census data. We exploit exogenous variation in the returns to migration, as well as substantial cross-sectional variation in a key determinant of the fixed cost of migration. Variation over time in the return to migration from home locations is generated by hurricanes, which exogenously lower the attractiveness of remaining at home.[2] Variation in the fixed cost of migration is generated by the size of migrant networks in the U.S. (the stock of previous migrants). We examine whether increases in the returns to migration driven by origin-country hurricanes have larger impacts on migration from countries that have larger pre-existing migrant networks in the U.S.

[2] Yang (2008a), Noy (2009), Strobl (2011), Imberman et al. (2012), Hsiang and Jina (2014), Boustan et al. (2017), and Franklin and Labonne (2019) among others.

We find that hurricanes cause immediate and substantial increases in U.S. immigration on average. A one-standard-deviation increase in our measure of hurricane affectedness increases migration to the U.S. (as a share of the home country population) by 0.021 percent, which is 11.8 percent of the sample mean annual migration rate. This effect is magnified among origin countries with larger pre-existing stocks of U.S. immigrants. The effect of hurricanes on migration is positive for countries with a migrant stock in the U.S. (as share of 1980 population) of at least 0.86 percent—roughly the 70th percentile across countries.[3] For a country at the 90th percentile of the prior migrant stock (5.6 percent of origin population), a one-standard-deviation increase in our measure of hurricane affectedness causes an inflow amounting to 0.029 percent of the origin population.

A key question is whether the migrant stock should be interpreted primarily as affecting migration-related fixed costs, or whether it stands in for some other omitted variable. We take two approaches to address this issue. First, we seek evidence for mechanisms behind the heterogeneous effect. We find that a key role played by migrant stocks is formally sponsoring relatives for legal, permanent immigration. If we replace our dependent variable of interest with legal immigration counts from the U.S. Department of Homeland Security (DHS), our coefficient estimates are very similar in magnitude. There is clearly a substantial legal, permanent immigration response to hurricanes. This legal, permanent immigration is driven primarily by different forms of family sponsorship. These findings strongly suggest that migrant stocks reduce the fixed cost of migration by facilitating legal immigration.

Our second approach to addressing omitted-variable concerns is to gauge the stability of our key parameter, the coefficient on the hurricane-migrant stock interaction term, to the inclusion of additional control variables for other origin country characteristics, such as per capita GDP, distance from the U.S., and land area.[4] We show that the coefficient on our interaction term of interest is highly robust to inclusion of interaction terms with these other country characteristics. Hurricanes appear to have heterogeneous effects across countries due to migrant stocks themselves, and not some other origin-country characteristic correlated with migrant stocks.

This chapter presents a set of interrelated empirical findings that, to our knowledge, are new to the literature. It is among the first to test whether the migration response to negative home-country shocks is larger when the stock of prior migrant compatriots is larger. Previous work has examined the relationship between migration and prior migrant stocks,[5] on the one

[3]In the regression where the effect of hurricanes is allowed to vary with respect to the size of a country's prior U.S. migrant stock, the main effect of hurricanes is negative (although not statistically significantly different from zero).

[4]Because our coefficient of interest is on an interaction term with hurricanes, these predetermined control variables must also be included as interaction terms in the same way.

[5]Studies using "shift-share" instruments (e.g., Card (2001) and many others) have established that

hand, or between migration and home-country shocks, on the other, but not the interaction (how the migration response to shocks is affected by prior migrant stocks).[6] This is an important question whose answer is not theoretically obvious. Our findings help us better understand the risk-coping role of prior migrant stocks. They also can help policy-makers predict how migration may be affected by future shocks affecting populations of potential migrants, as such responses will vary with the size of their prior migrant stocks.

While prior research highlights multiple ways in which migrant networks might facilitate new migration, our work contributes by providing clear evidence of a particular mode of assistance: providing a less costly, legal route to immigration for their compatriots through family reunification immigration policies. This was not obvious ex-ante; prior studies have highlighted other factors such as financial assistance, information provision, and social support.[7] In addition, no prior work has shed light on the extent to which shock-induced migrants enter via legal or undocumented channels, or shown that home-country shocks affect new migration on both permanent (green card) and temporary (non-immigrant visa) margins.

Another distinguishing feature of our work is that our focus is the U.S., the world's largest migration destination country. Migration to the U.S. accounted for 18.5 percent of international migration flows from 1990-2010 (Abel and Sander, 2014). Our U.S. focus reduces concern about external validity of the findings. One of our empirical results that is not as new or unique takes on added relevance due to the importance of the U.S. as a migration destination. While other studies have estimated how negative shocks in home areas affect outmigration, it is important to know that, on average across all origin countries, a very important type of negative shock (hurricanes) increases migration to the world's largest migration destination.[8]

This study also makes advances related to data. To our knowledge, this is the first empirical analysis of U.S. immigration that uses restricted-access U.S. Census data to construct

aggregate immigration inflows tend to be apportioned to sub-national locations based on the geographic distribution of previous migrants. In contrast to most studies using shift-share instruments, we examine an aggregate shifter, hurricanes, which is clearly exogenous. That said, in this chapter we are not instrumenting for total immigration into particular labor markets. Hurricanes are not likely to provide sufficient statistical power for analyzing impacts of immigration on local labor market outcomes.

[6]Clemens (2017) finds that previous migration facilitates new migration of Central Americans in response to violence in home localities.

[7]Key references include Hatton and Williamson (1994), Massey (1988), Orrenius (1999), Munshi (2003), Amuedo-Dorantes and Mundra (2007), Dolfin and Genicot (2010), Beaman (2012), Docquier et al. (2014), and Blumenstock et al. (2018). None of these other factors are ruled out, of course, and our study is not well-positioned to shed light on these other factors.

[8]Focusing on the U.S. also has the advantage of providing us considerable cross-sectional and temporal variation across countries in hurricane-induced shocks to migration returns, as well as cross-sectional variation in migrant networks across origin locations.

country-by-year inflow estimates. Based on 1-in-6 long-form Census responses, our migration measures are more precise than any previous survey-based estimates used to examine the causal determinants of U.S. immigration. Relatedly, we are able to analyze migration flows from a larger sample of countries than is available in the public U.S. Census data, and this provides additional identifying variation because many small countries (e.g., island nations) are also hurricane-prone. We supplement these data with administrative immigration data from the DHS, which are also rarely used in economic analyses of migration. Finally, we construct hurricane-affectedness measures from satellite-based meteorological data, which are less prone to measurement error and biases than other disaster data sources (Yang, 2008a).

Our work is related to research on migration responses to the returns to migration, and in particular to work emphasizing causal identification. Past studies have found that increases in the returns to migration, driven by shocks in either sending or receiving areas, do increase net outmigration. In some studies the identifying variation comes from shocks in the source locations,[9] while in others the variation in returns is generated by shocks in destination locations.[10] Other studies have found the opposite—that increases in the returns to migration driven by negative shocks in home areas lead to *less* outmigration (Halliday, 2006; Yang and Choi, 2007; Yang, 2008c)—which may reflect the importance of migration fixed costs in combination with liquidity or credit constraints.[11]

We proceed next to Section 3.1, where we outline a simple theoretical framework to guide interpretation of findings. Section 3.2 then provides an overview of relevant U.S. immigration policy. Section 3.3 describes the data used in the analyses, and Section 3.4 reports empirical results. Section 3.5 concludes.

[9]For example, Hatton and Williamson (1993), Munshi (2003), Hanson and McIntosh (2012), Hornbeck (2012), Marchiori et al. (2012), Bohra-Mishra et al. (2014), Gröger and Zylberberg (2016), Abarcar (2017), Baez et al. (2017), Boustan et al. (2017), Clemens (2017), Kleemans and Magruder (2018), and Minale (2018).

[10]For example, Yang (2006), Wozniak (2010), McKenzie et al. (2014), Bertoli et al. (2016), Fajardo et al. (2017).

[11]Consistent with liquidity constraints inhibiting migration, Ardington et al. (2009), Bryan et al. (2014), and Angelucci (2015) find that cash transfers increase migration. Chernina et al. (2014) similarly find that the easing of liquidity constraints generated by titling reforms in early 20th Century Russia facilitated outmigration. Stecklov et al. (2005) and Imbert and Papp (2018) find contrary results to these. Bazzi (2017) finds that positive income shocks in origin areas in Indonesia lead to less migration in wealthier areas and more in poorer ones. Boustan et al. (2017) find that hurricanes in the U.S. lead to more internal migration, particularly among those with higher incomes.

3.1 Conceptual Matters

We first take a moment to consider theoretically how hurricanes might affect outmigration decisions, and how the effects of hurricanes might differ with respect to the size of prior migrant networks. We discuss these issues qualitatively here, and provide a simple theoretical framework in the Online Appendix.

Consider individuals in a migrant-origin area choosing whether to stay in home locations or to bear a fixed cost and migrate to a more attractive destination (which in practice we will take to be the U.S.) The fixed cost of migration is a function of the size of a migrant's network. Prior research suggests that the fixed cost of migration is lower when an individual has a larger migrant network (e.g., Bauer et al., 2005), for a number of potential reasons. Networks could help reduce search and information costs (e.g., related to legal and illegal modes of entry, employment, housing, etc.), provide social support during adjustment (a reduction in psychic costs), and sponsor relatives for legal immigration (allowing migrants to avoid costlier illegal entry routes and costly wait times imposed by quotas).[12]

Now consider a negative shock to economic conditions in the home country, such as a hurricane, which makes the destination country more attractive in relative terms. If the negative home-country shock has no effect on migration costs, the prediction is straightforward: migration will increase.

However, the hurricane's effect becomes ambiguous if the negative shock to the home country does affect migration costs. It is most plausible that negative home-country shocks would raise migration costs. Increased demand for legal migration assistance as well as illegal migration services (migration smugglers or coyotes) could raise equilibrium prices for those services. In addition, loss of assets due to hurricanes could make it more difficult for credit-constrained households to pay fixed migration costs. Negative shocks at home could make it more difficult to obtain credit to pay for the fixed costs of migration (Yang, 2008c), or could raise the opportunity cost of departure (Halliday, 2006). Negative aggregate shocks could also have general equilibrium effects that make it more difficult to pay fixed migration costs, such as reductions in asset prices (Rosenzweig and Wolpin, 1993) or wages (Jayachandran, 2006).

So far, we have emphasized migrant networks helping those in the home country respond to shocks by helping them migrate. But migrants can of course also assist in other ways. Migrants' geographic separation means their shocks are less correlated with those in the home

[12]These points have been emphasized by Massey (1988), Jasso and Rosenzweig (1989), Orrenius (1999), Orrenius and Zavodny (2005), Dolfin and Genicot (2010), and Comola and Mendola (2015). Networks could also provide financial assistance with paying fixed migration costs, which would be important in contexts where potential migrants are liquidity or credit constrained.

area, and so they are valued as members of the informal insurance network.[13] Migrants are therefore in a good position to provide insurance, by sending financial assistance (remittances) in response to negative shocks at home.[14] Better insurance at home can reduce the impact of home-area shocks on migration, because those affected by a shock can cope "in place," without migrating to escape the consequences (Morten, 2019). The extent of insurance (the fraction of the loss replaced) is likely to be larger when the migrant network is larger (as a share of home country population): when migrant networks are larger, more individuals in the home country should have a migrant social network member, and the financial burden of supporting disaster-affected home-country residents can be spread across more migrants. Therefore, the possibility of migrants sending shock-coping remittances attenuates the effect of shocks on new migration.

In sum, then, the theoretical predictions are ambiguous: negative shocks to economic conditions in the home country could increase migration by increasing the return to migration. It is also possible for negative home-country shocks to *reduce* migration, if such shocks themselves increase the fixed costs of migration or reduce ability to pay migration fixed costs. Even if the shocks themselves do not make it more difficult to pay migration fixed costs, the extent to which migrant networks facilitate migration in response to shocks is unclear, because prior migrants can send remittances to those in the home country instead of helping them migrate in response to shocks.

To resolve these theoretical ambiguities, we turn to empirical tests in Section 3.4.

3.2 Immigration Policy During the Sample Frame

Before moving to our analysis, we summarize U.S. immigration policy from 1980 through 2004. The workings of U.S. immigration policy help us highlight features of immigrant stock networks that have the potential to facilitate immigration.

The outline of today's U.S. immigration policy regime has its origins in the 1965 Immigration and Nationality Act. This legislation abolished preferential treatment for Europeans and created a system in which a majority of visas were allocated to relatives of U.S. citizens or residents. It was also the first law to distinguish between immediate relatives (spouses, children under age 21, and parents) of U.S. citizens, who became exempt from quotas, and other types of immigrants who fell into one of seven new preference tiers subject to numerical

[13]There is a large body of work on how households in developing countries cope with risk (Morduch, 1995). Lucas and Stark (1985), Rosenzweig and Stark (1989), and Munshi and Rosenzweig (2016) have emphasized the role of migration and remittances for informal risk-coping strategies.

[14]Jayachandran (2006), Yang and Choi (2007), Yang (2008a), Jack and Suri (2013), Blumenstock et al. (2016), De Weerdt and Hirvonen (2016), and Clemens (2017).

limitations (Kandel, 2016). Further, by 1979, all country-specific quotas were abandoned in favor of an overall quota. In 1981, the overall quota stood at 270,000 for all those subject to the cap (Clark et al., 2007). Among the capped tiers, first preference goes to unmarried adult sons and daughters of U.S. citizens, second preference goes to spouses and children of green card holders (LPRs), third preference goes to married sons and daughters of U.S. citizens, and fourth preference goes to siblings of U.S. citizens. Thus, while green card holders can sponsor a limited set of relatives from home, they are substantially constrained in this ability relative to naturalized immigrants.

The major change to policy that occurred during our sample period was the Immigration Act of 1990, which increased allowable total immigration to 675,000 and increased the limit of family-based immigrants subject to quotas from 290,000 to 480,000 (Kandel, 2016). Technically, immediate relatives of U.S. citizens came under this 480,000 cap for the first time, but in practice, the cap is "permeable" and inflows of such migrants remain *de facto* uncapped to the present day. The remaining 195,000 allotments are slotted for employment visas (140,000) and a new category of "diversity" visas (55,000) allocated to countries that did not send many migrants to the U.S. between 1965 and 1990 (Clark et al., 2007).

An additional change that occurred during our sample period was the 1986 Immigration Reform and Control Act (IRCA), which granted legal status to millions of undocumented workers. While this legislation had many consequences, it mainly affects our results through its disproportionate legalization of migrants from certain countries, perhaps creating a positive shock in the *effective* stock of network capital in the United States for these countries. This is especially true given how important legal status and citizenship are to being able to serve as a beach head for compatriots under the current regime. A more minor point is that the legal permanent resident (LPR) status granted to these previously undocumented workers clearly did not result from new entries into the United States. We will thus subtract these "inflows" from our overall measure of LPR admissions in the DHS data.

3.3 Data

3.3.1 Sample Definition

Our sample consists of foreign territories listed in Table 3.1. Given how often many of these areas are hit by hurricanes and because of the level of detail our data affords us, we treat many non-sovereign territories as separate countries (e.g., Guadeloupe or Martinique).[15] We drop countries that are U.S. territories because of their preferential treatment in immigration

[15]From this point forward, use of the word "country" includes these non-sovereign territories.

policy. We also drop countries from the former Soviet Union and the European land mass.[16] North Korea and Eritrea are excluded because of a lack of reliable migration information for the entire sample period. Additionally, some countries that contain inconsistent migration information due to border redefinition are combined to retain consistency throughout the sample period. These include the Netherlands Antilles minus Aruba,[17] Sudan,[18] and Guadeloupe.[19] Finally, we also drop any country without an immigrant stock estimate from the 1980 Census. This left us with a balanced panel of 159 countries.

3.3.2 Hurricane Index

Hurricanes are storms that originate over tropical oceans with wind speeds above 33 knots.[20] These severe storms create damages through storm surges, strong winds, and flooding, and their radius of impact can be anywhere from 60 to 900 miles. Thus, depending on the severity of the storm, there is a wide scope for hurricanes to inflict extensive damage, particularly when infrastructure is weak and production is agriculture-oriented. Hurricanes occur in six basins: Atlantic, East Pacific, West Pacific, South Pacific, South Indian, and North Indian. Yang (2008a) provides a more detailed definition of hurricanes and their architecture.

We construct a hurricane index representing the average hurricane exposure of residents in a given country-year following Yang (2008a). This index uses data from meteorological records, rather than impact estimates compiled from news reports, governments, or other similar sources due to concerns about measurement error and potential misreporting of hurricane damages (motivated, for example, by a desire to attract greater international disaster assistance). The meteorological data on hurricanes consists of "best tracks" compiled by Unisys from the National Oceanic and Atmospheric Administration's Tropical Prediction Center (for the Atlantic and East Pacific hurricane basins) and the Joint Typhoon Warning Center (for the West Pacific, South Pacific, South Indian, and North Indian hurricane basins). The best tracks contain information on the hurricane's maximum wind speed and the geographic coordinates of its center (or "eye") at six-hour intervals. Figure 3.1 displays all hurricane best tracks from 1980 through 2004.

[16]The splitting of the Soviet Union does not enable us to have reliable migration information for these countries throughout the sample period. Europe is rarely hit by hurricanes, and because it contains mostly developed countries is not likely to provide a useful migration counterfactual.

[17]Curacao, Bonaire, Saba, St. Eustatius, and Sint Maarten. The Netherlands Antilles was not dissolved until 2010.

[18]South Sudan and Sudan. South Sudan broke off from Sudan in 2011.

[19]Guadeloupe and St. Barthelemy. St. Barthelemy broke off from Guadeloupe in 2003.

[20]Hurricanes are also known in different regions as typhoons and cyclones. For simplicity, in this chapter hurricanes, typhoons, and cyclones will all be referred to as hurricanes.

The best track data naturally take hurricanes as the unit of analysis, and so in their raw form give no indication of countries affected. The Online Appendix describes in detail how we turn this best track data into a country-by-year index. Other papers have utilized similar hurricane indices to study their impacts on various outcomes on land masses (Belasen and Polachek, 2009; Hsiang, 2010; Strobl, 2011; Hsiang and Jina, 2014). All use a model based on best tracks to simulate the wind speed faced by geographical areas a certain distance away from the best track line.[21]

The resulting index can be described as "intensity-weighted hurricane events per capita," in which intensity is a nonlinear function of hurricane-force wind speed. The key features of this index are that it measures the average "affectedness" by hurricanes for residents of a country in a given year. It rises in the number of hurricanes affecting a country, the share of the population affected, and in the intensity (wind speed) of the hurricanes to which people were exposed. In Table 3.2 we provide basic summary statistics of the hurricane index. Out of 3,895 country-year observations, 641 have non-zero values of the index. The standard deviation of the non-zero values is 0.0542.

3.3.3 Immigrants in the United States: Stocks and Inflows

3.3.3.1 U.S. Census Bureau

The primary source for our immigration data is confidential data provided by the U.S. Census Bureau, who granted us access to the full set of responses from the 1980 and 2000 Long Form Censuses along with the 2005 through 2015 American Community Survey (ACS) 1-year files. The 1980 and 2000 Census Long Form provide 1 in 6 counts of all persons living in the United States along with demographic information.[22] The ACS 1-year files provide a one percent sample of all persons living in the United States in a given year. The Online Appendix describes how we utilize these data sources to construct two key variables: sending-country-by-year estimates of migration inflow rates (m_{jt}) and sending country estimates of 1980 U.S. immigrant stocks ($s_{j,1980}$).

[21] Strobl (2011) uses population weights when measuring the effect of hurricanes on economic activity, while Hsiang and Jina (2014) do not.

[22] In 1970, the Census Bureau began sending both a Short and Long Form questionnaire to households. The Long Form is sent to roughly 1 in 6 households, and remaining households are sent the Short Form. Many demographic variables of interest are only contained in the responses to Long Form questionnaires—the recent controversy surrounding a 2020 citizenship question on the Short Form notwithstanding. Most importantly here, the 1980 Long Form Census questionnaire contained questions on place of birth and citizenship, while the 1980 Short Form Census questionnaire did not.

3.3.3.2 Department of Homeland Security (DHS)

Our second source of migration inflow data comes from the DHS. In addition to producing the annual *Yearbook of Immigration Statistics* (1996-2015), the DHS houses the records of the former Immigration and Naturalization Service (INS), who produced similar publications for past years titled the *Statistical Yearbook of the Immigration and Naturalization Service* (prior to 1996). Starting in 1982, these annual publications contain counts of legal permanent residence (LPR) statuses granted by country of last residence, which we use to construct an alternate measure of migration inflows. They also contain information on non-immigrant entries into the U.S. by country of birth and class of admission starting in 1983, which we use to construct a new panel that measures potentially temporary migration.[23] Data through 1996 are available only as hard-copy portable documents. We thus double-entered and cross-checked each relevant table to ensure accuracy in these outcome variables.[24]

The DHS data provides some important advantages over our confidential Census data beyond their use as a robustness check. First, the counts were all taken officially during the year of a given immigrant's receipt of LPR status or non-immigrant entry and thus do not suffer from attrition due to death or return migration. Second, in the case of LPR entries, country of last residence provides a more direct indicator of hurricane-induced migration than country of birth. Third, the DHS data allows us to separate classes of LPR admission, such as uncapped family reunification, capped family sponsorship, and refugees. This allows us to examine whether eligibility for immigration due to family-reunification policies is a mechanism through which our effects operate.

Finally, the non-immigrant entry panel allows us to understand two additional facets of hurricane-induced migration into the United States. First, it helps us assess whether there is a component of such migration that is potentially temporary. Second, it helps us elucidate the phenomenon of conditional entry followed by either a switch of status or an overstay on a temporary visa, a process through which much legal and illegal permanent migration occurs.

There are, however, also drawbacks to the DHS data that highlight its complementarity with our estimates from the confidential Census Bureau data. First, the DHS LPR measures do not distinguish between new inflows and changes in status from temporary to permanent residence. Second and relatedly, backlogs and backlog reduction efforts create uncertainty around how reliably the DHS estimates can be used to measure changes in actual entries—as

[23] According to the DHS Office of Immigration Statistics, non-immigrant data is not available in 1997 due to concerns about data quality in that year.

[24] The hard copies are available at in the U.S. Citizenship and Immigration Services (USCIS) Historical Library's General Collection.

compared to switches in status from temporary to permanent—over time. Third, the DHS data cannot shed light on undocumented entries, while these may be captured by the Census and ACS surveys (which purposely do not inquire about legal status).[25] Fourth, while it contains information about class of admission, the DHS does not allow us to examine many other important demographic characteristics of migrants, such as age. Finally, neither the Census nor the DHS data can correct for migrants who still live abroad but whom obtain a green card (LPR status) to engage in repeated circular migration.

3.4 Analysis

3.4.1 Specification

In order to test the theoretical implications described in Section 3.1, we exploit the exogeneity of our objective hurricane index and conduct reduced form analyses that test its impact on migration inflows to the U.S. For this purpose, we rely primarily on two specifications:

$$y_{jt} = \beta_0 + \beta_1 H_{jt} + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{3.4.1}$$

$$y_{jt} = \gamma_0 + \gamma_1 H_{jt} + \gamma_2 (H_{jt} \times s_{j,1980}) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{3.4.2}$$

where y_{jt} is an outcome and t runs from 1980 through 2004. Our primary results are for $y_{jt} = m_{jt}$ where m_{jt} is the number of immigrants from country j to the U.S. in year t as a proportion of country j's population in 1980. Analogously, $s_{j,1980}$ is the stock of immigrants from country j already in the U.S. in 1980 as a proportion of country j's population in 1980. Including stocks as a proportion of 1980 population also allows us to interpret $s_{j,1980}$ as a rough measure of likelihood a given migrant knows someone in the U.S. H_{jt} is the hurricane index for country j in year t.

The inclusion of year fixed effects (δ_t) accounts for time-varying changes in the overall ability of foreigners to migrate to the United States. Common issues such as changing demand in the U.S. economy and back-logs in the immigration system that are not country-specific are important components of δ_t. Country fixed effects (η_j) control for fixed factors that affect how likely denizens of country j are to migrate to the U.S., such as distance. They also absorb the main effect of $s_{j,1980}$. We also allow for differential country-specific linear time trends with the inclusion of $\phi_j t$, which account for long-run linear trends in migration from country

[25] Individuals who are captured in the DHS non-immigrant data may enter legally and then later overstay their visas, becoming undocumented.

j to the U.S. Standard errors are clustered at the country level .

A key coefficient of interest is β_1, the average effect of hurricanes across all origin countries. In addition, we are interested in γ_1 and γ_2. γ_1 is the effect of hurricanes on migration in countries with zero prior migrant stock. γ_2 captures the heterogeneity in the effect of hurricanes on migration with respect to a country's prior migrant stock. As discussed in Section 3.1, the signs of all these coefficients are theoretically ambiguous, motivating our empirical analysis.

3.4.2 Results

In the Online Appendix, we first establish that our hurricane index captures events that create economically relevant losses in potential sending countries. In the context of our theoretical framework from Section 3.1, we interpret these losses as an increase in the return to migration to the U.S. by generating asset losses, personal harm, and longer-run declines in economic growth. We focus here on our primary results, with m_{jt}—immigrant inflows from country j in year t as a proportion of country j's 1980 population—as the outcome of interest. As described in the Online Appendix, m_{jt} is created using access to confidential data from the U.S. Census Bureau. These data allow us to create accurate counts of immigrant inflows to the U.S., even for small countries that often go overlooked in such studies.

3.4.2.1 Primary Results on Migration

Table 3.3 presents the results of estimating Equations (3.4.1) and (3.4.2) with m_{jt} as the outcome. Column 1 of Panel A demonstrates that, on the whole, hurricanes induce positive levels of migration across our sample of 159 countries ($\beta_1 > 0$, statistically significantly different from zero at the 5 percent level). Column 2 illustrates that this effect operates largely through the stock channel: $\gamma_2 > 0$ (statistically significantly different from zero at the 5 percent level), suggesting that the ability of sending-country denizens to use migration as an ex-post response to hurricanes relies heavily on the presence an established network. This indicates a potentially crucial role for family reunification and other forms of sponsorship from within the U.S. in response to natural disasters abroad, motivating further investigation along these margins below.

We further split m_{jt} into separate age bins to investigate the characteristics of these hurricane-induced migrants. Table 3.4 shows that the youngest migrants—aged 0 to 12—as well as prime-aged migrants—aged 18 to 44—account for the majority of the effect seen in Table 3.3. Qualitatively, this aligns with the notion that working-aged adults and their children are most likely to respond to the combined impetus of an income shock and the pre-existence of a migration network.

The average effect of hurricanes in the first column implies that a one-standard-deviation hurricane (0.054) would increase migration to the U.S. by 0.00022 as a share of the home-country population (22 individuals per 100,000, or 0.022 percent). This is a substantial effect compared to sample statistics of the annual migration rate (Table 3.2), amounting to 11.8 percent of the sample mean (0.00183), 7.3 percent of the standard deviation (0.00296), and 10.8 percent of the inter-quartile range (0.002).

Column 2 of Table 3.3 reveals that the effect of hurricanes is magnified among origin countries with larger pre-existing stocks of U.S. immigrants. In this regression, the main effect of hurricanes is negative (but not statistically significantly different from zero at conventional levels). The effect of hurricanes on migration becomes positive for countries with a migrant stock in the U.S. (as share of 1980 population) of at least 0.86 percent—roughly the 70th percentile across countries. For a country at the 90th percentile of the prior migrant stock (6.1 percent of origin population), a one-standard-deviation increase in our measure of hurricane affectedness causes an inflow amounting to 0.033 percent of the origin population. This is also a substantial effect compared to sample statistics of the annual migration rate. It is 18.2 percent of the sample mean, 11.2 percent of the standard deviation, and 16.7 percent of the inter-quartile range.

3.4.2.2 Citizenship Status of Stock

To begin exploring how these networks operate, we exploit the fact that citizenship of respondents was recorded in the 1980 Long Form Census. We thus examine how the citizenship status of the 1980 stock affects the response to hurricanes. Differences in the ability of citizens versus non-citizens in promoting immigration allow us to roughly distinguish between different types of migrant network benefits. While both citizens and non-citizens can provide informational, financial, or psychic benefits, prior migrants who are citizens have the greatest ability to sponsor relatives for legal immigration (legally enshrined in the 1965 Amendments to the Immigration and Nationality Act.) For example, in 2004, 42.9 percent of the 946,142 legal immigrants admitted to the U.S. were able to bypass numerical quotas because they were immediate relatives of U.S. citizens. Another 12 percent were subject to numerical limitations, but also gained entry due to family sponsorship by a U.S. citizen (Department of Homeland Security 2006).[26] Thus, in the specification

[26]Note that these "admissions" include new arrivals and changes of status.

$$m_{jt} = \pi_0 + \pi_1 H_{jt} + \pi_2(H_{jt} \times s^{\text{citizen}}_{j,1980}) + \pi_3(H_{jt} \times s^{\text{non-cit}}_{j,1980}) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{3.4.3}$$

we expect $\pi_2 > \pi_3$. Table 3.5 shows evidence for this differential effect: only the interaction term on the U.S.-citizen portion of the migrant stock has a positive and statistically significant coefficient. This motivates a deeper look into how different classes of legal entrants respond to natural disasters.

3.4.2.3 DHS Results

For this purpose, we turn to data from the DHS *Yearbook of Immigration Statistics*, and the former Immigration and Naturalization Service's annual *Statistical Yearbook of the Immigration and Naturalization Service*, which allow us to separately examine entries of legal permanent residents (LPR) and legal non-immigrants—those who are only granted temporary visas. This generates two new outcome variables, m^{DHS}_{jt} where $DHS = \{LPR, \text{non-imm}\}$. Our specification remains the largely the same, with one exception. The DHS data does not allow us to distinguish between new entries and changes of status. Well-known back-logs in the immigration processing system can therefore create lag between shocks in sending countries and the enumeration of a migrants who gain LPR status if they enter as temporary residents first. In 2013, for example, 54 percent of family-based immigrants adjusted status from temporary to LPR compared to 46 percent who actually represented new entries (Kandel, 2016). We therefore increase the lag order in our specification by taking a simple average of H_{jt} and $H_{j,t-1}$, which we denote $H_{j,t,t-1}$. Our modified specifications become:

$$m^{DHS}_{jt} = \beta_0 + \beta_1 H_{j,t,t-1} + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{3.4.4}$$

$$m^{DHS}_{jt} = \gamma_0 + \gamma_1 H_{j,t,t-1} + \gamma_2(H_{j,t,t-1} \times s_{j,1980}) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{3.4.5}$$

The results from these models are presented in Table 3.6, where Panel A presents the results using our restricted-access estimates of migration inflows for comparison.[27]

There is a robust, positive effect of the stock interaction term on legal migration: γ_2 is estimated to be positive for both immigrant and non-immigrant entries.

In the row titled "Prop. of Census Inflows" of each panel, we calculate the proportion of inflows implied by the second column, produced by restricted-access migration counts m_{jt} in Panel A, that can be explained by inflows reflected in Panels B and C, produced by data from

[27] The set of countries has been restricted to be the same across all estimated specifications. We lose three countries to lack of data availability from the DHS.

the DHS (m_{jt}^{DHS}). This is done by obtaining predicted values from Equation (3.4.5), then multiplying by 1980 country population and summing over these fitted values to produce aggregate inflow estimates implied for each outcome. We then divide these aggregate inflow estimates by the result of the same calculation from the second column of Panel A.

This calculation reveals that DHS-recorded entries immediately following hurricanes are larger than those that are enumerated in later Census data. LPR hurricane-driven entries account for 2.47 times the corresponding number of entries in the Census data. We interpret this ratio of 2.47 to indicate that the hurricane-responsive migration effects that we observe in the Census data can be fully explained by entries of legal permanent residents. The fact that this ratio exceeds unity should not be cause for concern, because individuals who enter as LPRs immediately following hurricanes could fail to appear in later Census data for a number of reasons. Most prominently, large shares of immigrants to the U.S. do return to their home countries (or remigrate to third countries). Jasso and Rosenzweig (1982) estimate that up to 50 percent of the 1971 U.S. immigrant cohort had remigrated by January 1979.[28] On top of this, there is simple mortality, which we would expect to lead roughly one in ten LPR entries seen in the DHS results to not appear in the Census data from 2000 and after.[29] Another contributing but smaller factor is undercount of minorities in the Census.

The combination of these factors can bring the initial ratio of 2.47 very close to 1. If we then consider statistical noise and measurement error (e.g., incorrect reported years of entry or country of origin), the magnitude of the migration impacts on LPR entries seen in the DHS data are quite consistent with magnitudes in regressions using Census data. We conclude that the effects found in Table 3.3 and the second column of Table 3.6 are not in contradiction with one another, particularly when one considers the that these point estimates are each accompanied by 95 percent confidence intervals.

It is also important to keep in mind that the immediate impact of hurricanes on non-immigrant (mainly tourist and business visa) entries (Panel C of Table 3.6) is 50.51 times greater than the magnitudes seen in the regressions using Census data. We interpret this very high ratio as implying that the vast majority of the individuals (essentially all, to a first approximation) who enter the U.S. on non-immigrant visas in the wake of hurricanes stay only temporarily in the U.S., eventually returning to their home countries (or perhaps to going to third countries).

The detail of the DHS data allows us to further probe some of the mechanisms implied by our results thus far. In particular, the citizenship results from Table 3.5, the large response

[28]Similarly high return or remigration rates from the U.S. have been estimated in earlier decades of the 20th century as well (Chiswick and Hatton, 2003; Bandiera et al., 2013).

[29]Assuming an age-adjusted annual mortality rate of 1 percent, and immigrants entering the U.S. relatively uniformly over the 25-year period of analysis.

of legal, permanent inflows from Table 3.6, and the realities of the U.S. immigration system described in Section 3.2 suggest that family sponsorship may play a crucial role in allowing immigration to serve as an ex-post response to natural disaster shocks in sending countries. Table 3.7 suggests that this is the case. More than a third of the network interaction effect detected for LPRs in Table 3.6 can be traced to parents, spouses, or children of U.S. citizens—classes of immigrants who are not subject to numerical limitations. We further find that among immigrants who are subject to numerical limitations ("Capped Categories"), the network effect is especially salient for family-sponsored entrants.[30] Meanwhile, the effects of hurricanes on categories of entry that should not be affected by hurricanes in sending countries, such as employer-sponsored immigrants or diversity lottery winners, do not show the same heterogeneity with respect to migrant stocks.

3.4.3 Robustness and Mechanisms

The findings presented in Section 3.4 are consistent with immigrant stocks reducing the fixed cost of migration, allowing for a greater migratory response to hurricanes from source countries. There is, however, a concern of interpretation: the migrant stock could simply be correlated with omitted variables that are responsible for this observed heterogeneity. To gauge the robustness of our network-driven interpretation of the results to omitted variable concerns, we estimate regressions with the following specification

$$m_{jt} = \gamma_0 + \gamma_1 H_{jt} + \gamma_2 (H_{jt} \times s_{j,1980}) + \gamma_c (H_{jt} \times c_j) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{3.4.6}$$

This estimating equation is a modifies of our main specification, (3.4.2), by adding an additional set of interaction terms with time-invariant control variables c_j.[31]

Control variables c_j include a range of potential omitted variables. For example, $s_{j,1980}$ may proxy for sending country incomes (log real 1980 GDP per capita). Countries with higher incomes may be expected to both have higher $s_{j,1980}$ and more responsiveness to hurricanes if income makes credit constraints less binding for paying migration fixed costs. Financial development, measured by domestic credit as a proportion of GDP, may play a

[30] Note that these data are only available starting in 1992.

[31] We also include interaction terms with $c_j^{missing}$, dummy variables that account for some of these variables being unavailable for certain countries. When a variable is missing for a certain country, $c_j^{missing} = 1$ (and is 0 otherwise). When $c_j^{missing} = 1$, we replace the missing value of c_j with 0. The coefficient on the interaction term with $c_j^{missing}$ then represents heterogeneity in the responsiveness to hurricanes among all countries for which that variable is missing. Note the vector of main effects are not included in the regression because they are absorbed by the country fixed effects.

similar role. Migrant stocks may also proxy for distance to the U.S., with closer countries having both a higher $s_{j,1980}$ and lower migration fixed costs. We may expect that immigrant communities that are more concentrated geographically (say in migrant enclaves) may be better able to facilitate new immigration, perhaps due to closer social network connections. We thus include a measure of within-U.S. geographic concentration of immigrant stocks in 1980. Larger countries, either in population or area, may naturally offer more opportunities for internal migration, thus creating lower $s_{j,1980}$'s and lower responsiveness to hurricanes. Similarly, countries that have more alternate international migration destinations, such as those connected to popular destinations in Europe, may feature lower stocks and lower responsiveness, so we utilize a measure of 1990 immigrant stocks in non-U.S. destinations as a control variable.

The Online Appendix details the construction of each of these variables. Here, we focus on Table 3.8, which displays the results of estimating Equation (3.4.6) with each individual control variable as well as with the complete set. The estimated coefficient $\hat{\gamma}_2$ remains remarkably stable, and statistically significant, in each regression. There appears to be a robust effect of the stock of immigrants itself, as opposed to the many factors it may additionally proxy for. Additional results in the Online Appendix demonstrates that this robustness also applies to the results from Table 3.5, regarding the citizenship of the 1980 proportional stock.

The Online Appendix also conducts placebo exercises, demonstrating that future values of the hurricane index do not predict migration in prior years. Event-study specifications further confirm these placebo exercise results and validate our choice to employ either 0 or 1 lag in the estimating equations above.

3.4.4 Migration demand vs. migration "supply"

An important question is whether hurricane-induced migration should be thought of as occurring on the "demand" or "supply" sides of the migration "market". We argue that our findings should be interpreted as primarily a demand-side phenomenon: hurricanes induce greater desire to migrate (with a greater likelihood of actual migration from countries with larger stocks of prior migrants). But it is also important to ask whether the "supply" of migration possibilities may also similarly respond to hurricanes. One might imagine that the U.S. government could loosen immigration restrictions to make it possible for more people to migrate in response to hurricanes. If the hurricane-induced increase in the supply of immigration slots occurred differentially more when countries had larger migrant stocks in the U.S., this could provide an alternative (or complementary) explanation for our findings.

There are two mechanisms though which the U.S. government could increase the supply of migration slots: Temporary Protected Status (TPS) and Deferred Enforced Departure (DED).[32] These are temporary statuses granted by the executive branch of the federal government to nationals of select countries due to an adverse event (most commonly natural disasters, political unrest, or conflict) in their home country, and allow beneficiaries to live and work in the U.S. for a defined, limited period of time. While many entrants covered by these programs end up staying in the U.S. for multiple years, they are officially classified as non-immigrants. TPS and DED entries should therefore not affect our regression estimates for LPR entries in the DHS data.

It is also of note that in only one instance have either of these statuses been designated in response to a hurricane: TPS for Hondurans and Nicaraguans in the wake of Hurricane Mitch in 1998. Thus, to directly test whether increases in migration slots made available by TPS or DED are affecting our results, we reestimated Equations (3.4.2) and (3.4.5) dropping Honduras and Nicaragua from our estimation sample. Results in the Online Appendix shows that dropping these countries has a minimal effect on our estimated effects. This is true for LPR inflows (DHS), for which we shouldn't see any changes, but also for non-immigrant entries (DHS) and overall entries (using the U.S. Census RDC data).

In sum, then, there is no evidence that migration-supply-side responses are an important channel through which the immigration effects of hurricanes are operating. The hurricane-induced migration we document in this chapter appears to be a demand-side phenomenon.

3.4.5 The Role of Restricted-Access Census Data and of Small Countries

As noted in the Introduction, one of our key contributions is the use of restricted-access U.S. Census data. These data allow us to expand the sample of analysis to include small countries that are not included for confidentiality reasons in publicly available Census data (as discussed in the Online Appendix).

To illustrate the importance of these data, we estimate Equations (3.4.1) and (3.4.2) using publicly-available Census data in the Online Appendix. For both the average effect of hurricanes and the heterogeneity in this effect with respect to prior migrant stocks, coefficient estimates from publicly-available data are much smaller in magnitude compared to corresponding estimates presented in Table 3.3 (and are far from being statistically significantly different from zero at conventional levels).

The difference in results from analyses using restricted-access versus public-use Census data derives from the inclusion of small countries. Small countries added to the analysis provide key identifying variation. As small countries, they have higher hurricane indices on

[32]Prior to 1990, DED was called Extended Voluntary Departure (EVD). While similar in their practical aspects, the two programs are separate because they are implemented under different executive powers.

average: the index takes into account the share of land area covered by a hurricane, and so small countries can have larger hurricane indices. They also happen to have on average higher population shares of prior migrants in the US, but with substantial heterogeneity that also provides identifying variation. That is, they have significant "leverage" in our specifications. To aid in getting a picture of these small countries, the Online Appendix lists all countries in our sample, sorted by 1980 population, and provides the value of their mean hurricane index and population share of prior migrants in the U.S.

To directly explore the role of small countries in contributing to the estimated effects, we provide additional regression results in the Online Appendix. We show that small countries play a central role. There is no large or statistically significant effect of network-facilitated migration in response to hurricanes when we weight by population or restrict the sample to quartiles of countries above the bottom quartile of 1980 population. (That said, our main results are robust to weighting by log 1980 population).

Additionally, in the Online Appendix, we estimate our main regression specification when dropping progressively more smaller countries from the sample, starting with countries with the very smallest 1980 populations. The point estimate on the migrant stock interaction term does become smaller when more and more of the small countries are dropped from the regression. But the main results are not contingent on the presence of outliers in the sample, only disappearing when the smallest 20 countries (approximately) are dropped from the sample.[33]

Given that the key identifying variation is coming from small countries, one might raise concerns about external validity (whether empirical patterns identified off migration flows from the smallest countries would also be seen elsewhere). From a statistical standpoint, large countries could simply not be contributing to our estimated effects because we cannot split them into smaller administrative areas for analysis—we do not have data on migration from sub-national origin areas. A hurricane occurring in a large country (such as China, India, Mexico, or the Philippines) affecting a particular sub-national locality of a few thousand people may similarly generate migration to the U.S. (with heterogeneity with respect to prior migrant stocks from that locality) along the lines of what we see in our regressions, but this will not be detected by our estimation procedure because migration flows of the country as a whole will dwarf this response in the data.

[33]The U.S. Census does not allow us to specify exactly how many countries were excluded from each regression, for confidentiality reasons, so the number of dropped countries is approximate.

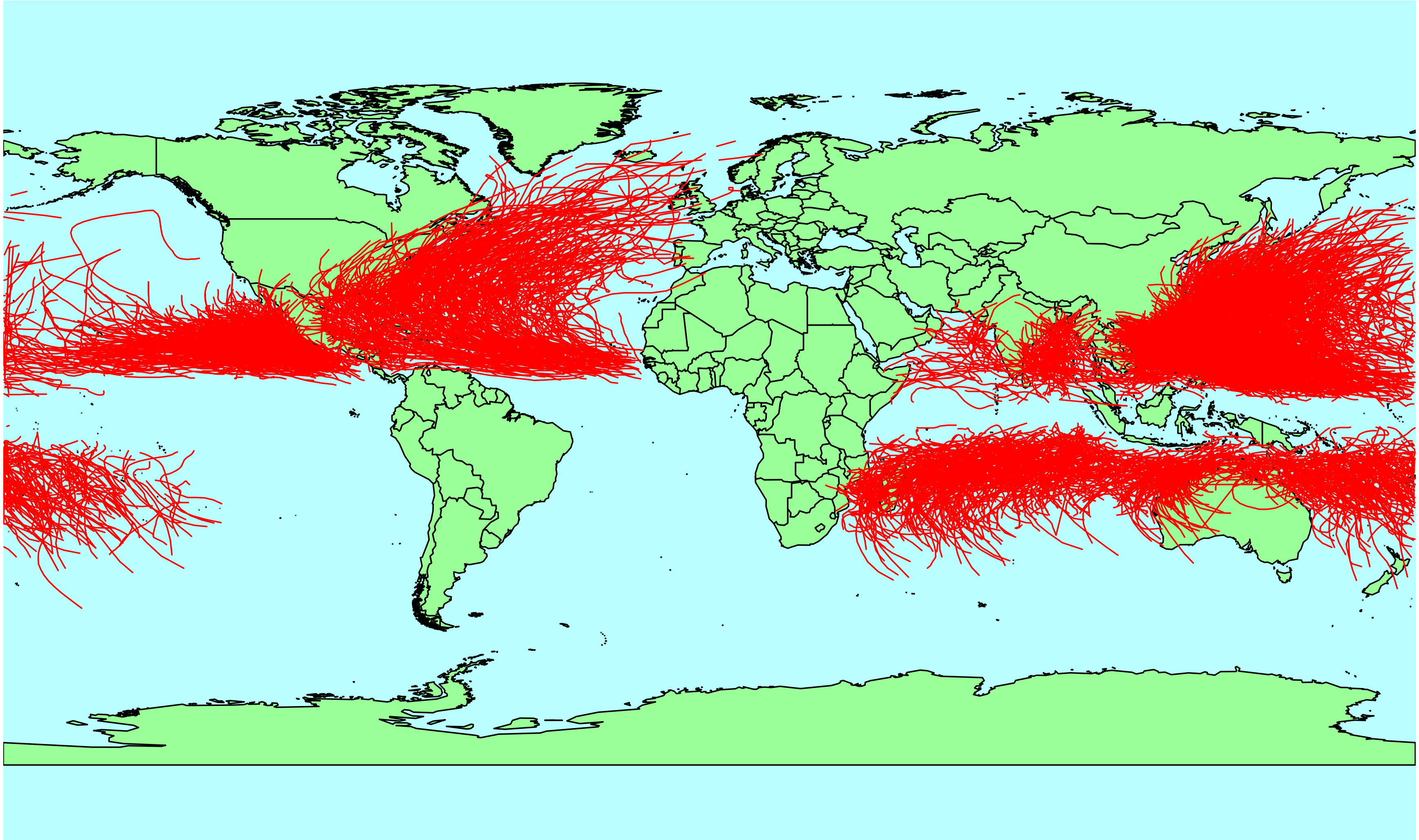
3.5 Conclusion

We examine how international migration responds to changes in the returns to migration, and how this response depends on the costs or barriers that migrants face in moving. We examine this question in the context of a quarter-century of migration to the U.S., the world's largest migration destination, from 159 origin locations worldwide. In our analysis, we exploit the occurrence of hurricanes, which exogenously increase the returns to migration by making origin areas less attractive, and ask whether the migration response to hurricanes depends on the size of prior migrant stocks from the same country. Our migration outcomes are unusually precise, measured either from from restricted-access U.S. Census data or actual legal immigration counts from U.S. government administrative data. We find that, on average, countries more affected by hurricanes see more migration to the U.S. This migration response is indeed larger (as a share of origin-country population) among countries with larger stocks of prior U.S. migrants. A key role played by previous migrant networks appears to be sponsoring relatives for legal immigration. There is also a substantial effect of hurricanes (and similar heterogeneity of effects with respect to migrant stocks) on legal entries via temporary or non-immigrant visas.

This study is among the first to test whether the immigration response to disasters in migrants' origin areas is larger when origin areas have larger stocks of prior migrants. We document an important role played by migrant networks: helping compatriots in the home country migrate themselves as a way of coping with negative shocks. We provide unique evidence that hurricane-induced flows of new migrants enter via legal, statutory immigration channels, and that there is an identifiable effect on permanent (not just temporary) migration.

These findings are of substantial policy interest. Immigration has long been one of the most contentious issues in the public realm, while at the same time shocks in migrant-source countries are pervasive. Of particular interest are shocks to economic and social conditions generated by climate change. The policy debate should be informed by a better understanding of how and when increasingly severe natural disasters in migrant-origin countries will actually lead to increased migration. Policy-makers in destination countries would benefit from an improved understanding of the determinants of migrant inflows that may result from such shocks.

Figure 3.1: Hurricane Best Tracks: 1980-2004



Source: Unisys Weather data, formerly available at http://weather.unisys.com/hurricane/. Raw data available upon request.

Table 3.1: List of Countries in Sample

Afghanistan	French Polynesia	Nigeria
Algeria	Gabon	Niue
Angola	Gambia	Oman
Anguilla	Ghana	Pakistan
Antigua & Barbuda	Grenada	Panama
Argentina	Guadeloupe	Papua New Guinea
Aruba	Guatemala	Paraguay
Australia	Guinea	Peru
Bahamas	Guinea-Bissau	Philippines
Bahrain	Guyana	Qatar
Bangladesh	Haiti	Reunion
Barbados	Honduras	Rwanda
Belize	Hong Kong	Samoa
Benin	India	Sao Tome & Principe
Bermuda	Indonesia	Saudi Arabia
Bhutan	Iran	Senegal
Bolivia	Iraq	Seychelles
Botswana	Israel	Sierra Leone
Brazil	Ivory Coast	Singapore
British Virgin Islands	Jamaica	Solomon Islands
Brunei	Japan	Somalia
Burkina Faso	Jordan	South Africa
Burma (Myanmar)	Kenya	South Korea
Burundi	Kiribati	Sri Lanka
Cambodia	Kuwait	St. Helena
Cameroon	Laos	St. Kitts-Nevis
Canada	Lebanon	St. Lucia
Cape Verde	Lesotho	St. Vincent & the Grenadines
Cayman Islands	Liberia	Sudan
Central African Republic	Libya	Suriname
Chad	Macau	Swaziland
Chile	Madagascar	Syria
China	Malawi	Taiwan
Colombia	Malaysia	Tanzania
Comoros	Maldives	Thailand
Congo	Mali	Togo
Cook Islands	Martinique	Tokelau
Costa Rica	Mauritania	Tonga
Cuba	Mauritius	Trinidad & Tobago
Cyprus	Mexico	Tunisia
Democratic Republic of Congo (Zaire)	Micronesia	Turkey
Djibouti	Mongolia	Turks & Caicos Islands
Dominica	Montserrat	Uganda
Dominican Republic	Morocco	United Arab Emirates
East Timor	Mozambique	Uruguay
Ecuador	Namibia	Vanuatu
Egypt	Nauru	Venezuela
El Salvador	Nepal	Vietnam
Equatorial Guinea	Netherlands Antilles	Wallis & Futuna Islands
Ethiopia	New Caledonia	Western Sahara
Falkland Islands	New Zealand	Yemen
Fiji	Nicaragua	Zambia
French Guiana	Niger	Zimbabwe

Notes: See Section 3.3 for details on sample selection.

Table 3.2: Summary Statistics

			Percentile						
	Mean	Std. Dev.	10	25	50	75	90	N	Source
Hurricane Index	0.00402	0.02373	0	0	0	0	0.00064	3,895	Unisys
Hurricane Index (if > 0)	0.02451	0.05417	0.00001	0.00014	0.00190	0.01903	0.07709	639	Unisys
1980 Population (thousands)	21,627	96,379	60	250	3,027	11,095	38,124	159	UN & Census IDB
As a Proportion of 1980 Population:									
Annual Migrants	0.00183	0.00296	0.00004	0.00010	0.00039	0.00210	0.00625	2,200	IPUMS[a]
Annual Immigrants	0.00133	0.00339	0.00001	0.00005	0.00020	0.00112	0.00418	2,582	DHS
Annual Non-Immigrant Entries	0.06127	0.22229	0.00019	0.00059	0.00413	0.02101	0.11062	2,214	DHS
Annual Immediate Family Immigrants	0.00051	0.00109	<0.00001	0.00002	0.00009	0.00046	0.00175	2,582	DHS
Annual Family-Sponsored Immigrants	0.00099	0.00246	0.00001	0.00003	0.00013	0.00078	0.00316	1,511	DHS
1980 Stock of Immigrants	0.01594	0.03100	0.00020	0.00037	0.00250	0.01573	0.06160	150[a]	IPUMS[a]
1980 Stock of Citizen Immigrants	0.00615	0.01471	0.00004	0.00009	0.00068	0.00469	0.02099	150[a]	IPUMS[a]
1980 Stock of Non-Citizen Immigrants	0.00979	0.01746	0.00012	0.00029	0.00165	0.01113	0.03320	150[a]	IPUMS[a]

Notes: See the Online Appendix for details on creation of hurricane index. The second row shows summary statistics for the hurricane index conditional on it being greater than zero. "Immediate Family" refers to parents, children, or spouses of U.S. citizens—these admissions are uncapped. "Family-Sponsored" immigrants are those whose admissions are capped, but who enter through family sponsorship.
Sources: DHS data obtained from electronic copies of the *Yearbook of Immigration Statistics* for 1996-2004 (Department of Homeland Security, 2004) and the *Statistical Yearboook of the Immigration and Naturalization Service* for prior to 1996 (Immigration and Naturalization Service, 1995). UN data obtained from the United Nations Population Division (United Nations, 2017). Census IDB data obtained from the Census Bureau's International Data Base (U.S. Census Bureau, 2).
[a]Statistics constructed using Census public-use microdata obtained from IPUMS-USA (Ruggles et al., 2019b) rather than RDC data to avoid confidentiality issues, which explains the loss in sample size. These are not the data used in regression model estimation.

Table 3.3: The Effect of Hurricanes on Migration, 1980-2004

	As a Prop. of 1980 Population	
	Migrants(t)	Migrants(t)
	(1)	(2)
Hurricane Index(t)	0.0040	-0.0010
	(0.0020)	(0.0010)
Hurricane Index(t) $\times$ 1980 Proportional Immigrant Stock		0.1163
		(0.0451)
Country-Years	3,900	3,900
R^2	0.4319	0.4409
Countries	159	159

Notes: Each column refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center.

Table 3.4: The Effect of Hurricanes on Migration by Age Group, 1980-2004

Panel A: Children	As a Prop. of 1980 Population			
Age Group of Migrants(t):	0 to 12	0 to 12	13 to 17	13 to 17
Hurricane Index(t)	0.0014	-0.0007	0.0006	0.0004
	(0.0007)	(0.0004)	(0.0005)	(0.0006)
Hurricane Index(t) $\times$ 1980 Proportional Immigrant Stock		0.0481		0.0057
		(0.0138)		(0.0076)
Country-Years	3,900	3,900	3,900	3,900
R^2	0.2293	0.2461	0.1794	0.1798
Countries	159	159	159	159
Panel B: Prime-Aged	As a Prop. of 1980 Population			
Age Group of Migrants(t):	18 to 24	18 to 24	25 to 44	25 to 44
Hurricane Index(t)	0.0009	-0.0005	0.0010	-0.0004
	(0.0007)	(0.0003)	(0.0007)	(0.0004)
Hurricane Index(t) $\times$ 1980 Proportional Immigrant Stock		0.0329		0.0306
		(0.0185)		(0.0150)
Country-Years	3,900	3,900	3,900	3,900
R^2	0.2953	0.3010	0.3123	0.3155
Countries	159	159	159	159
Panel C: Older	As a Prop. of 1980 Population			
Age Group of Migrants(t):	45 to 64	45 to 64	65 and older	65 and older
Hurricane Index(t)	0.0002	0.0006	-0.0001	-0.0005
	(0.0004)	(0.0006)	(0.0002)	(0.0004)
Hurricane Index(t) $\times$ 1980 Proportional Immigrant Stock		-0.0083		0.0072
		(0.0076)		(0.0047)
Country-Years	3,900	3,900	3,900	3,900
R^2	0.1884	0.1906	0.1359	0.1403
Countries	159	159	159	159

Notes: Each column within a panel refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center.

Table 3.5: The Effect of Hurricanes on Migration by Citizenship of Stock, 1980-2004

	As a Prop. of 1980 Population	
	Migrants(t)	Migrants(t)
Hurricane Index(t)	-0.0010	-0.0005
	(0.0010)	(0.0009)
Hurricane Index(t)	0.1163	
$\times$ 1980 Proportional Immigrant Stock	(0.0451)	
Hurricane Index(t)		0.4044
$\times$ 1980 Proportional Immigrant Citizen Stock		(0.2245)
Hurricane Index(t)		-0.1444
$\times$ 1980 Proportional Immigrant Non-Citizen Stock		(0.1661)
Country-Years	3,900	3,900
R^2	0.4409	0.4429
Countries	159	159
p-value: Equal Interaction Effect		0.1540

Notes : Each column refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.2) and (3.4.3). "Hurricane Index" refers to the hurricane index for a given country in year t. "Migrants", "1980 Proportional Stock," "1980 Proportional Citizen Immigrant Stock," and "1980 Proportinal Non-Citizen Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center.

Table 3.6: The Effect of Hurricanes on Migration—Comparing Census to DHS Data

Panel A: Census, 1980–2004	As a Prop. of 1980 Population	
	Migrants(t)	Migrants(t)
Hurricane Index($t, t-1$)	0.0046	-0.0012
	(0.0021)	(0.0016)
Hurricane Index($t, t-1$) × 1980 Proportional Immigrant Stock		0.1235
		(0.0427)
Country-Years	3,800	3,800
Countries	156	156
R^2	0.4426	0.4475
Prop. of Census Inflows		1
Panel B: DHS, 1982–2004	As a Prop. of 1980 Population	
	LPR(t)	LPR(t)
Hurricane Index($t, t-1$)	0.0023	-0.0035
	(0.0040)	(0.0039)
Hurricane Index($t, t-1$) × 1980 Proportional Immigrant Stock		0.1266
		(0.0402)
Country-Years	2,600	2,600
Countries	156	156
R^2	0.2954	0.2966
Prop. of Census Inflows		2.47
Panel C: DHS, 1983–2004	As a Prop. of 1980 Population	
	Non-Immi(t)	Non-Immi(t)
Hurricane Index($t, t-1$)	0.2193	-0.0627
	(0.0788)	(0.0689)
Hurricane Index($t, t-1$) × 1980 Proportional Immigrant Stock		5.7883
		(2.3536)
Country-Years	2,200	2,200
Countries	156	156
R^2	0.4485	0.4495
Prop. of Census Inflows		50.51

Notes: Each column within a panel refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. "LPR(t)" refers to the number of individuals granted lawful permanent resident status from a given country in year t. "Non-Immi(t)" refers to the number of individuals that entered the U.S., enumerated by the DHS, from a given country in year t that were not granted lawful permant residence (e.g., tourists and those on student visas). "Hurricane Index($t, t-1$)" refers to the average of a hurricane index for a given country across years t and $t-1$. "Prop. of Census Inflows" calculated by multiplying the estimated coefficients by each country's specific "Hurricane Index($t, t-1$)" and 1980 Proportional Immigrant Stock, summing them across country-years, then dividing by the same calculation made using the results from the second "Census" column.
Sources: Outcomes in Panels B and C obtained from electronic copies of the *Yearbook of Immigration Statistics* for 1996-2004 (Department of Homeland Security, 2004) and the *Statistical Yearboook of the Immigration and Naturalization Service* for prior to 1996 (Immigration and Naturalization Service, 1995).

Table 3.7: The Effect of Hurricanes on LPR Entries by Category—DHS Data

Panel A: Uncapped, Immediate Relatives of U.S. Citizens	Legal, Permanent Immigration by Category As a Prop. of 1980 Population			
	Parents	Spouses	Children	Parents, Spouses, and Children
Hurricane Index$(t, t-1)$	-0.0002	-0.0007	-0.0007	-0.0014
	(0.0003)	(0.0006)	(0.0008)	(0.0014)
Hurricane Index$(t, t-1)$ $\times$ 1980 Proportional Immigrant Stock	0.0130	0.0150	0.0234	0.0457
	(0.0069)	(0.0065)	(0.0167)	(0.0189)
Country-Years	2,600	2,600	2,600	2,600
R^2	0.1609	0.1970	0.1200	0.1435
Years	1982 to 2004	1982 to 2004	1982 to 2004	1982 to 2004
Countries	156	156	156	156
Panel B: Capped Categories	Legal, Permanent Immigration by Category As a Prop. of 1980 Population			
	Refugee	Family Sponsored	Employer Sponsored	Diversity Lottery
Hurricane Index$(t, t-1)$	0.0003	-0.0019	0.0001	-0.0001
	(0.0002)	(0.0029)	(0.0002)	(0.0003)
Hurricane Index$(t, t-1)$ $\times$ 1980 Proportional Immigrant Stock	-0.0013	0.1630	-0.0225	-0.0005
	(0.0027)	(0.0660)	(0.0054)	(0.0044)
Country-Years	2,600	1,500	1,500	1,500
R^2	0.3309	0.1667	0.4223	0.4218
Years	1982 to 2004	1992 to 2004	1992 to 2004	1992 to 2004
Countries	156	156	156	156

Notes: Each column within a panel refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). "Hurricane Index$(t, t-1)$" refers to the average of a hurricane index for a given country across years t and $t-1$. "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center.
Sources: Outcomes obtained from electronic copies of the *Yearbook of Immigration Statistics* for 1996-2004 (Department of Homeland Security, 2004) and the *Statistical Yearboook of the Immigration and Naturalization Service* for prior to 1996 (Immigration and Naturalization Service, 1995).

Table 3.8: Robustness

	Outcome for all columns: Migrants(t) as a Prop. of 1980 Population								
Hurricane Index(t)	-0.0018	0.0124	0.0094	-0.0045	-0.0034	-0.0010	-0.0028	-0.0012	0.0541
	(0.0010)	(0.0175)	(0.0064)	(0.0030)	(0.0034)	(0.0010)	(0.0021)	(0.0022)	(0.0334)
Hurricane Index(t) $\times$	0.1199	0.1213	0.1001	0.1249	0.1175	0.1159	0.1291	0.1032	0.1094
1980 Proportional Immigrant Stock	(0.0453)	(0.0494)	(0.0369)	(0.0462)	(0.0428)	(0.0451)	(0.0483)	(0.0399)	(0.0331)
Hurricane Index(t) $\times$	0.3709								-0.0122
Immigrant Concentration Index	(0.2488)								(0.0088)
Hurricane Index(t) $\times$		-0.0015							-0.0049
log(1980 Real GDP Per Capita)		(0.0020)							(0.0033)
Hurricane Index(t) $\times$			-0.0802						-0.1165
log(1980 Population)			(0.0517)						(0.0717)
Hurricane Index(t) $\times$				-0.0474					-0.2181
[1970s Remittances as Prop. of GDP]				(0.0700)					(0.1040)
Hurricane Index(t) $\times$				0.0040					0.0034
$\mathbb{1}$[Missing: Remittances]				(0.0030)					(0.0043)
Hurricane Index(t) $\times$					-0.0015				0.0026
[1970s Dom. Credit as Prop. of GDP]					(0.0061)				(0.0088)
Hurricane Index(t) $\times$					0.0038				0.0126
$\mathbb{1}$[Missing: Dom. Credit]					(0.0038)				(0.0064)
Hurricane Index(t) $\times$						-0.0014			0.0102
[Land Area (mil. sq. km)]						(0.0041)			(0.0080)
Hurricane Index(t) $\times$							0.1952		0.2579
[Distance to U.S. (mil. km)]							(0.1618)		(0.1686)
Hurricane Index(t) $\times$								0.0055	-0.0117
[1990 Prop. non-U.S. Emigrant Stock]								(0.0068)	(0.0079)
Hurricane Index(t) $\times$								0.0004	-0.0122
$\mathbb{1}$[Missing: non-U.S. Emigrant Stock]								(0.0029)	(0.0056)
Country-Years	3,900	3,900	3,900	3,900	3,900	3,900	3,900	3,900	3,900
R^2	0.4412	0.4413	0.4422	0.4423	0.4424	0.4409	0.4412	0.4412	0.4495
Countries	159	159	159	159	159	159	159	159	159

Notes: This table is intended for comparison with Column 2 of Table 3.3. Each column refers to a different OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equation (3.4.6). 1970s Domestic Credit as Prop. of GDP and 1970s Remittances as a Prop. of GDP divide averages of non-missing data of Domestic Credit and Remittances from 1970 through 1979 by 1980 GDP. "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center.

APPENDICES

APPENDIX A

Appendix to Chapter 1

A.1 Case Studies: Details

A.1.1 Synthetic Control Method: an Overview

The Synthetic Control Method (SCM) is a data driven approach for constructing control groups when treatments occur in specific geographies. By incorporating elements of matching, it relaxes the traditional difference-in-differences assumption that confounding factors must be time invariant in order to estimate unbiased treatment effects. It allows common shocks across regions to affect different regions differently. In short, SCM utilizes the information at its disposal, including pre-event outcomes, to generate a control group most similar to the treated unit in its pre-treatment behavior. Specifically, given a set of untreated, "donor" geographies, D, a set of M control variables for each of these donors, $X_{M\times 1}$, and a year of treatment t^*, SCM solves

$$W^*(V) = \arg\min_{\boldsymbol{W}} ||X_{\text{treated},m} - \mathbf{X}'W||_{v_m}$$

where $\mathbf{X}_{D\times M}$ stacks the vectors of control variables over donors. $W^*(V)_{(D\times 1)}$ is a set of weights for donor units that minimizes the distance between the control variables in the treated and donor units. Like with matching, X can and should include pre-treatment outcomes, but should not include the full set of pre-treatment outcomes (Kaul et al., 2015). Denoting y_t as the outcome of interest and y_{Dt} as the $(D \times 1)$ vector of outcomes in the donor units at time t,

$$W^*(V)' \times y_{Dt} \equiv y_{t,\text{Synthetic}}$$

gives the synthetic control value at each time t. In order to trace out year-over-year dynamic responses to shocks, the parameters of interest for a given outcome y take the form

$$\delta_{yt} \equiv (y_{\text{treated},t} - y_{\text{synthetic},t}) - (y_{\text{treated},t^*-1} - y_{\text{synthetic},t^*-1}) \tag{A.1}$$

For $t < t^* - 1$, δ_{yt} serves as a test for parallel trends and for $t \geq t^*$, δ_{yt} is akin to an event study treatment effect estimate. At $t^* - 1$, the estimate is normalized to 0.

A.1.2 Choosing Predictors Using Cross-Validation

Because of the wide range of covariates that can be used in a given SCM model, the methodology can be susceptible to cherry-picking in terms of which variables go into X (Ferman et al., 2016). In order to circumvent these concerns, I choose each synthetic control model using a cross-validation procedure proposed by Dube and Zipperer (2015). The procedure is essentially a pseudo-out-of-sample test of how well a given candidate set of predictors does in forecasting *post-treatment* outcomes for donor (untreated) geographies. It creates these forecasts by constructing a synthetic control for each of the donor units. That is, in order to select X, the following procedure is implemented:

1. Choose a candidate X
2. Run the synthetic control method for each donor unit d and each event e
3. Choose the X^* such that $X^* = \arg\min_X \sum_e \sum_{d=1}^{D} \sum_{t \geq t^*} \left(Y_{dt} - \sum_{q \neq d} w_q^*(X) Y_{qt}(X) \right)^2$

where the term $\sum_{q \neq d} w_q^*(X) Y_{qt}(X)$ represents the synthetic control constructed for donor unit d. The treated units are excluded from this entire process. Additionally, I only run this process on the reduced form outcome establishments per $t^* - 1$ worker. The first stages for Phoenix and Arizona are estimated using the chosen X^* from the cross-validation procedure run on the reduced form.

The chosen X^* for SCM estimation were: sector shares, the under-40 year-old share, the self-employed share, and the college share, all in $t^* - 1$. I include log employment in the year before the event and five-year pre-treatment averages of the outcome, establishments per time $t^* - 1$ worker, in all models. Note that the under-40 year-old share, the self-employed share, and the college share are only able to be included for Phoenix and Arizona. The chosen donor units and their weights can be seen below, in Table A.1.

Table A.1: Synthetic Control Weights

Geography	Weight
Panel A: Phoenix MSA, Immigrants per Initial Worker*	
Cape Coral-Fort Myers, FL	0.280
Daphne-Fairhope-Foley, AL	0.159
Santa Fe, NM	0.105
Merced, CA	0.057
Madera, CA	0.023
Port St. Lucie, FL	0.010
Salinas, CA	0.010
Panel B: Phoenix MSA, Establishments per Initial Worker	
Austin-Round Rock-Georgetown, TX	0.595
Salisbury, MD-DE	0.094
Cape Coral-Fort Myers, FL	0.089
Port St. Lucie, FL	0.066
Ocala, FL	0.053
San Diego-Chula Vista-Carlsbad, CA	0.047
Riverside-San Bernardino-Ontario, CA	0.036
Barnstable Town, MA	0.021
Panel C: Miami MSA, Establishments per Initial Worker	
San Diego-Chula Vista-Carlsbad, CA	0.381
San Francisco-Oakland-Berkeley, CA	0.205
New York-Newark-Jersey City, NY-NJ-PA	0.145
Houston-The Woodlands-Sugar Land, TX	0.132
Sacramento-Roseville-Folsom, CA	0.084
Ocean City, NJ	0.052

*The Phoenix MSA has several other donor units with weights of less than 0.01 that are not listed here.

A.1.3 Inference

In order to conduct inference and construct confidence intervals for the pooled SCM case studies, I harmonize the analysis to a set of 214 metropolitan statistical areas delineated by the Office of Management and Budget in 2013. Note that all Arizona cities are excluded from the SCM estimates for Miami and all other Arizona cities and Miami are excluded from the SCM estimates for Phoenix. I then once again follow procedures laid out in Dube and Zipperer (2015).

First, for each event e, I calculate δ_t^{RF} for the treated unit as well as a "placebo" version of this estimator for each donor unit as if the donor unit was treated in the same year, $\delta_{t,\text{donor}}^{RF}$.[1] I then divide all of these estimates by the δ^{1S} from the treated unit. The result is an actual estimate of the effect of immigrant inflows on firm presence in the treated unit, β_t^e as well as a set of placebo estimates, $\beta_{t,\text{donor}}^e$ that are used to construct an empirical distribution for inference.

The first step in the inference process involves ranking the estimates of β_t^e among the 212 estimates of $\beta_{t,\text{donor}}^e$. The estimated rank percentile is this rank divided by 213. The test statistic for combined inference is then simply the average of these percentile ranks. The critical values needed to conduct inference are given in Table A1 of Dube and Zipperer (2015) for $N = 2$. Inverting this hypothesis test to generate confidence intervals, like those shown in Figure 1.2, simply involves finding the mean of the empirical distributions at these critical values, $\frac{1}{E}\sum_e \beta_{t,\text{donor}}^e(crit)$. For example, the lower bound of the 95 percent critical value is the estimate β_t plus this mean at the 2.5 percent critical value (0.111) from the aforementioned Table A1 of Dube and Zipperer (2015).

A.1.4 Results from State-Level Case Studies

As described in Section 1.3.1, one way to pool the Mariel and Arizona cases is to focus on Arizona's largest city, Phoenix, and Miami. The other alternative is to consider Arizona and Florida together. This section conducts this alternative for completeness.

Interestingly, Florida shows a stronger relationship in the state case study than Miami did in the individual city case study. This somewhat aligns with previous evidence stemming from the Mariel Boatlift case study. Card (1990), for example, documents that Florida as a whole experienced less population decline than did Miami in the years following the Boatlift. Indeed, most previous Mariel Boatlift studies find that the absorption of Marielitos was only accompanied by a short run increase in population, wholly consistent with net emigration to (or less inmigration to Miami from) the rest of Florida (see, e.g., Figure 2 in Peri and

[1] I exclude the treated unit when calculating the placebo estimates.

Yasenov, 2015). These relative expansions in the rest of the state could mean that it was really the state of Florida that more completely absorbed the Marielitos rather than Miami itself (see here for anecdotal evidence that the whole state was affected). This, in turn, would lead to larger establishment presence in the rest of the state. Relative to the rest of this chapter, this provides moderate caution regarding the magnitudes of the effects found in Section 1.3. Given the possibility population displacement, particularly when immigration shocks are large, these estimates should be viewed as a lower bound. Smaller shocks that are most likely used for identification from the instrumental variable are less likely to create these issues (see, e.g., Card and DiNardo, 2000).

Table A.2: Synthetic Control Weights

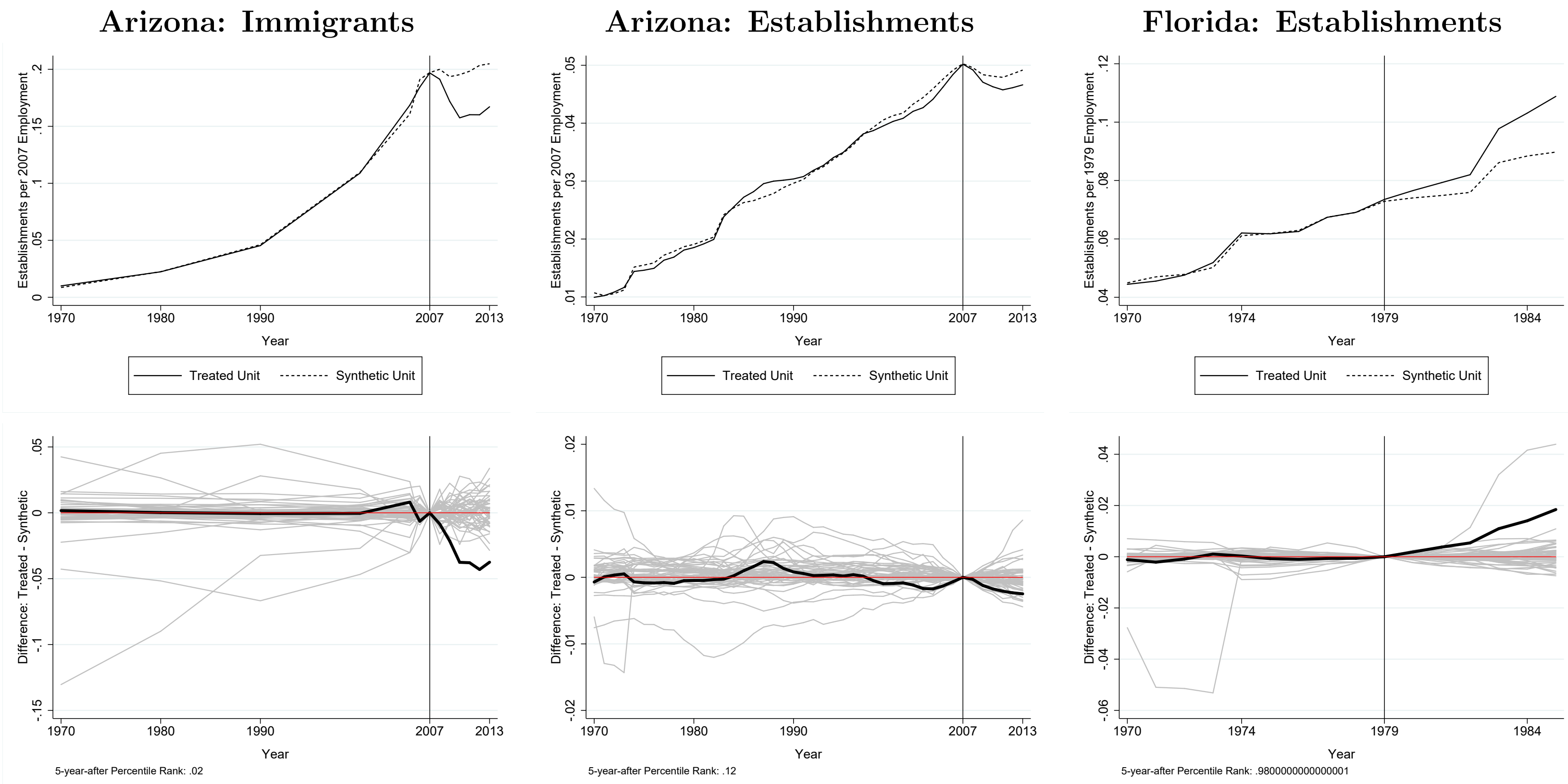
Geography	Weight
Panel A: Arizona, Immigrants per Initial Worker	
Alaska	0.066
Delaware	0.025
Florida	0.05
Georgia	0.111
Maryland	0.043
Nevada	0.566
Utah	0.138
Panel B: Arizona, Establishments per Initial Worker	
Alaska	0.088
Maryland	0.155
Nevada	0.613
Tennessee	0.01
Utah	0.134
Panel C: Florida, Establishments per Initial Worker	
Alaska	0.035
Arizona	0.057
Hawaii	0.230
Maine	0.002
Maryland	0.086
Nebraska	0.008
Rhode Island	0.272
Wyoming	0.300

Table A.3: Synthetic Control First Stage Δ_I^{DD}

Case	Treatment Start (t^*)	Variation Source	Δ_I^{DD}	Empirical p-value
Arizona	2008	Arizona LAWA	-0.043	0.038
Florida	1980	Mariel Boatlift	0.036	—

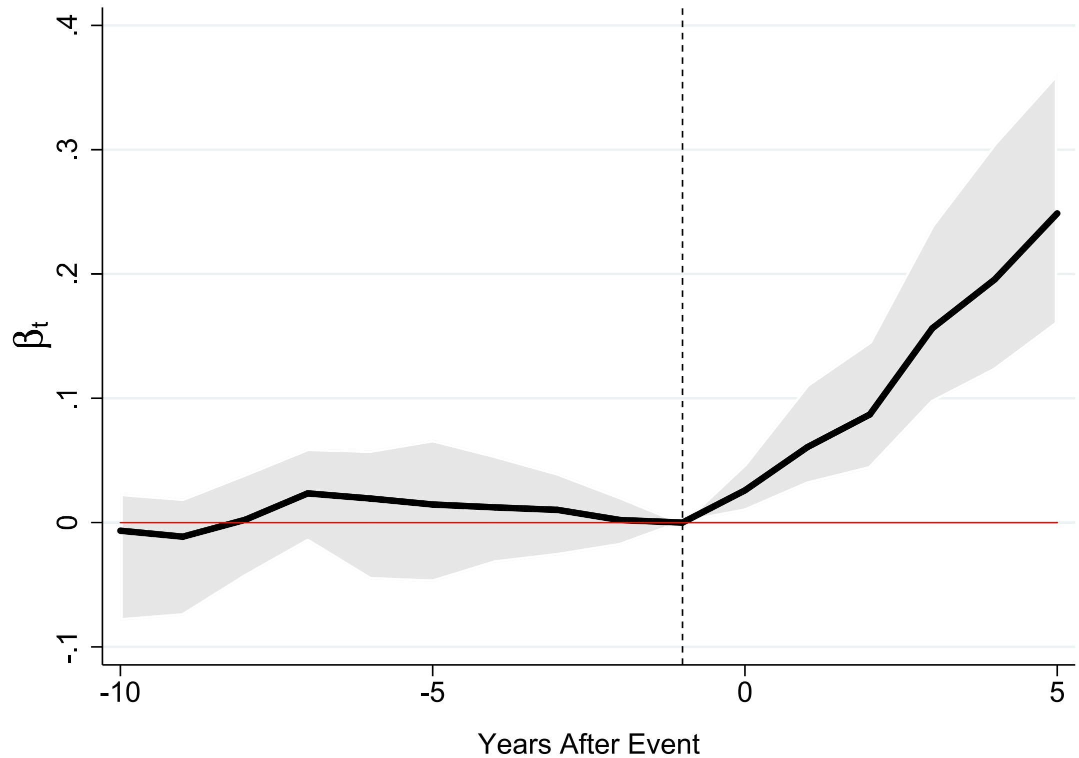
Source: First row—author's calculations from IPUMS-USA and County Business Patterns. Second row—Card (1990) p. 248 and author's calculations from Current Population Survey (CPS) Merged Outgoing Rotation Group accessed using IPUMS-CPS; assumes all 125,000 Marielitos stayed in Florida in order to be conservative.

Figure A.1: Individual Synthetic Control Results—Measures per Initial ($t^* - 1$) Worker



Source: Author's calculations from IPUMS-USA and County Business Patterns.

Figure A.2: Pooled SCM Results, Establishments per Immigrant (β_t)



Source: Author's calculations from IPUMS-USA and County Business Patterns.
Notes: 95% confidence intervals represented by gray area around point estimates indicated by solid black line. These confidence intervals only apply to the "reduced form"—"first stage" only used as a scaling factor. Combined estimates reflect the cases of Arizona with 2008 as year 0 and Florida with 1980 as year 0.

A.2 Instrument Vetting

This section presents a variety of analyses and visualizations that bolster the case for $z_{gkt}^{\text{Emigrants}}$ as a relevant and valid instrument in Sections 1.3 and 1.4. As of this draft, these analyses are done using publicly-available data in order to limit the size of disclosure requests from the U.S. Census Bureau. That means independent variables like I_{gkt} are measured from IPUMS-USA (Ruggles et al., 2019a) and industry groups k are simply 1-digit SIC classifications to accommodate the loss in precision and retain consistency over time. Estimated specifications are from Equation (1.3.1), but with y_{gkt} corresponding to establishment presence, as measured in the County Business Patterns (CBP), instead of LBD-measured firm presence.

A.2.1 Educational Decomposition of Inflows

The following table decomposes immigrant inflows in ΔI_{gkt} into three education categories: high school degree, some college, and college degree or more. As with the Results seen in Table 1.2, it demonstrates that immigrant inflows tend to be tilted towards workers with lower educational attainment, and that the push represented by $\Delta z_{gkt}^{\text{Emigrants}}$ does not substantially differ from the general inflow represented by ΔI_{gkt} on this front. Each immigrant pushed by $\Delta z_{gkt}^{\text{Emigrants}}$, for example, pushes 0.654 immigrants with at most a high school degree into a commuting zone-sector on average (Column 4), while this number is 0.745 for general inflows that may be contaminated by pull factors (Column 1).

Table A.4: Educational Content of Immigrant Inflows (Publicly Available Data, k = SIC Sector)

	OLS			Emigrants Instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
ΔI_{gkt}: Immigrant Inflows per Initial Worker	0.745***	0.121***	0.134***	0.654***	0.167***	0.180***
	(0.023)	(0.009)	(0.016)	(0.023)	(0.010)	(0.019)
Education	≤H.S. Deg.	Some College	College+	≤H.S. Deg.	Some College	College+
Within R^2	0.849	0.381	0.221	0.836	0.328	0.195
α_{gt}, α_{kt}	✓	✓	✓	✓	✓	✓
Region × SIC Sector × Year FE	✓	✓	✓	✓	✓	✓
1980 Controls × Year FE	✓	✓	✓	✓	✓	✓
Observations	15,162	15,162	15,162	15,162	15,162	15,162

Notes: See Equation (1.3.1) for specification. Data obtained from IPUMS-USA. College equivalent refers to 0.5 times the number of workers with Some College plus all workers with a College Degree or More—see, e.g., Doms et al. (2010). All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A.2.2 Comparability of Results: Immigration and Establishment Presence Using Publicly-Available Data

This section compares results from regressions using publicly-available data in a commuting zone-SIC sector panel both to results in Table 1.5 and across use of different instruments. Of note, I introduce a third instrument along with $z_{gkt}^{\text{Standard}}$ and $z_{gkt}^{\text{Emigrants}}$. The new instrument is $\Delta z_{gkt}^{\text{Births}}$:

$$\Delta z_{gkt}^{\text{Births}} = \frac{1}{E_{g,1980}} \left[\sum_{o} \pi_{go,1980} \times \rho_{(-g)okt} \times \left(\sum_{s=t-40}^{t-31} \text{Births}_{os} \right) \right]$$

The impetus behind this instrument comes from Hanson et al. (2017), who demonstrate that birth cohort sizes in foreign countries relative to those in the U.S. explain migration of young, Central American workers to the U.S. Here, I expand this idea to cover how differential timing of demographic transitions abroad can generate pushes to migrate to the U.S., given that the U.S. is the largest immigration destination in the world. Given that lagged births in sending countries present a compelling case for satisfying random assignment, even relative to the emigrant instrument (e.g., if there are concerns about correlated demand across OECD countries), results using $\Delta z_{gkt}^{\text{Births}}$ serves as a natural benchmark for results from Equation (1.3.1). However, using long-run factors like demographic transitions to instrument for immigration inflows can only be utilized for decadal or longer-horizon specifications. Thus, the primarily utility of $\Delta z_{gkt}^{\text{Births}}$ is to show that it delivers similar results to $\Delta z_{gkt}^{\text{Emigrants}}$.Then, $z_{gkt}^{\text{Emigrants}}$ can be utilized in Section 1.4 where the time horizon shortens to every 5 years.

Table A.5 demonstrates that this is the case. Results using $\Delta z_{gkt}^{\text{Emigrants}}$ and $\Delta z_{gkt}^{\text{Births}}$ cluster around 0.55, while results using $\Delta z_{gkt}^{\text{Standard}}$ are substantially attenuated. Of note, the results from $\Delta z_{gkt}^{\text{Emigrants}}$ and $\Delta z_{gkt}^{\text{Births}}$ are nearly identical, but slightly larger than results from Table 1.5, which utilizes smaller industry groups, restricted access demographic data and firms instead of establishments as the outcome. Thus, while future drafts will conduct the checks contained in Section A.2 using restricted-access data, Table A.5 indicates that the checks using publicly-available data presented in this draft are likely still informative.

Table A.5: Results Across Instruments (Publicly Available Data, k = SIC Sector)

	Outcome: Change in Establishments per Initial Worker			
	(1)	(2)	(3)	(4)
ΔI_{gkt}: Immigrant Inflows per Initial Worker	0.0393***	0.0539***	0.0563***	0.0224**
	(0.0044)	(0.0062)	(0.0073)	(0.0107)
Instrument	None—OLS	Emigrants	Lagged Births	Standard
1st Stage F Statistic	—	63.42	24.76	20.67
Within R^2	0.056	0.048	0.045	0.045
α_{gt}, α_{kt}	✓	✓	✓	✓
Region × SIC Sector × Year FE	✓	✓	✓	✓
1980 Controls × Year FE	✓	✓	✓	✓
Observations	5,054	5,054	5,054	5,054

Notes: See Equation (1.3.1) for specification. Data obtained from IPUMS-USA and County Business Patterns. All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A.2.3 Country-Level Pushes

In order for $z_{gkt}^{\text{Emigrants}}$ to provide a good source of exogenous variation for I_{gkt}, we not only need it to be randomly assigned, but also to be relevant. In this section, I show simple evidence that $M_{ot}^{\text{non-US}}$ is a relevant predictor of I_{ot}—total immigrant stock in the U.S. from country o. Crucially, this is true both in levels and changes (including a country fixed effect)—outflows to non-U.S. OECD countries predict inflows to the U.S. Thus, $z_{gkt}^{\text{Emigrants}}$ is not simply replacing the "shift" component of the "shift-share" instrument with noise. Instead, it is replacing it with a relevant factor that is plausibly exogenous relative to commuting zone-industry level outcomes in the U.S. The Regressions take the form of

$$f(I_{ot}) = \alpha + \beta^{\text{Push}} f(M_{ot}^{\text{non-US}}) + g(\alpha_o, \alpha_t) + \varepsilon_{ot}$$

where $f(x)$ either logs x, keeps x level, or divides x by origin o's 1980 population. $g(\cdot)$ represents a linear combination. Table A.6 presents the results. Note that because these results use publicly-available they do not reflect the full set of over 150 countries used to generate $z_{gkt}^{\text{Emigrants}}$ in the main results using restricted-access data. Future drafts will update these results accordingly.

Table A.6: Push Factors

	Outcome: $\frac{I_{ot}}{\text{Population}_{o,1980}}$		Outcome: $\log(I_{ot})$		Outcome: I_{ot}	
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{M_{ot}^{\text{non-US}}}{\text{Population}_{o,1980}}$	0.6861***	0.7719***				
	(0.1196)	(0.1128)				
$\log(M_{ot}^{\text{non-US}})$			0.4430***	0.5516***		
			(0.0666)	(0.0616)		
$M_{ot}^{\text{non-US}}$					0.1883**	0.3724**
					(0.0748)	(0.1889)
Within R^2	0.413	0.400	0.247	0.377	0.017	0.012
Year FE	✓	✓	✓	✓	✓	✓
Country FE		✓		✓		✓
Observations	420	420	420	420	420	420
Countries	105	105	105	105	105	105

Notes: Dependent variables obtained from IPUMS-USA. Indpendent variables obtained from the IAB, and the United Nations' World Population Prospects 2017 for population denominator. Standard errors clustered at the country level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A.2.4 Pre-Trends and Balance Tests

A distinct advantage of the County Business Patterns (CBP) relative to the restricted-access Longitudinal Business Database (LBD) is that the CBP stretches farther back in time. This enables a pure pre-trends test that is not often available in panel, shift-share settings. This pre-trends test takes the following regression form:

$$\Delta\text{Establishments}_{gk,1980} = \alpha + \pi_t^{\text{Pure}} \left[\Delta z_{gkt}\right] + \Gamma X_{gkt} + \epsilon_{gkt}$$

where a rejection of the null hypothesis $\pi^{\text{Pure}} = 0$ indicates the presence of pre-trends in the instrumental variable. In addition, and as seen in Table 1.5, I operationalize a pre-trends test within the analysis panel, following suggestions made in Borusyak et al. (2018). This version of the pre-trends test first runs the reduced form regression implied in Equation (1.3.1):

$$\Delta\text{Establishments}_{gkt} = \alpha + \beta^{\text{RF}} \left[\Delta z_{gkt}\right] + \Gamma X_{gkt} + \alpha_{gt} + \alpha_{kt} + u_{gkt}$$

and collects residuals $\hat{u}_{gkt}$. Then, the pre-trends test is conducted through the regression

$$\hat{u}_{gkt} = \alpha + \pi^{\text{Panel}} \left[\Delta z_{gk,t+10}\right] + \eta_{gkt}$$

where rejection of the null hypothesis $\pi^{\text{Panel}} = 0$ is once again evidence of deleterious pre-trends. Table A.7 presents the p-values from these tests. $\Delta z_{gkt}^{\text{Emigrants}}$ out-performs $\Delta z_{gkt}^{\text{Standard}}$ across the two tests, and $\Delta z_{gkt}^{\text{Births}}$ performs well, as expected. Notably $\Delta z_{gkt}^{\text{Standard}}$ displays evidence of pre-trends that could confound analysis.

Table A.7: Pre-Trends Test p-values (Publicly Available Data, k = SIC Sector)

	Standard	Emigrants	Births
Pure Pre-Trends Test p-value, 1980s ($H_0: \pi_{t=1990}^{\text{Pure}} = 0$)	0.046	0.320	0.081
Pure Pre-Trends Test p-value, 1990s ($H_0: \pi_{t=2000}^{\text{Pure}} = 0$)	<0.001	0.053	0.135
Pure Pre-Trends Test p-value, 2000s ($H_0: \pi_{t=2010}^{\text{Pure}} = 0$)	0.735	0.141	0.158
Panel Pre-Trends Test p-value ($H_0: \pi^{\text{Panel}} = 0$)	0.098	0.241	0.710

Notes: Data obtained from IPUMS-USA and County Business Patterns. Standard errors that underly p-values in final row have *not yet* been bootstrapped to account for two-step estimation.

A.2.5 Correlated Shocks Across gk with Similar Shares (Adao et al., 2019)

Adao et al. (2019) find that regression residuals can be substantially correlated across areas with similar "share" components in shift-share instruments, invalidating standard inference procedures. Importing their concerns to the current study, any industry-country level shocks that affect outcomes through the presence of base year shares $\pi_{go,1980}$, even if not related to immigration itself, can generate correlated outcomes across areas with similar $\pi_{go,1980}$.

A simple example that could apply here would be a sector-specific trade shock in a given origin country. For example, if Syria experiences a positive trade shock that is independent of Syrian emigration forces, this can affect firm presence in areas heavily populated by Syrians in the U.S.—e.g., Detroit and Boston—through trade linkages, but it is unlikely to have any affect on firm presence in areas like Atlanta or Miami with low Syrian populations. These correlated shocks would not create a bias in β, but would require a modification of standard errors beyond clustering at the commuting zone-industry level, since Detroit and Boston are not even in the same region.

Adao et al. (2019) illustrate this issue by conducting placebo tests in which they replace the "shift" component with white noise and assessing the resulting false rejection rate after multiple simulations of the reduced form regression model. Here, the instrumental variable takes the form

$$\Delta z_{gkt} = \frac{1}{E_{g,1980}} \sum_{o} \pi_{go,1980} \left(\rho_{ok(-g)t} \times \text{Push}_{ot} - \rho_{ok(-g),t-10} \times \text{Push}_{o,t-10}\right)$$

Thus, the analogous placebo exercise uses instruments of the form

$$z_{gkt}^{\text{Placebo}} \frac{1}{E_{g,1980}} \sum_{o} \pi_{go,1980} \times \omega_{okt}$$

where ω_{okt} is a random draw from a normal distribution with variance of $\rho_{ok(-g)t} \times \text{Emigrants}_{ot}$. Each placebo instrument is then used in the reduced form estimating equation:

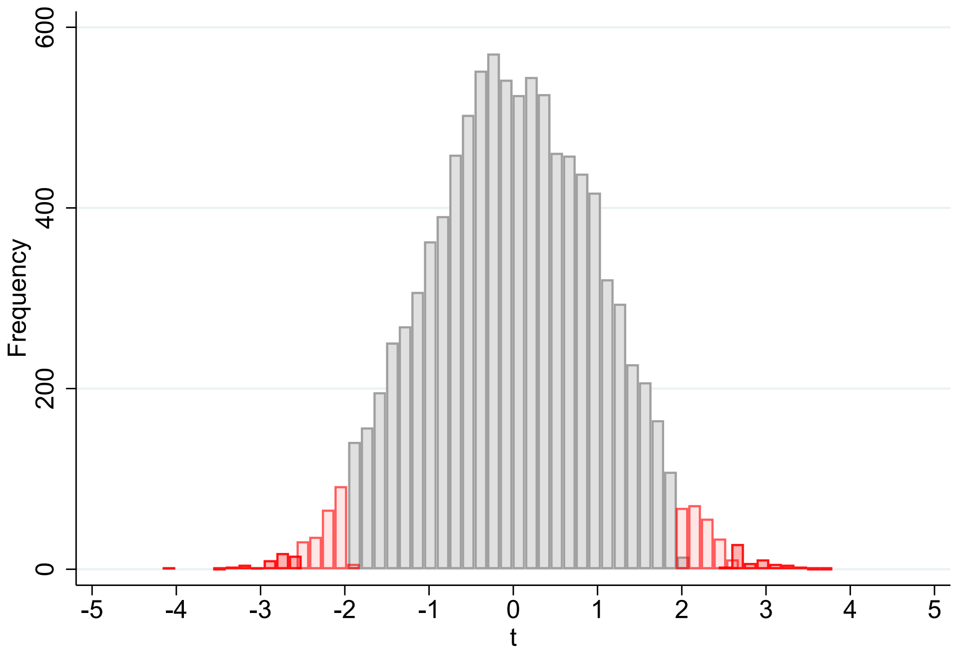
$$\frac{\Delta \text{Establishments}_{gkt}}{\text{Workers}_{gk,t-10}} = \alpha + \beta^{\text{Placebo}} \left[\Delta z_{gkt}^{\text{Placebo}}\right] + \Gamma X_{gkt} + \alpha_{gt} + \alpha_{kt} + u_{gkt}$$

with errors clustered at the commuting zone-sector level, as in the main analysis.

Results from 1,000 placebo simulations are presented in Figure A.3. There is not strong evidence of any bias in the standard error estimates when clustering at the commuting zone-industry level. The false rejection rate at the 95% confidence level is 0.0592 (should be 0.05) and the false rejection rate at the 99% confidence level is 0.0121 (should be

0.01). While these results are reassuring from the standpoint of the significance levels and confidence intervals presented in this chapter, future drafts will still incorporate standard error estimators suggested by Adao et al. (2019) for completeness.

Figure A.3: Distribution of t-Statistics from Adao et al. (2019) Placebo Simulations (Publicly Available Data, k = SIC Sector)



Notes: Data obtained from IPUMS-USA and County Business Patterns. It only applies to sectors and commuting zones included in analysis presented below. Light red bars are false rejections at the 95% level. Darker red bars are false rejections at the 99% level.

A.2.6 Confounding Short and Long-Run Responses (Jaeger et al., 2018)

Jaeger et al. (2018) broach a concern that arises from the serial correlation in the "shift" component of "shift-share" instrumentation. When this shift component is excessively serially correlated, estimated parameters like β in in Equation (1.3.1) can confound short- and long-run responses to immigrant inflows. Though this concern is particularly deleterious when wages are the primary outcome variable of interest, it merits consideration in any setting where prior shocks may affect current outcomes. Jaeger et al. (2018) propose a data-demanding procedure to both test for and account for such concerns, which is to include both the independent variable and its lag, and to instrument for both (i.e., here, include ΔI_{gkt} and $\Delta I_{gk,t-10}$ and use both $\Delta z_{gkt}^{\text{Emigrants}}$ and $\Delta z_{gk,t-10}^{\text{Emigrants}}$ as instruments. Table

Table A.8: Correlations in Instruments

	Instrument:		
	Standard (1)	Emigrants (2)	Births (3)
Correlation with:			
Lagged Instrument	0.8595	0.7331	0.0469
Immigrant Inflows	0.5120	0.5088	0.4059
Lagged Immigrant Inflows	0.5472	0.4434	0.2033

Notes: Observations weighted by 1980 workforce size.

1.5 already demonstrates that $\Delta z_{gkt}^{\text{Emigrants}}$ passes this rigorous test, and additionally provides evidence that the effect of immigrant inflows on firm presence is largely confined to the decade of the inflow, at least when comparing across industries within commuting zones.

Here, I present evidence that, within the specification laid out in Equation (1.3.1), this robustness is a unique feature to instruments that have plausible exogenous components in the "shift" component, instead of aggregate inflows. In line with Jaeger et al. (2018), I hypothesize that the reason is that these exogenous components are less serially correlated over time than aggregate immigrant inflows. Note once again that the results below in Table A.9 are not identical to Table 1.5 when using $\Delta z_{gkt}^{\text{Emigrants}}$ because the results here use publicly-available data. While both $\Delta z_{gkt}^{\text{Emigrants}}$ lose a substantial amount of first stage relevance, they both retain enough to make reasonable inferential statements about estimated β coefficients. On the other hand, the specification with $\Delta z_{gkt}^{\text{Standard}}$ breaks down on multiple fronts when employing the double instrumentation strategy.

Table A.8 presents some simple evidence for why this might be the case. As found in Jaeger et al. (2018), the standard instrument $\Delta z_{gkt}^{\text{Standard}}$ is more correlated with $\Delta I_{gk,t-10}$ than it is with contemporaneous ΔI_{gkt}, which it is supposed to instrument for. The same cannot be said for either $\Delta z_{gkt}^{\text{Emigrants}}$ or $\Delta z_{gkt}^{\text{Births}}$. $\Delta z_{gkt}^{\text{Standard}}$ also has the highest serial correlation among the three instruments. Interestingly, the relatively small improvements $\Delta z_{gkt}^{\text{Emigrants}}$ make over $\Delta z_{gkt}^{\text{Standard}}$ are enough to pass the double-instrumentation test presented in Table A.9. As expected, $\Delta z_{gkt}^{\text{Births}}$ performs very well on these checks because of varying timing in demographic transitions in sending countries. This is why it sees the smallest First Stage F statistic drop-off in Table A.9 between single- and double-instrumentation specifications. Similar evidence regarding $\Delta z_{gt}^{\text{Standard}}$ is shown in Jaeger et al. (2018).

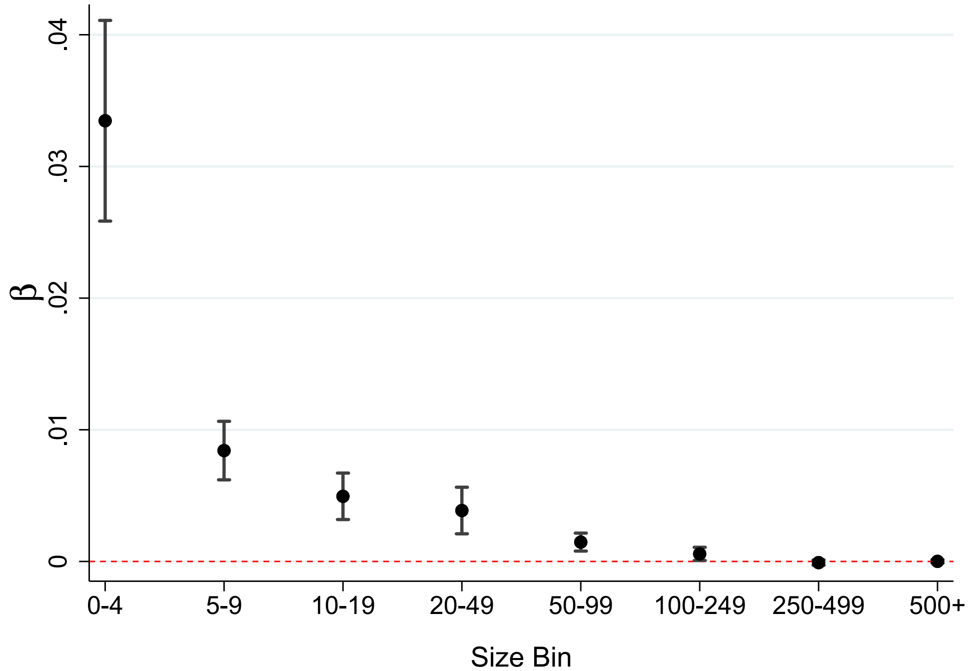
Table A.9: Results Across Instruments, with Double Instrumentation (Publicly Available Data, k = SIC Sector)

	Outcome: Change in Establishments per Initial Worker							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔI_{gkt}	0.0393***	0.0427***	0.0539***	0.0677***	0.0563***	0.0721***	0.0224**	0.1270
	(0.0044)	(0.0058)	(0.0062)	(0.0158)	(0.0073)	(0.0149)	(0.0107)	(0.1870)
$\Delta I_{gk,t-10}$		-0.0130***		-0.0383*		-0.0488**		-0.1550
		(0.0049)		(0.0232)		(0.0190)		(0.2570)
Instrument	None—OLS	None—OLS	Emigrants	Emigrants	Lagged Births	Lagged Births	Standard	Standard
1st Stage F Statistic	—	—	63.42	9.431	24.76	11.98	20.67	0.10
Within R^2	0.056	0.064	0.048	0.036	0.045	0.018	0.045	-0.532
α_{gt}, α_{kt}	✓	✓	✓	✓	✓	✓	✓	✓
Region × SIC Sector × Year FE	✓	✓	✓	✓	✓	✓	✓	✓
1980 Controls × Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	15,162	10,108	15,162	10,108	15,162	10,108	15,162	10,108

Notes: See Equation (1.3.1) for specification. All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A.3 Firm Size Results in Levels

Figure A.4: The Effect of Immigration on Firm Presence—Size Decomposition



Notes: See Equation (1.3.1) for specification. Data accessed and analyzed in Michigan Census Research Data Center.

A.4 Helpful Relationships for Back of the Envelope Calculations

A.4.1 Census Self-Employed Workers versus Business Registrar-Based Establishment Presence

This sections helps us interpret how much of the effect of immigration on firm presence comes from immigrant entrepreneurship. First, Table A.10 below shows that the emigrants instrument $\Delta z_{gkt}^{\text{Emigrants}}$ brings in roughly 0.86 employees and 0.14 self-employed workers for every immigrant it pushes into a given gk. Next, in Table A.11 I regress $\frac{\Delta \text{Self-Employed}_{gkt}}{\text{Workers}_{gk,t-10}}$ (from the ACS/Census) on $\frac{\Delta \text{Establishments}_{gkt}}{\text{Workers}_{gk,t-10}}$ (from the County Business Patterns) using Equation (1.3.1) to get a ballpark estimate of how many Longitudinal Database (LBD) establishments correspond to each self-employed individual measured in the ACS/Census. Note that the County Business Patterns and the LBD have the same source data: the Business Registrar. Each self-employed worker is associated with roughly 0.16 CBP establishments. Thus, a ball-park estimate of the entrepreneurship effect each immigrant has on establishment presence is $0.14 \times 0.16 \approx 0.022$.

A.4.2 Census Worker Counts and Business Registrar-Based Employment

This section uses publicly available data from the Decennial Censuses and American Community Surveys (ACS) to construct the independent variable $\frac{\Delta \text{Workers}_{gkt}}{\text{Workers}_{gk,t-10}}$ and the County Business Patterns to construct the numerator of the dependent variable $\frac{\Delta \text{Employment}_{gkt}}{\text{Workers}_{gk,t-10}}$. The result from estimating equation (1.3.1) using these variables, in Table A.12, is a ballpark estimate of how many LBD jobs each Decennial Census worker corresponds to. In turn, this helps us interpret the results in Section 1.3.5. Full absorption of immigrant workers, as measured by the Decennial Census and ACS would correspond to at least 0.64 LBD jobs per immigrant worker.

Table A.10: Class Content of Immigrant Inflows (Publicly Available Data, k = SIC Sector)

	OLS		Emigrants Instrument	
	(1)	(2)	(3)	(4)
ΔI_{gkt}: Immigrant Inflows per Initial Worker	0.890***	0.110***	0.861***	0.139***
	(0.009)	(0.009)	(0.019)	(0.019)
Class of Worker	Employee	Self-Employed	Employee	Self-Employed
Within R^2	0.969	0.320	0.968	0.298
α_{gt}, α_{kt}	✓	✓	✓	✓
Region × SIC Sector × Year FE	✓	✓	✓	✓
1980 Controls × Year FE	✓	✓	✓	✓
Observations	15,162	15,162	15,162	15,162

Notes: See Equation (1.3.1) for specification. Data obtained from IPUMS-USA. All specifications include control variables for 1980 log employment, 1980 establishments per worker, 1980 self employment share, 1980 college share, and 1980 under-40 share in the commuting zone-sector interacted with year fixed effects. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table A.11: Census Self-Employed and CBP Estabs. (Publicly Available Data, k = SIC Sector)

	$\frac{\Delta \text{CBP Establishments}_{gkt}}{\text{Census Workers}_{gk,t-10}}$
$\frac{\Delta \text{Census Self-Employed Workers}_{gkt}}{\text{Census Workers}_{gk,t-10}}$	0.1615***
	(0.0069)
Within R^2	0.2039
α_{gt}, α_{kt}	✓
Region × SIC Sector × Year FE	✓
1980 Controls × Year FE	✓
Observations	15,162

Notes: Data obtained from IPUMS-USA and County Business Patterns. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table A.12: Census Workers and CBP Employment (Publicly Available Data, k = SIC Sector)

	$\frac{\Delta \text{CBP Employment}_{gkt}}{\text{Census Workers}_{gk,t-10}}$
$\frac{\Delta \text{Census Workers}_{gkt}}{\text{Census Workers}_{gk,t-10}}$	0.6431***
	(0.0222)
Within R^2	0.2455
α_{gt}, α_{kt}	✓
Region × SIC Sector × Year FE	✓
1980 Controls × Year FE	✓
Observations	15,162

Notes: Data obtained from IPUMS-USA and County Business Patterns. Observations weighted by 1980 workforce size. Standard errors clustered at the commuting zone-sector level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

A.5 Model Additions and Extensions

A.5.1 Comparing the Immigrant Surplus in a Representative Firm Model to a Model with Firm Heterogeneity

As a starting point, consider the following, constant elasticity of substitution (CES), model of production for a representative firm[2] in a local economy:

$$Q = \bar{z}\left(aI^{\frac{\sigma_I-1}{\sigma_I}} + \bar{N}^{\frac{\sigma_I-1}{\sigma_I}}\right)^{\frac{\sigma_I}{\sigma_I-1}} \tag{A.1}$$

where I represents immigrant labor, $\bar{N}$ represents a fixed stock of native labor, $\bar{z}$ is total factor productivity, and σ_I is the elasticity of substitution between immigrant and native employees. While a more realistic version of the production function would start with nests for education and eventually work its way down to nativity (see, e.g., Ottaviano and Peri, 2008), this model delivers much of the intuition for the rest of this section in a simple way. In this context, σ_I can be thought of as a reduced form parameter that aggregates all the different reasons why an average immigrant worker may be different than an average native worker (including average educational attainment—see Figure 1.3 and Table 1.2).

Let native wages, w_N, be the numeraire. There is only one good, so consumer welfare is simply a function of its price and wages. When I increases in this model, the economy expands automatically (we can think of this as long run capital adjustment with a fixed rental rate in the background), but what matters to native workers is how it affects the price of the good. Denoting immigrant wages as w_I, P as the price of the good, taking first order conditions, and rearranging yields the following expression:

$$\mathcal{W}_I \equiv -\frac{d\log(P)}{dI} = -\frac{d\log(c)}{dI} \tag{A.2}$$

where

$$c \equiv \left(a^{\sigma_I} w_I^{1-\sigma_I} + 1\right)^{\frac{1}{1-\sigma_I}}$$

are labor costs to the firm. $\mathcal{W}_I$ is proportional to the immigrant surplus—the surplus accruing to the native workers as a result of immigrant inflows. Here, it is directly tied to the labor cost savings that occur as immigrant wages decline in response to I rising. Using similar, CES production function models, both Borjas (2014) and Ottaviano and Peri (2008) find that $\mathcal{W}_I$ is small and positive.

Sections 1.3 and 1.4 will show that immigrants have large and heterogeneous effects on extensive margin decisions by firms. This suggests accounting for these effects in our

[2] Or, the aggregation of many, small, identical firms.

theoretical analysis of immigration. This point has been made explicitly by di Giovanni et al. (2014), who feature gains from variety in their global welfare analysis of immigration. Melitz (2003) offers a simple way to incorporate extensive margin firm responses in long-run, steady state, general equilibrium analysis. The key features of this model are 1) consumer taste for variety, 2) a non-degenerate total factor productivity distribution across firms, and 3) a fixed cost of production. Combined, these features generate monopolistic competition that results in a non-trivial, but finite, firm mass, with heterogeneity across each firm's total factor productivity level.

We can gain simple insights into how our analysis of immigration may change when we account for these features by placing the production function from (A.1) into the closed economy Melitz (2003) framework, where each firm is indexed by its own total factor productivity, z:

$$q(z) = zL(z)$$
$$L(z) = \left[a\left(I(z)\right)^{\frac{\sigma_I-1}{\sigma_I}} + \left(N(z)\right)^{\frac{\sigma_I-1}{\sigma_I}}\right]^{\frac{\sigma_I}{\sigma_I-1}}$$

Consumers have elasticity of substitution μ across firms, leading to a conventional pricing rule for each firm:

$$p(z) = \left(\frac{\mu}{\mu-1}\right)\left(\frac{c}{z}\right) \tag{A.3}$$

where w_I is the immigrant wage and we once again set native wages w_N to be the numeraire. The overall price index is given by

$$P^{1-\mu} = n_e \int_{z^*}^{\infty} p(z)^{1-\mu} g(z) dz$$

where n_e is the entrepreneur mass and z^* is the cutoff productivity, below which these entrepreneurs suffer losses and therefore exit in the long run. $g(z)$ is the distribution of productivity across firms in the local economy. I follow convention in assuming it is Pareto:

$$g(z) \equiv \phi m^{\phi} z^{-\phi-1}$$

The mass of firms in the local economy is simply $F \equiv n_e \int_{z^*}^{\infty} g(z)dz = n_e m^{\phi}\left(z^*\right)^{-\phi}$.

With these simple ingredients in place, we can derive the following expression:

$$\mathcal{W}_I \equiv \underbrace{-\frac{d\log(c)}{dI}}_{\text{Same as above}} \qquad + \qquad \underbrace{\left(\frac{1}{\mu-1}\right)\frac{d\log(F)}{dI}}_{\text{Increased variety through more firms}} \qquad + \qquad \underbrace{\frac{d\log(z^*)}{dI}}_{\text{Productivity pass-through to prices}} \tag{A.4}$$

In this setup, the immigration surplus has two additional terms compared to the representative firm model, in Equation (A.2). The first represents welfare gains in the form of increased consumer variety. The second represents welfare gains that arise from an increase in the productivity bar that entrepreneurs must cross in order to operate in the market. As seen in Equation (A.3), firms with higher z charge lower prices to consumers in order to compete away market share from their competitors. Thu\phi-(\mu-1)s, when z^* rises and lower productivity firms exit the market, consumers benefit through lower prices. The signs of each of these additional reduced form parameters are explored in detail in the empirical analyses of this chapter. Section 1.3 finds that increased exposure to immigrant workers generates an increase in firm presence in local labor markets, while Section 1.4 finds that low productivity firms are culled from the market—in the language of this model, the latter indicates an increase in z^*.

A.5.2 Derivation of Labor Market Equilibrium

In this section, I derive labor market equilibrium for low-education immigrant labor. An analogous derivation gives us labor market equilibrium for low-education native labor. Skilled labor's price is set to be the numeraire, as described in the text. Firm's maximize the following expression for profits:

$$\pi_j(z) = p_j(z)q_j(z) - \sum_i \sum_e w_{ie} i_e - c_j \kappa_j^f$$

where $i \in \{I, N\}$ and $e \in \{U, S\}$. They take their demand curves, $q_j(z) = p_j(z)^{-\mu} P^{\mu-1} Y$ and their production functions $q_j(z) = zL_j(z)$ as given. Thus, first order conditions yield:

$$\begin{aligned} w_{IU} &= c_j L^{1/\sigma_E} a U_j^{1/\sigma_I - 1/\sigma_E} b_{Uj} I_{Uj}^{-1/\sigma_I} \\ w_{NU} &= c_j L^{1/\sigma_E} a U_j^{1/\sigma_I - 1/\sigma_E} N_{Uj}^{-1/\sigma_I} \end{aligned} \tag{A.5}$$

So, we have the familiar relative wage expression among low-education immigrant and native workers:

$$\frac{w_{IU}}{w_{NU}} = \left(\frac{I_{Uj}}{N_{Uj}}\right)^{-1/\sigma_I} b_{Uj}$$

Solving for N_{Uj} and plugging into the low-education aggregate U_j then yields the following expression for U_j in terms of I_{Uj}:

$$U_j = I_{Uj} b_{Uj}^{-\sigma_I} w_{IU}^{\sigma_I} c_{Uj}^{-\sigma_I}$$

Plugging this expression back into (A.5) yields the following expression for I_{Uj} in terms of wages and L_j:

$$I_{Uj} = w_{IU}^{-\sigma_I} b_{Uj}^{\sigma_I} c_j^{\sigma_E} c_{Uj}^{\sigma_I - \sigma_E} a^{\sigma_E} L_j \equiv I_{Uj}^{\text{unit}} L_j$$

Firms use L_j to produce output and to cover their fixed costs. Thus, integrating across firms yields the following expression that equates low-education immigrant labor supply, I_U, with labor demand:

$$I_U = n_e \left[\int_{z_0^*}^{z_1^*} I_{U0}^{\text{unit}} \left(\frac{q_0^*(z)}{z} + \kappa_0^f \right) g(z) dz + \int_{z_1^*}^{\infty} I_{U1}^{\text{unit}} \left(\frac{q_1^*(z)}{z} + \kappa_1^f \right) g(z) dz \right]$$

where $q_0^*(z)$ and $q_1^*(z)$ are optimal output choices—plugging in the pricing rule $p_j(z) = \left(\frac{\mu}{\mu-1}\right)\left(\frac{c_j}{z}\right)$ into the demand expression. After some algebra, the final expression for I_U becomes:

$$I_U = a^{\sigma_E} w_{IU}^{-\sigma_I} c_0^{-1} Y \Bigg\{ b_{U0}^{\sigma_I} c_0^{\sigma_E} c_{U0}^{\sigma_I - \sigma_E} \left[\left(\frac{\mu}{\mu - 1} \right)^{-1} (\theta)^{-1} \left(1 - R_z^{-(\phi - (\mu - 1))}\right) + \left(\frac{\phi - (\mu - 1)}{\phi\mu} \right) \left(1 - R_z^{-\phi}\right) \right] +$$
$$b_{U1}^{\sigma_I} c_1^{\sigma_E} c_{U1}^{\sigma_I - \sigma_E} \left(\frac{c_1}{c_0} \right)^{-\mu} \left[\left(\frac{\mu}{\mu - 1} \right)^{-1} (\theta)^{-1} \left(R_z^{-(\phi - (\mu - 1))}\right) + \tau \left(\frac{(\phi - (\mu - 1))}{\phi\mu} \right) \left(R_z^{-\phi}\right) \right] \Bigg\}$$

APPENDIX B

Appendix to Chapter 2

B.1 Decomposition of Recession-Induced Losses

This appendix explains how we decompose recession-induced wage losses into three different components.

Start with the AKM/CHK decomposition (Equation (2.3.1)) in period t

$$\log(\text{wage}_{it}) = \alpha_i + \psi_j \mathbf{1}\{i \text{ works at } j \text{ in } t\} + \boldsymbol{x}'_{it}\beta + r_{it}.$$

Taking averages on both sides of Equation (2.3.1) in period t:

$$E\left[\log(\text{wage}_{it})\right] = E\left[\alpha_i\right] + E\left[\psi_j\right] + E\left[\boldsymbol{x}'_{it}\beta\right]. \tag{B.1}$$

Define the average wage in period t if the same group of individuals were to experience a 1 point increase in the unemployment rate at entry as

$$E\left[\log(\text{wage}^R_{it})\right] = E\left[\alpha^R_i\right] + E\left[\psi^R_j\right] + E\left[\boldsymbol{x}'_{it}\beta\right]. \tag{B.2}$$

In Equation (B.2), α^R_i and ψ^R_j represent estimated person and establishment fixed effects estimated for the same underlying individuals if they entered the labor market with a 1 point higher unemployment rate. Notably, person fixed effects in the AKM/CHK framework are subject to scarring and will be lower for otherwise identical individuals who, for exogenous reasons, have lower lifetime earnings.

Now, we can write $E\left[\log(\text{wage}_{it})\right] - E\left[\log(\text{wage}^R_{it})\right]$ as:

$$\underbrace{\beta_t^{\text{Wage}}}_{\text{\% wage differential}} \quad = \quad \underbrace{E\left[\alpha_i\right] - E\left[\alpha_i^R\right]}_{\text{\% due to non-employer factors}} \quad + \quad \underbrace{E\left[\psi_j\right] - E\left[\psi_j^R\right]}_{\text{\% due to employer-related factors}} \quad . \tag{B.3}$$

Next, define

$$\beta_t^{\text{Non-employer}} = E\left[\alpha_i\right] - E\left[\alpha_i^R\right], \tag{B.4}$$

$$\beta_t^{\text{Employer}} = E\left[\psi_j\right] - E\left[\psi_j^R\right], \tag{B.5}$$

as the non-employer-specific and employer-specific components of recession-induced wage differentials. Estimating our main specification (Equation (2.4.1)) using CHK establishment fixed effects (ψ_j) on the left-hand-side yields $\left\{\hat{\beta}_0^{\text{Employer}}, \ldots, \hat{\beta}_{10}^{\text{Employer}}\right\}$.

Having defined recession-induced employer-specific losses, we now partition these losses into rents and compensating differentials using the decomposition in Sorkin (2018). Equation (2.3.3) splits establishment fixed effects into rents, which are explained by value, and compensation differentials which are orthogonal to value. Taking expectations on both sides, we can write:

$$E\left[\psi_j\right] = \beta E\left[V_j\right] - E\left[a_j\right]. \tag{B.6}$$

Establishment fixed effects for otherwise identical individuals who enter the labor market when unemployment rates are 1 point higher can be written as

$$E\left[\psi_j^R\right] = \beta E\left[V_j^R\right] - E\left[a_j^R\right]. \tag{B.7}$$

Subtracting (B.7) from (B.6), the employer-specific pay reduction is

$$\beta_t^{\text{Employer}} = \underbrace{\beta\left(E\left[V_j\right] - E\left[V_j^R\right]\right)}_{\text{\% due to rents}} - \underbrace{\left(E\left[a_j\right] - E\left[a_j^R\right]\right)}_{\text{\% due to amenities}}. \tag{B.8}$$

Next, define

$$\beta_t^{\text{Rent}} = \beta\left(E\left[V_j\right] - E\left[V_j^R\right]\right), \tag{B.9}$$

$$\beta_t^{\text{Amenity}} = E\left[a_j\right] - E\left[a_j^R\right], \tag{B.10}$$

Combining Equations (B.3) and (B.8) we can write

$$\begin{aligned}\beta_t^{\text{Wage}} &= \beta_t^{\text{Non-employer}} + \beta_t^{\text{Employer}} \\ &= \beta_t^{\text{Non-employer}} + \left(\beta_t^{\text{Rent}} - \beta_t^{\text{Amenity}}\right)\end{aligned} \tag{B.11}$$

Because β_t^{Wage} ,β_t^{Rent}, and β_t^{Amenity} are estimated for $t \in \{0, \ldots, 10\}$, we can recover $\beta_t^{\text{Non-employer}}$ using Equation (B.11).

Finally, define the present value of wages for the first decade of labor market experience as

$$PV = \bar{w}_0 + \bar{w}_1(1+r)^{-1} + \cdots + \bar{w}_{10}(1+r)^{-10}, \tag{B.12}$$

where $\bar{w}_t$ represents the average daily wage earned in year t. The PDV of wages for workers who face a 1 point increase in the unemployment rate at entry is

$$PV^R = \bar{w}_0(1 - \beta_0^{\text{Wage}}) + \bar{w}_1(1 - \beta_1^{\text{Wage}})(1+r)^{-1} + \cdots + \bar{w}_{10}(1 - \beta_{10}^{\text{Wage}})(1+r)^{-10} \tag{B.13}$$

We use our estimates $\left\{\hat{\beta}_0^{\text{Wage}}, \ldots, \hat{\beta}_{10}^{\text{Wage}}\right\}$ to quantify the loss in the present value of wages attributable to a 1 point change in the unemployment rate. We then scale the resulting estimate by a one standard deviation increase in the unemployment rate which reflects a typical recession. Similar calculations with $\left\{\hat{\beta}_0^{\text{Employer}}, \ldots, \hat{\beta}_{10}^{\text{Employer}}\right\}$, $\left\{\hat{\beta}_0^{\text{Non-employer}}, \ldots, \hat{\beta}_{10}^{\text{Non-employer}}\right\}$, $\left\{\hat{\beta}_0^{\text{Rent}}, \ldots, \hat{\beta}_{10}^{\text{Rent}}\right\}$, $\left\{\hat{\beta}_0^{\text{Amenity}}, \ldots, \hat{\beta}_{10}^{\text{Amenity}}\right\}$ yield estimates of the loss in the PV of wages attributable to employer-specific factors, non-employer specific factors, rents, and amenities.

B.2 Accounting for Measurement Error in Establishment Values

When estimating equation (2.3.3), we find that almost 96 percent of the variation in establishment fixed effects is in the residual, which is substantially larger than the 72 percent estimate obtained by Sorkin using U.S. data. The key reason for this difference is that our establishment value estimates are derived from a 2 percent sample of workers — and therefore more likely to be affected by measurement error — while Sorkin's estimates are obtained using the full population of workers in several U.S. states.

To evaluate the quantitative implications of measurement error in our estimates, we randomly partition individuals in the SIAB into two groups and re-estimate the establishment values using moves within each partition separately. For each establishment, this exercise yields two different value estimates, each of which is based on an independent sample of worker moves. Denote the error-free value of establishment j by V_j^*; the split sample estimates V_j^1 and V_j^2 are given by

$$V_j^1 = V_j^* + u_j^1 \tag{B.1}$$
$$V_j^2 = V_j^* + u_j^2. \tag{B.2}$$

Because independent samples of worker moves are used to estimate V_j^1 and V_j^2, we assume that

$$E\left[u_j^1 u_j^2\right] = 0. \tag{B.3}$$

We then re-estimate equation (2.3.3) using an instrumental variables (IV) approach by employing V_j^1 as an instrument for V_j^2 and vice-versa. Assuming that each value estimate exhibits classical measurement error (i.e. that $E\left[V_j^* u_j^1\right] = E\left[V_j^* u_j^2\right] = 0$), one would expect OLS to yield estimates biased toward zero relative to IV.

Table B.1 presents estimates of β based on the pooled set of workers, as well as estimates based on sample of establishments for which we estimate two sets of value estimates based on independent sets of worker moves. Using OLS to estimate β on the pooled sample yields a coefficient of 0.212. Estimating β using OLS in the smaller sample of establishments for which we obtain two independent estimates of value, yields coefficients of 0.239 and 0.220 which are similar in magnitude to the pooled sample estimate. In contrast, the IV-based slope estimates are approximately two times larger than those obtained using OLS, a result which is consistent with classical measurement error induced attenuation.[1]

[1]The two IV-based estimates reverse the order of the instrumental and the instrumented variable which is arbitrary. Randomly partitioning the pooled sample reduces the number of workers available to estimate

We also note that our interpretation of Figure 2.6 does not change in the presence of measurement error. While the average relationship plotted in the dashed line would tilt upwards when measurement error is corrected for, so would the implied career paths, by roughly the same amount. This is the visual manifestation of the logic that recessions do not differentially send workers to firms with more or less measurement error in their estimated values.

Table B.1: Classical Measurement Error Induced Attenuation in β

	Pooled Sample	Split Sample			
	OLS	OLS(1)	OLS(2)	IV(1)	IV(2)
$\hat{\beta}$	0.212	0.239	0.220	0.397	0.432
	(0.002)	(0.005)	(0.005)	(0.009)	(0.009)
R^2	0.04	0.06	0.05	0.03	0.004
N	171,640	59,762	59,762	59,762	59,762

Notes: All models include gender fixed effects. OLS(1) and OLS(2) use the value estimates obtained using each sample split. IV(1) and IV(2) reverse the order of the instrumental and instrumented variable which is arbitrary.

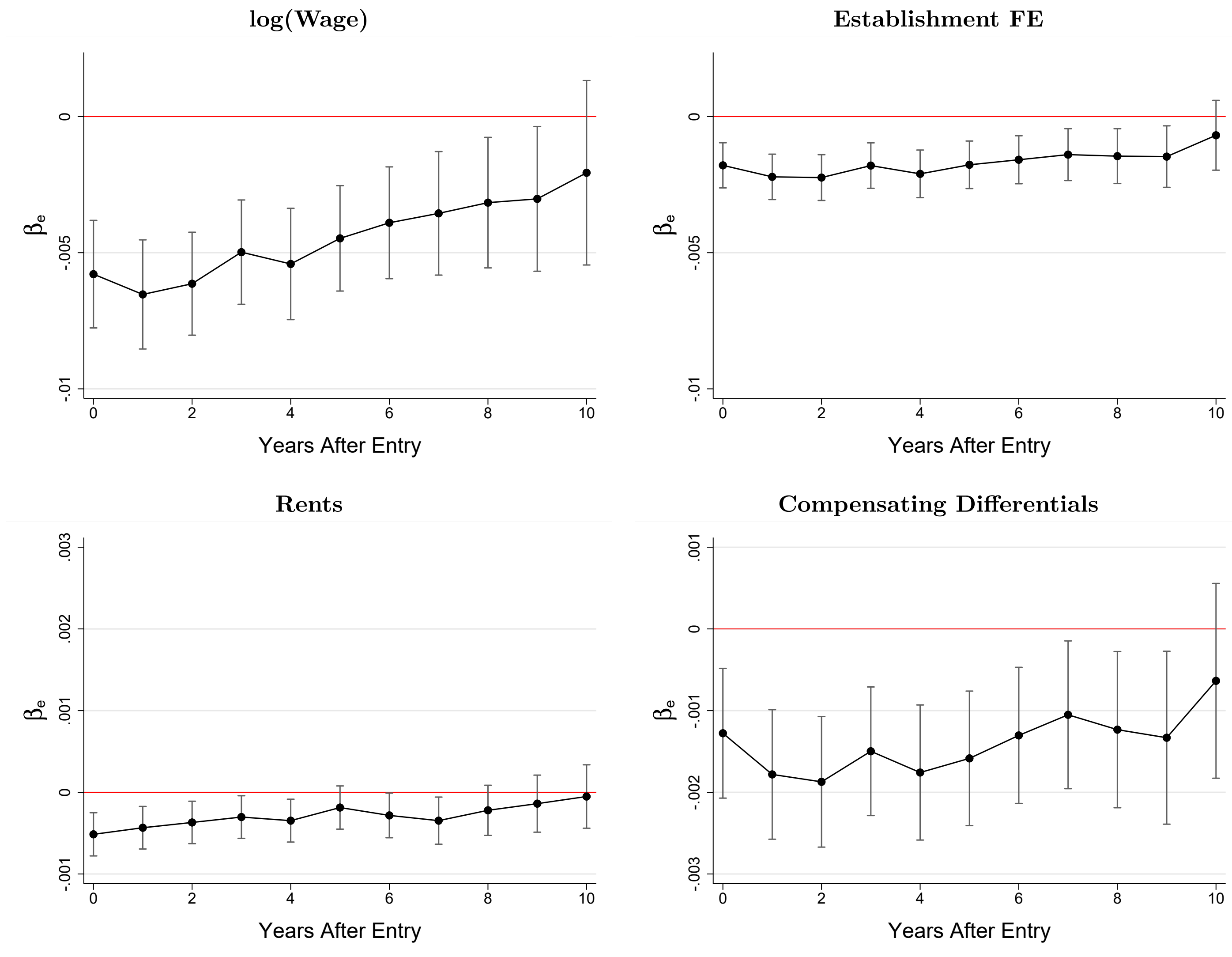
values for any given establishment, which disproportionately eliminates smaller establishments since they may no longer have sufficient flows to be in the strongly connected set. The slightly larger OLS-based estimates in the split sample versus the pooled sample are a by-product of this selectivity.

B.3 Additional Analyses

B.3.1 Using Actual Training Duration

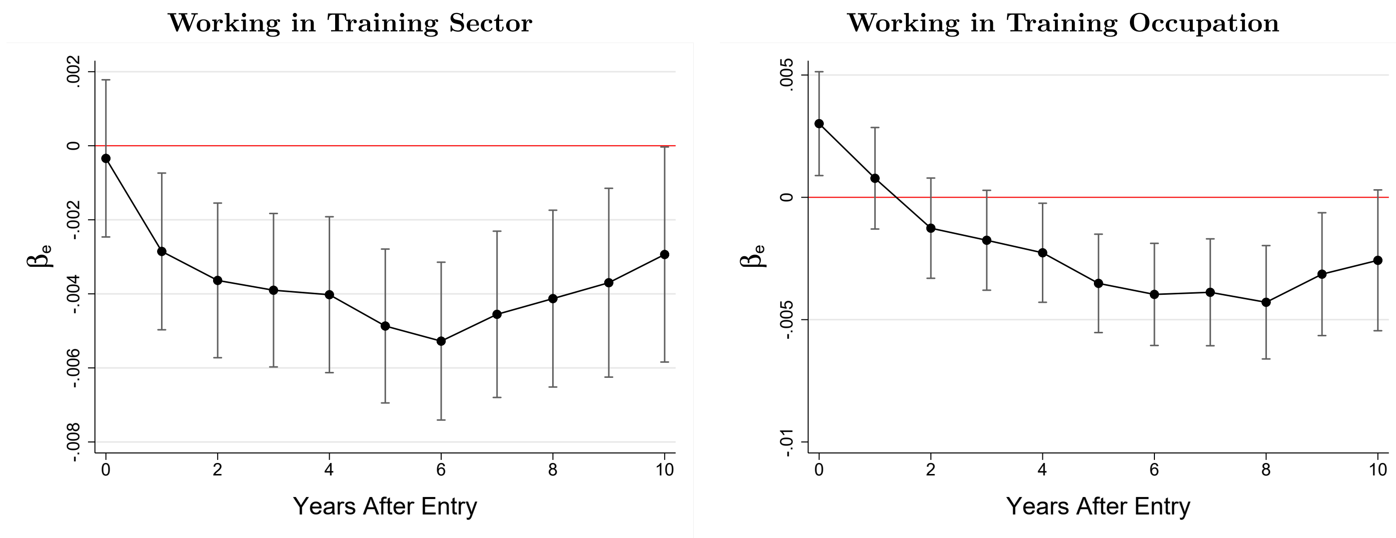
This appendix shows results obtained by re-estimating Equation (2.4.1) using actual rather than predicted year of entry. Actual year of entry is defined by the last day of vocational training.

Figure B.1: The Effect of Entry Conditions (U_{osc}) on Early Career Outcomes Using Actual Year of Entry



Notes: Sample size for each specification is 125,363. 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, predicted year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German indicator variable, and a female indicator variable.

Figure B.2: The Effect of Entry Conditions (U_{osc}) on Early Career Mobility Using Actual Year of Entry



Notes: Sample size for each specification is 125,363. 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, predicted year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German indicator variable, and a female indicator variable.

APPENDIX C

Appendix to Chapter 3

C.1 Theoretical Framework

We follow in the tradition of Sjaastad (1962) by modeling migrants as agents who compare the present discounted value of net income streams in destination areas and origin areas. A substantial subsequent literature has built on this starting point with the primary aim of examining migrant selectivity.[1] A subset of the literature explicitly takes account of migration fixed costs.[2] McKenzie and Rapoport (2010) adapt the notation of Chiquiar and Hanson (2005) to consider migration fixed costs that decline in the size of the migrant network at destination, and we follow their formulation. The literature tends to focus on implications of the theory for migrant selectivity (the extent to which the migration decision depends on relative returns to skill across migrant origin and destination). Instead, we focus on a key prediction of this model that has been under-emphasized: that the migration response to changes in the returns to migration will depend on the size of migration fixed costs. Because it is not our focus, we suppress consideration of migrant selectivity.

[1] Key previous works include Borjas (1987) seminal adaptation of the Roy (1951) model, as well as Greenwood (1985), Taylor (1987), Borjas (1991), Stark (1991), Chiswick (1999), Beine et al. (2001), Feliciano (2005), Chiquiar and Hanson (2005), Orrenius and Zavodny (2005), Clark et al. (2007), Ibarraran and Lubotsky (2007), Beine et al. (2008), Dolfin and Genicot (2010), McKenzie and Rapoport (2010), Akee (2010), Abramitzky et al. (2012), Ortega and Peri (2013), Bertoli et al. (2013), and Bertoli et al. (2016).

[2] Key works in the literature that explicitly consider the fixed cost of migration to be a central aspect of the migration decision include Borjas (1987), Carrington et al. (1996), Chiquiar and Hanson (2005), Ibarraran and Lubotsky (2007), Gathmann (2008), McKenzie and Rapoport (2010), Grogger and Hanson (2011), Bertoli et al. (2013), Belot and Hatton (2012), Bertoli and Rapoport (2015), Kennan and Walker (2011), Chen et al. (2019), and Boustan et al. (2017). Empirical studies on the association between pre-existing migrant stocks and subsequent migration flows include Winters et al. (2001), Clark et al. (2007), Pedersen et al. (2008), Zavodny (1997), Hanson and McIntosh (2012), McKenzie and Rapoport (2010), Collins (1997), Collins and Wanamaker (2015), and Orrenius and Zavodny (2005).

C.1.0.1 Basic setup

Consider an individual in their "home" (non-U.S.) country deciding whether or not to migrate to the "foreign" country (the U.S.). Let w_h be the present value of the flow of the individual's future income in the home country, and w_f be the corresponding value for the foreign country. To simplify matters, we consider a one-time decision to migrate permanently to the foreign country. Migration involves a fixed cost C, which we presume is a function of the migrant's network n. Let the fixed cost of migration be lower when an individual has a larger migrant network, meaning $C' < 0$. Express migration costs in "time-equivalent" units (as a fraction of the present value of income flows in the foreign country):

$$\pi(n) = \frac{C(n)}{w_f}.$$

Assuming π is small, individuals migrate if:

$$\ln(w_f) - \pi(n) > \ln(w_h).$$

Because migration costs $C(n)$ decrease with migrant network size, so do time-equivalent migration costs $\pi(n)$. Express the natural log of time-equivalent migration costs as $\ln(\pi) = \mu - \gamma n$, where $\gamma > 0$. Now, the condition for migration can be written as:

$$\ln(w_f) - e^{\mu - \gamma n} > \ln(w_h). \tag{C.1}$$

In this set-up, we can represent the individual's choice graphically. In Figure C.1, the size of the migrant network n is on the horizontal axis, while the vertical axis is monetary value in logs. The right hand side of inequality (C.1) is the solid line at $\ln(w_h^0)$, which is horizontal because home-country income does not depend on network size. The left hand side of inequality (C.1) is represented by the solid upward-sloping curve: because migration costs decline in n, the net present value of the income stream in the foreign country rises in n. Individuals who choose to migrate are those with network size above the threshold $\underline{n}^0$, whose migration fixed costs are low enough to make migration worthwhile.

Now consider the impact of a negative shock to home economic conditions, so that the present value of the home income stream declines from w_h^0 to w_h^1. (In the empirics, we will interpret hurricanes as having this effect.) This is represented by a downward shift of the horizontal line representing the value of not migrating to the horizontal dashed line at $\ln(w_h^1)$.

C.1.0.2 Negative home shock does not affect migration costs

If the negative home-country shock has no effect on migration costs, the analysis is straightforward. This leads a new set of individuals to choose to migrate, since now the threshold network size for migration has fallen from $\underline{n}^0$ to $\underline{n}^1$ in Figure C.1. Within the population of those who had not migrated prior to the negative shock, those migrating will be those with differentially higher network size (in the range from $\underline{n}^1$ to $\underline{n}^0$). Those with lower network size (below $\underline{n}^1$) will continue to remain in the home country.

C.1.0.3 Negative home shock affects migration costs

The hurricane's effect becomes ambiguous if the negative shock to the home economy does affect migration costs. Imagine simply that the negative shock, a hurricane, raises the natural log of time-equivalent migration costs by H, so that $\ln(\pi) = \mu - \gamma n + H$. We can rewrite this as $\pi = e^{\mu-\gamma n+H}$, so the condition determining migration becomes:

$$\ln(w_f) - e^{\mu-\gamma n+H} > \ln(w_h) \tag{C.2}$$

It now becomes possible for a negative shock to either increase or decrease migration. These possibilities are also represented in Figure C.1. A negative shock now also leads the curved line (the left hand side of inequality C.2) to shift downward. If the increase in the log of time-equivalent migration costs is low (say H_{lo}), the downward shift is small, illustrated by the shift to the dashed curve labeled $\ln(w_f) - e^{\mu-\gamma n+H_{lo}}$. The net effect is still for migration to increase: the threshold network size for migration falls from $\underline{n}^0$ to $\underline{n}^2$. On the other hand, if the shift is large enough (such as to the dotted curve in Figure C.1, representing a larger increase in the log of time-equivalent migration costs H_{hi}), then, migration can actually decline—the threshold for migration actually rises from $\underline{n}^0$ to $\underline{n}^3$.

C.1.0.4 Migrant networks provide insurance

Now consider the possibility that migrants can provide insurance in the form of remittances in response to negative shocks such as hurricanes.

In the context of our theoretical framework, we can represent the insurance provided by the migrant network as replacing a fraction of home-area income losses caused by a negative shock. The income loss due to a negative shock is the difference between pre- and post-shock home wages, $\ln(w_h^0) - \ln(w_h^1)$. Let $\alpha(n)$ be the fraction of this loss that is replaced by migrant remittances. Let $\alpha' > 0$, to represent that the extent of insurance (the fraction of the loss replaced) is larger when the migrant network is larger (as a share of home country population). This is sensible, because when migrant networks are larger, more individuals in the home country should have a migrant social network member, and the financial burden of supporting disaster-affected home-country residents can be spread across more migrants.

After remittances from migrants in the wake of a negative shock, the relevant measure of well being in the home country is log wages plus remittances, $\ln(w_h^1)+\alpha(n)\left[\ln(w_h^0)-\ln(w_h^1)\right]$. Log wages plus remittances are presented graphically in Figure C.2 as the upward-sloping heavy dashed line between the horizonal lines at $\ln(w_h^1)$ and $\ln(w_h^0)$.[3] It is now the intersection of this line with the foreign net wage function that determines the threshold network size above which people in the home country choose to migrate, in this case $\underline{n}^4$. New migration occurs for individuals with migrant networks in the range of $\underline{n}^4$ to $\underline{n}^0$.

This range is smaller than the range of new migrants if their migrant networks did not send remittances in response to negative shocks (that range is $\underline{n}^1$ to $\underline{n}^0$). Therefore, the possibility of migrants sending shock-coping remittances attenuates the effect of shocks on new migration. The attenuation can be arbitrarily large. As $\alpha(n)$ approaches 1, new migration in response to home-country shocks goes to zero.

C.1.0.5 In sum

Theoretical predictions are ambiguous: negative shocks to economic conditions in the home country could increase migration by increasing the return to migration. It is also possible for negative home-country shocks to *reduce* migration, if such shocks themselves increase the fixed costs of migration, or reduce ability to pay migration fixed costs. Migrants' ability to send remittances in response to negative shocks introduces further ambiguity, potentially attenuating further any positive migration response to hurricanes.

[3]For the purpose of this figure, we have specified $\alpha(n)$ as a logistic function bounded between $\ln\left(w_h^1\right)$ and $\ln\left(w_h^0\right)$.

Figure C.1: Negative Shocks and Migration

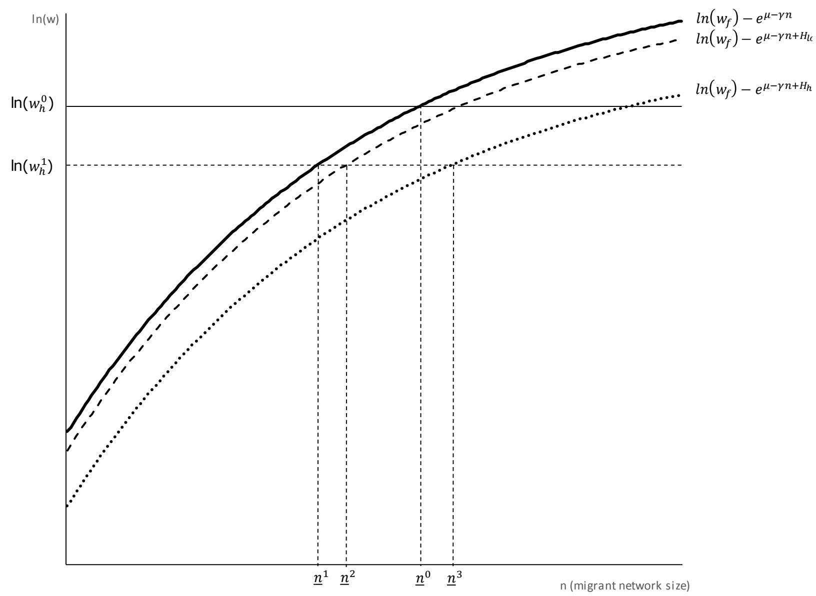
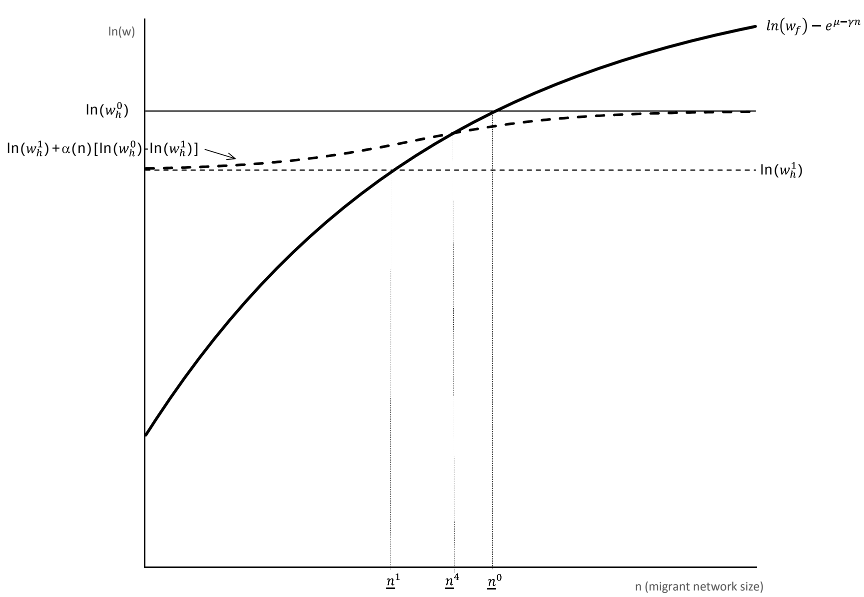


Figure C.2: Negative Shocks and Migration, When Migrants Provide Insurance



C.1.1 Construction of the Hurricane Index

The damage caused by hurricanes depends on the intensity of the hurricane (in particular, wind speed). In addition, hurricanes should cause more damage if they strike in more populated areas. An index H_{jt} for country j in year t that has these features is as follows:

$$H_{jt} = \frac{\sum_i \sum_s x_{isjt}}{N_{jt}}$$

where x_{isjt} is a measure of person i's "affectedness" by hurricane s in country j, year t. Affectedness is summed over hurricanes and over individuals, and then divided by total population N_{jt}. We define a person's hurricane "affectedness" in a particular storm is a nonlinear function of the wind speed to which the individual was exposed.[4] There is no data source for individual-level hurricane affectedness (x_{isjt}), and so we approximate the numerator in the hurricane index H_{jt} by estimating wind speeds at evenly-spaced points on a country's land area, and combining this with population estimates at these points.

The first step in this process is the creation of a 0.25 by 0.25 degree grid of latitude and longitude points that fall inside large countries and 2.5 minute by 2.5 minute latitude and longitude points that fall inside small countries.[5] Then, we predict the wind speed of each hurricane segment (a connected set of points from the best tracks) using a model from Dilley et al. (2005):

$$pw_{gjst} = \mathbb{1}\{w_{gjst} > 33\}\left[33 + (w_{gjst} - 33)\left(1 - \frac{d_{gjst}}{prad_{gjst}}\right)\right] \tag{C.3}$$

Here, pw_{gjst} is the predicted wind speed (in knots) felt at grid point g in country j from storm s, w_{gjst} is the actual wind speed recorded at the beginning of the storm segment from the best track, d_{gjst} is the distance between the grid point and the storm segment, and $prad_{gjst}$ is the predicted radius of the hurricane segment, where we only calculate pw_{gjst} for grid points for which $d_{gjst} < prad_{gjst}$.[6]

As an example of a pw_{gjst} calculation, consider Figure C.3, which shows both the best track for Hurricane Mitch and its radius of hurricane-force winds. The black grid points are points in Honduras that did not experience hurricane-force winds, while the yellow grid

[4]The pressure exerted by winds is commonly modeled in climatology as rising in the square of wind speed (Emanuel, 2005).

[5]"Large" countries are defined as those that have at least two 0.25 by 0.25 degree grid points, and "small" countries are defined as the converse of this large set of countries. Country delineations are provided by the `maptools` package in *R*.

[6]$prad_{gjst}$ is calculated based on a model of wind-speed decay given distance from the hurricane, as in Dilley et al. (2005).

points did experience such winds. Consider the grid point highlighted in blue, $g*$. We first calculate the shortest distance between this point and the nearest storm segment from the Hurricane Mitch best track, represented by the blue line from the point to the storm best track. This distance is $d_{g*,Honduras,Mitch,1998}$. Then, since this distance is less than the predicted radius ($prad_{g*,Honduras,Mitch,1998}$) of the closest storm segment—represented by the red width surrounding the storm best track—we proceed to calculating $pw_{g*,Honduras,Mitch,1998}$ using Equation (C.3), where wind speed also comes from this nearest storm segment.

The effect of hurricane s at grid point g in country j during year t is then:

$$x_{gjst} = \mathbb{1}\{pw_{gjst} > 33\} \left[\frac{(pw_{gjst} - 33)^2}{(w^{max} - 33)^2} \right]$$

where w^{max} is the maximum wind speed observed in the dataset (166.65 knots). Finally, to aggregate this information up to a population-weighted, country-year level, we utilize the 1990 gridded population data for each 0.25 degree and 2.5 minute grid point from Columbia University's Socioeconomic Data and Applications Center (SEDAC).[7] This allows us to create the final hurricane index H_{jt} for country j in year t:

$$H_{jt} = \frac{\sum_g \sum_s x_{gsjt} N_{g,1990}}{\sum_g N_{g,1990}}$$

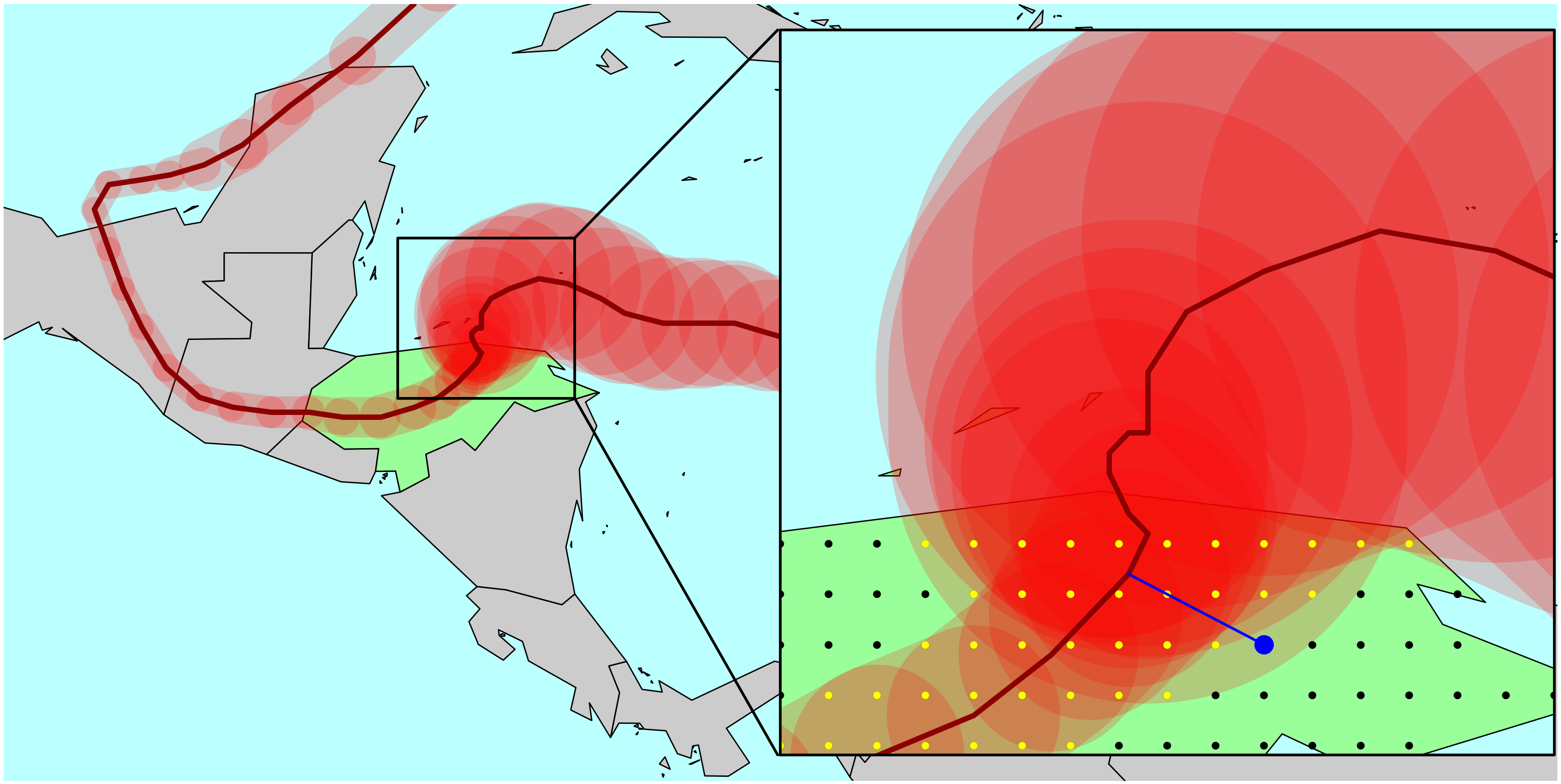
where $N_{g,1990}$ is the grid point's population 1990 given from SEDAC. That is, we sum up a measure of how affected each country grid point is by each storm across storms to get each grid point's affectedness, then take a weighted sum of these grid points (by population), to obtain the intensity-weighted hurricane events per capita measure.

Three additional issues merit mention with respect to the construction if H_{jt}. First, 1990 is the earliest date for which we have access to worldwide gridded population from SEDAC. Since our sample period is 1980 to 2004, there is the potential for our estimate to reflect reverse causality created by hurricane-induced migration from grid points affected in the 1980s. In this case, within-country areas most likely to be hit by hurricanes would receive weights that are too low, creating values of H_{jt} that are also too low. This reverse causality would generate a downward bias on our estimated effect of hurricanes on emigration, making our estimates conservative. Second, because of a lack of reliable wind speed information in the best tracks, we only have H_{jt} for countries affected by North Indian basin hurricanes starting in 1981 and South Indian and South Pacific basin hurricanes starting in 1983. We therefore drop any observations from countries affected by North Indian hurricanes prior to

[7]http://sedac.ciesin.columbia.edu/data/collection/gpw-v3

1981 and any countries affected by southern hemisphere hurricanes prior to 1983. Finally, the hurricane season in the southern hemisphere starts in November. For ease of comparison within year across countries, we include hurricanes from November and December in the following year's hurricane index for countries in the southern hemisphere.

Figure C.3: Hurricane Mitch over Honduras



Source: Unisys Weather data (http://weather.unisys.com/hurricane/) processed in R.

C.2 Census Bureau: 1980 Stocks and 1980-2004 Inflows

In order to estimate migration inflows, we construct retrospective estimates using the 2000 Census and 2005 through 2015 ACS 1-year files. This methodology utilizes the combination of questions that asks survey respondents where they were born and what year they came to live in the United States. Aggregating person weights by country of birth and year of entry within a given survey thus generates a set of initial country-year migration inflow estimates for all years before the survey. That is,

$$M_{jt}^{\text{survey}} = \sum_{i \in \text{survey}} \left[\mathbb{1}\{\text{Person } i \text{ is from country j}\} \times \mathbb{1}\{\text{Person } i \text{ entered in year } t\} \times pwgt_i^{\text{survey}} \right]$$

where i is an individual respondent to a given survey (2000 Census, 2005 ACS, 2006 ACS, ..., 2015 ACS) and $pwgt_i$ is that individual's person weight assigned by that survey. Given the sheer sample size of the 2000 Census, we use these aggregated estimates to infer migration inflows for the years 1980 through 1999. In order to extend our annual sample to 2004 while retaining relatively low levels of noise in our estimates, we average the estimates generated by the 11 ACS surveys from 2005 through 2015 for the years 2000 to 2004:

$$M_{jt} = \begin{cases} M_{jt}^{\text{2000 Census}} & \text{if } t \leq 1999 \\ \frac{1}{11} \sum_{r=2005}^{2015} M_{jt}^{\text{ACS year } r} & \text{if } 2000 \leq t \leq 2004 \end{cases}$$

Given this methodology, the key advantage of access to confidential data comes in estimating migration inflows from small countries. Use of smaller Census samples available publicly can generate accurate estimates of migrant inflows for large countries with many immigrant survey respondents that appear consistent across surveys. However, small countries, many of which are heavily affected by hurricanes, often either contain relatively few observations per year of entry or are aggregated into categories like "Other Caribbean" in publicly available data. This would generate substantial imprecision in the annual migration estimates. The 1-in-6 count provided by the confidential 2000 Census and aggregation of multiple ACS surveys alleviates this issue.

Despite this novel use of confidential data, a few concerns merit further consideration with this methodology. First, by using the 2000 Census and to look at inflows as far back as 1980, we are focusing on permanent migrants to the U.S.—those who remain living in the U.S. (or connected enough through repeated return trips) to be enumerated by the Census Bureau up to 20 years after arrival. As estimates from the 2000 Census roll forward from the starting point of 1980, underestimation due to death and re-migration give way to overestimation of permanent migrants due to the presence of more temporary migrants closer to the year 2000. Nonetheless, Passel and Suro (2005) find that this methodology tracks other migration estimates well for large countries in publicly available data, and thus we find its broader use with confidential data to be appropriate. Furthermore, as described in Section 3.3.3.2, we complement these estimates with data from the DHS that counts legal *permanent* resident entries *at the time of entry* in order to ensure that our results are robust to these concerns. In this sense, the results from the Census/ACS panel can be viewed as incorporating undocumented and temporary migrant response to hurricanes.

Second, as elucidated by Redstone and Massey (2004), in the presence of circular migration, the interpretation of year of entry provided by survey respondents in the Census is not clear. Specifically, in cases where immigrants reported multiple entries and exits in the New Immigrant Survey, Redstone and Massey (2004) find that 45 percent of immigrants report a "year that they came to live" that was not their first entry, and 54 percent of immigrants report a "year that they came to live" that was not their final entry.[8] The answers to this Census question appear to largely be a combination (across respondents) of first year of entry and the mental decision to make the United States their permanent home. Given the nature of our empirical strategy, we understand this as an issue of interpretation rather than bias. Any effect found on migrant inflows using the Census data should be interpreted as an effect on the decision to stay permanently in the U.S.—including both literal, one-time moves and the decision to turn repeated circular migration into permanent residency in the United States. Furthermore, remaining, pure noise created by inaccuracy in recalling year of entry causes larger standard errors in our coefficient estimates, making our estimates of precision conservative.We also use access to the confidential, full version of the 1980 Census Long Form responses to construct a measure of immigrant stocks from each country in 1980, the base year of our analysis:

$$S_{j,1980} = \sum_{i \in 1980 \text{ Census}} \left[\mathbb{1}\{\text{Person } i \text{ is from country j}\} \times pwgt_i^{1980 \text{ Census}}\right]$$

These estimates have the advantage of producing more accurate stocks for small countries due to the large, 1-in-5 count sample size of the confidential data and do not suffer from either of the concerns of year-by-year migration estimates mentioned above.

[8] The wording "year you came to live in the U.S." used by Redstone and Massey (2004) exactly mimics the Census wording in order to make this comparison.

C.3 Income and Damages in Sending Countries

We establish here that the hurricane index captures events that have tangible, negative consequences in sending countries. We estimate the long-run response of incomes in sending countries to hurricane events, as in Hsiang and Jina (2014). We obtain year-by-year real GDP per capita estimates from the World Bank's World Development Indicators (WDI), enabling us to estimate the long-run effect of hurricanes on income.[9] Following Hsiang and Jina (2014), our regression specification is:

$$g_{jt} = \alpha + \sum_{\ell \neq -1, \ell=-5}^{10} \theta_\ell H_{j,t-\ell} + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{C.1}$$

$$g_{jt} = \alpha + \sum_{\ell \neq -1, \ell=-5}^{10} \alpha_\ell H_{j,t-\ell} + \sum_{\ell \neq -1, \ell=-5}^{10} \alpha_\ell^{stock} (H_{j,t-\ell} \times s_{j,1980}) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \tag{C.2}$$

$$g_{jt} = \log(\text{Real GDP per capita})_{jt} - \log(\text{Real GDP per capita})_{j,t-1}$$

We add the α_ℓ coefficients from Equation (C.1) starting at ℓ =0 to unravel the impulse response of log real GDP per capita to the hurricane index (calibrated to $\sigma_H = 0.02$).

The results are shown in Figure C.4, where we see a robust, long-run effect. Ten years later, a one standard deviation increase in the hurricane index leads to 5 to 10 percent lower in GDP per capita. This kind of permanent economic impact buttresses the notion that hurricanes can cause the kind of permanent migration we observe. We also estimate Equation (C.2) in order to determine whether the interaction between hurricanes in sending countries and immigrant stocks in the United States alters the impact of hurricanes on sending country economic activity. Figure C.5 shows that the impulse responses of GDP per capita implied by α_ℓ^{stock} coefficients does not contain any evidence of such an interaction.[10] Meanwhile, constructing the impulse response based on the α_ℓ coefficients from Equation (C.2) yields similar results to doing so without the stock interaction effect, as in Equation (C.1). This strengthens our interpretation of $s_{j,1980}$ as a pure pull factor for potential migrants. That is, the stock operates as a network effect, facilitating migration as a response to hurricanes, but does not appear to alleviate damages at home to the point of dampening the push factor caused by hurricane-induced income losses.

Note that, for completeness, we can construct similar graphs for our main outcome of interest, migration m_{jt}. Figures C.6 and C.7 thus show the results of estimating Equations (C.1) and (C.2) with m_{jt} as the outcome. Because our primary outcome of interest is

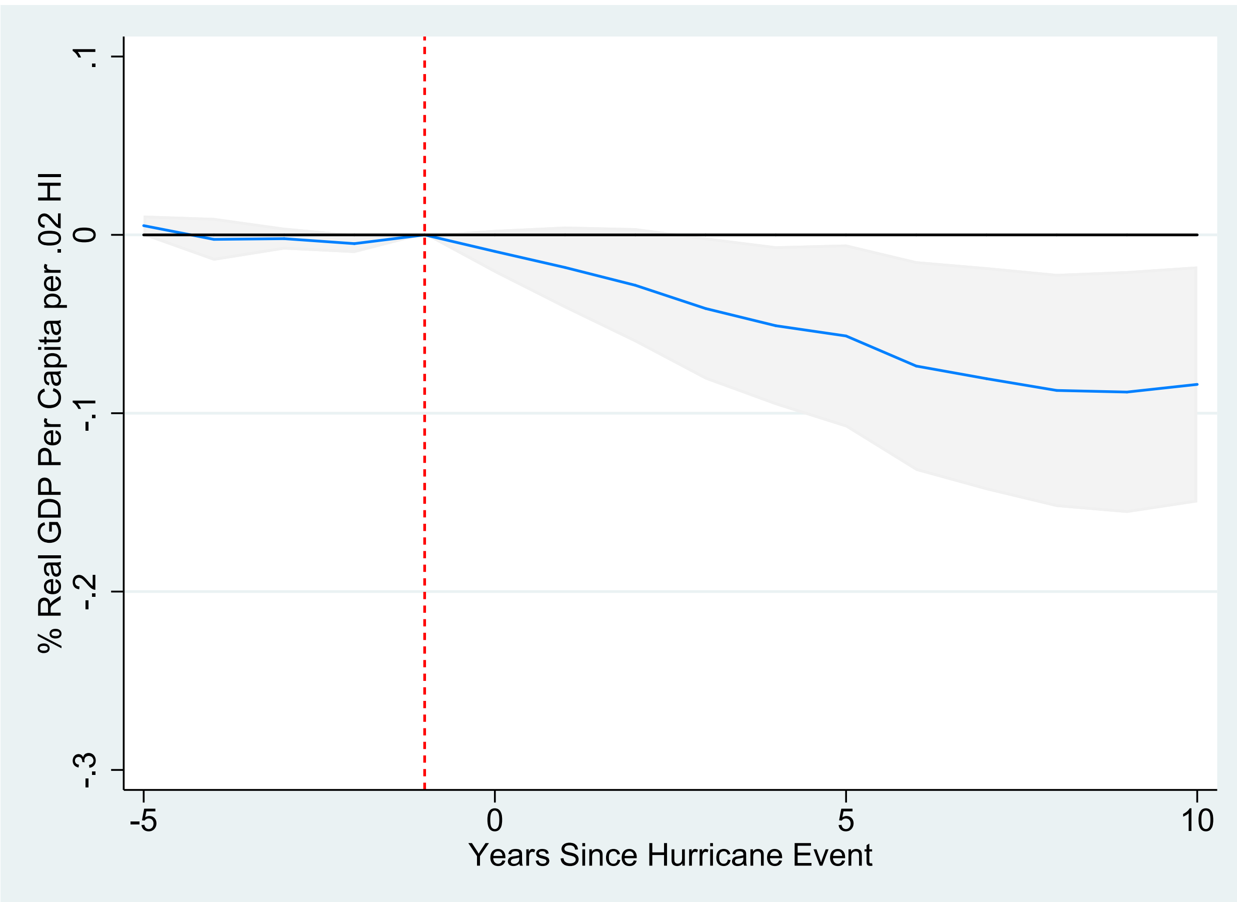
[9] See Table C.2 for summary statistics.

[10] The impulse responses for the stock interaction effect are multiplied by the standard deviation of $s_{j,1980}$, 0.03 to retain consistency in units.

migration flows, we do not cumulate responses in this case, and instead directly plot the resulting coefficients. As seen from the estimates of θ_ℓ in Figure C.6, the only detectable migration response to hurricanes appears in the year of the hurricane itself. This justifies our use of a specification without lags in the hurricane index in Equations (3.4.1) and (3.4.2). Splitting this effect into its interaction through previous migrant stock and a direct, level effect, Figure C.7 further reveals that this "Year 0" response is entirely driven by the interaction effect. The lack of response prior to a given hurricane event serves as another placebo test. The data fail to reject the null hypothesis of no pre-trends.

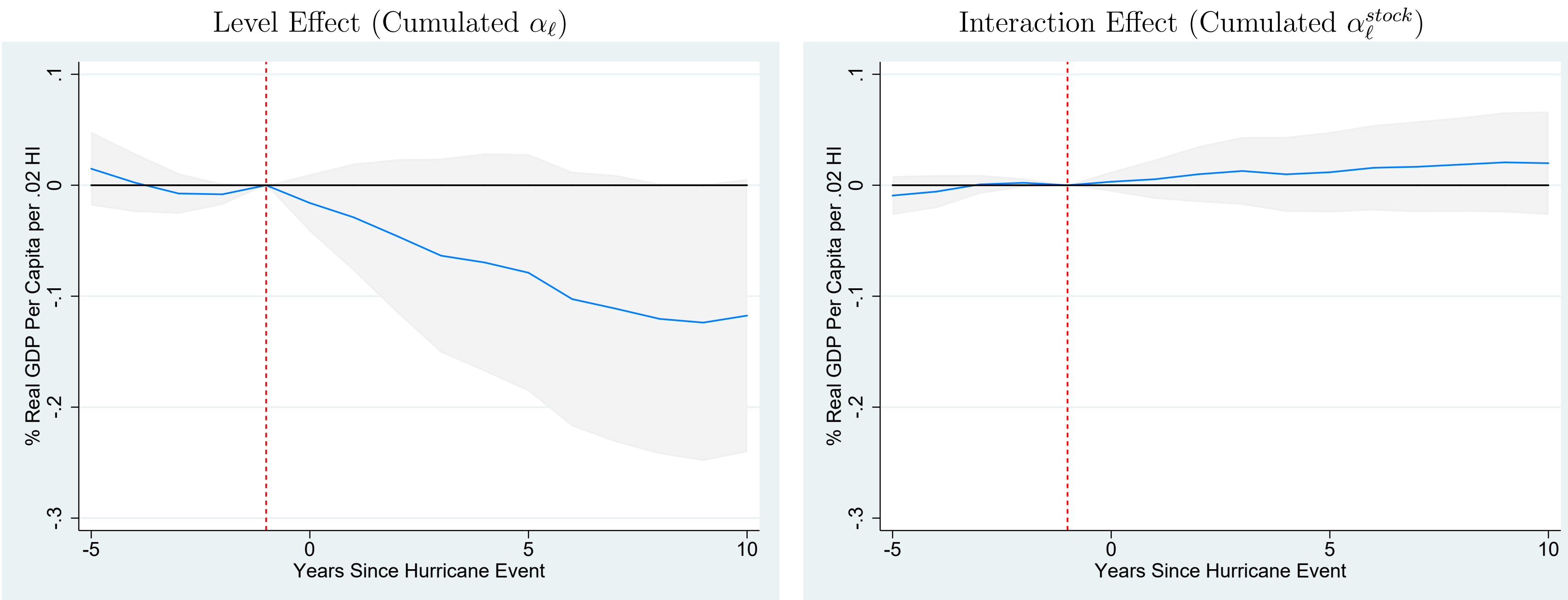
Another source of data on impact in sending countries is EM-DAT, as described in Section 3.3. Table C.1 presents results from estimating Equations (3.4.1) and ((3.4.2)) with damages as a proportion of 1980 real per capita GDP, as well as deaths, injuries, and total number of people affected as a proportion of 1980 population due to meteorological disasters as outcomes. Table C.1 shows a strong, robust effect of hurricanes on damages reported in potential sending countries. A one standard deviation increase in hurricane incidence in a given year corresponds to a 7.80 percent increase in damages as a proportion of 1980 GDP. As with our results from estimating Equation (C.2), we find no evidence of a stock interaction effect that mitigates the effect of hurricanes on sending country damages.

Figure C.4: Long Run Effect of Hurricanes on GDP Per Capita (Cumulated θ_ℓ)



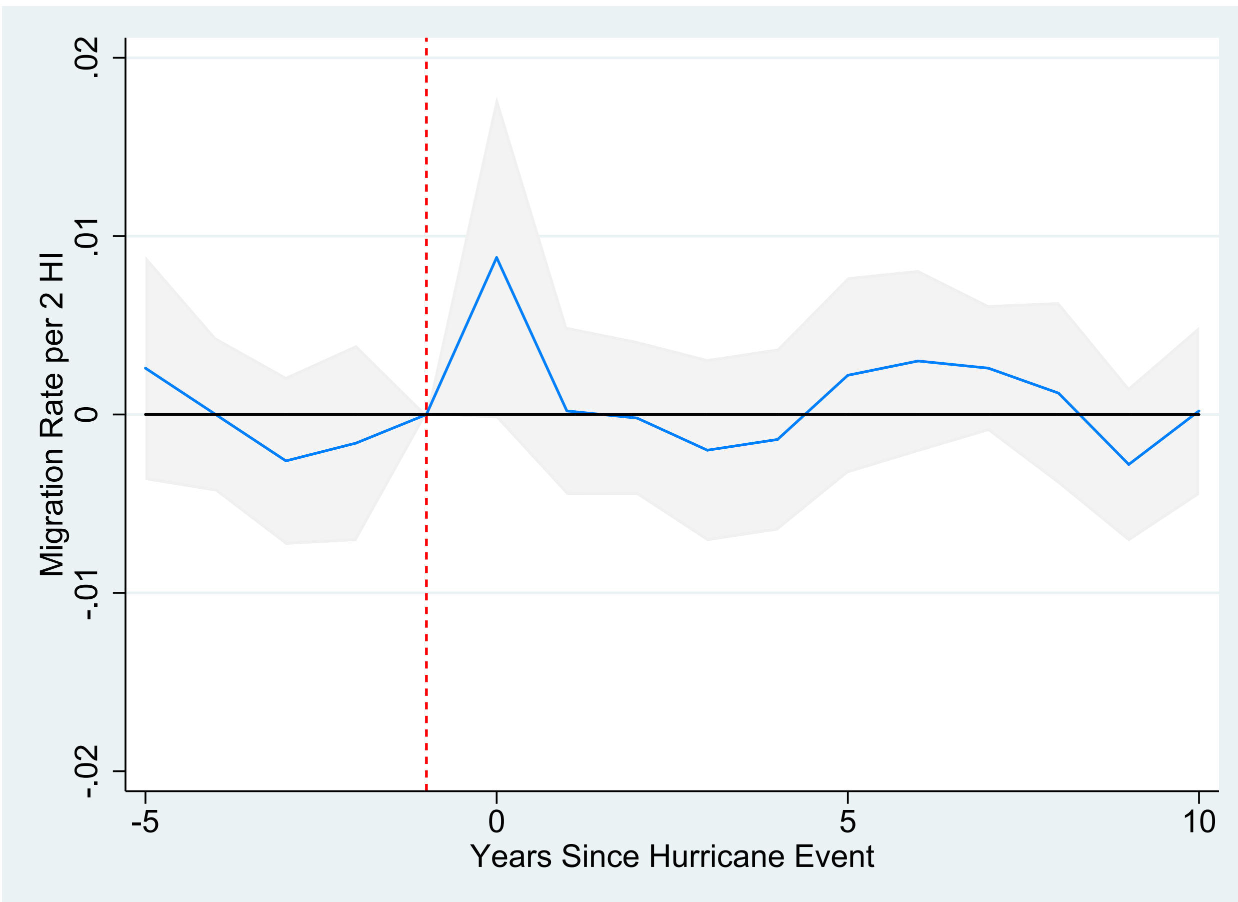
Notes: This figure represents an impulse response function generated by adding the coefficients α_ℓ that are estimated using Equation (C.1) before being multiplied by the standard deviation of the hurricane index.

Figure C.5: Long Run Effect of Hurricanes on GDP Per Capita, with Stock Interaction



Notes: Each figure represents an impulse response function generated by adding the coefficients α_ℓ (Left Panel) and α_ℓ^{stock} (Right Panel) that are estimated using Equation (C.2) before being multiplied by the standard deviation of the hurricane index and, in the case of the Right Panel, the standard deviation of the 1980 immigrant stock as a proportion of 1980 sending country population.

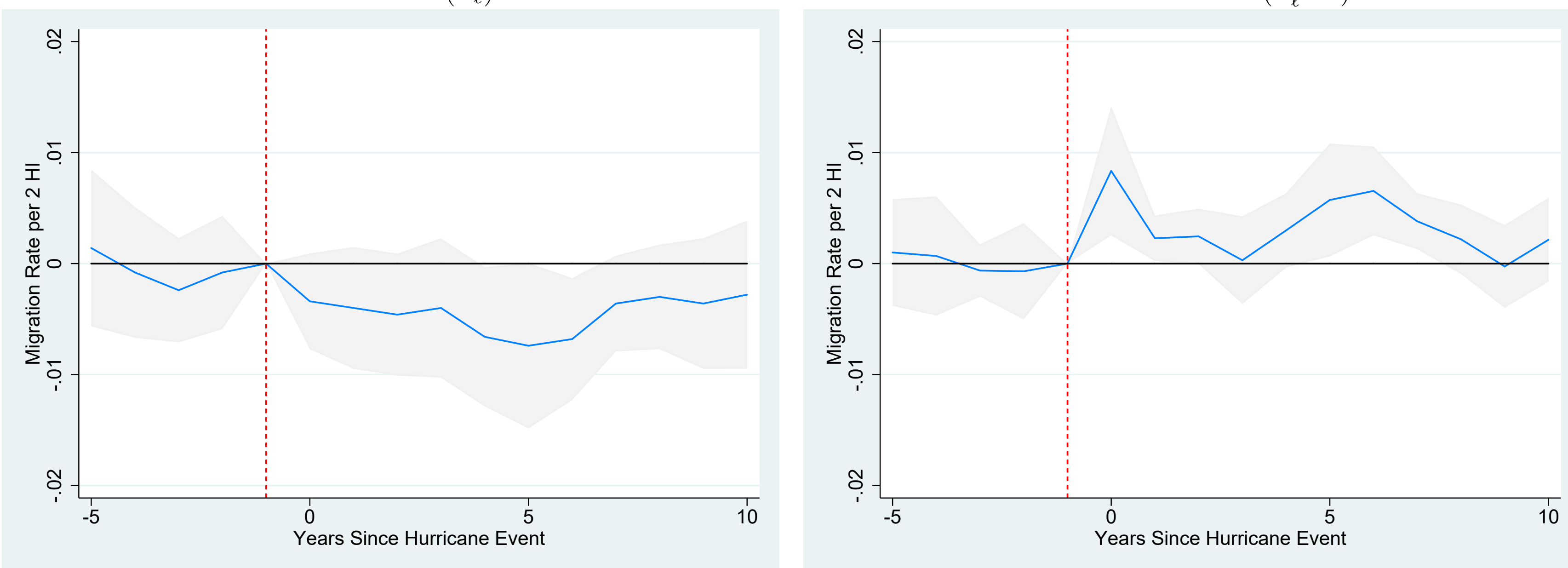
Figure C.**6:** Long Run Effect of Hurricanes on Migration (θ_ℓ)



Notes: This figure plots the coefficients that are estimated using Equation (C.1) before being multiplied by 100 times the standard deviation of the hurricane index.

Figure C.7: Long Run Effect of Hurricanes on Migration, with Stock Interaction

Level Effect (α_ℓ) Interaction Effect (α_ℓ^{stock})



Notes: This figure plots the coefficients that are estimated using Equation (C.2) (Bottom Panel) before being multiplied by 100 times the standard deviation of the hurricane index and, in the case of the Right Panel, the standard deviation of the 1980 immigrant stock as a proportion of 1980 sending country population.

Table C.1: The Effect of Hurricanes on Sending Country Damages, 1980-2004

			As Proportion of 1980 Population					
Outcome:	$\frac{\text{Damages}}{\text{1980 GDP}}$	$\frac{\text{Damages}}{\text{1980 GDP}}$	Deaths	Deaths	Injured	Injured	Affected	Affected
Hurricane Index(t)	3.8980***	4.2642***	0.0004**	0.0002	0.0009	0.0004	0.3492***	0.3465**
	(1.0114)	(1.5097)	(0.0002)	(0.0001)	(0.0007)	(0.0005)	(0.1132)	(0.1390)
Hurricane Index(t) $\times$		-8.4283		0.0040		0.0128		0.0625
1980 Proportional Immigrant Stock		(21.5619)		(0.0042)		(0.0158)		(2.3058)
Country-Years	3900	3900	3900	3900	3900	3900	3900	3900
R^2	0.0987	0.0987	0.0443	0.0466	0.1193	0.1194	0.0878	0.0878
Countries	159	159	159	159	159	159	159	159

Notes: Each column refers to a different OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). Outcome variables obtained from the Center for Research on Epidemiology of Disasters International Disaster Database. "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

C.4 Control Variables and their Sources

This section describes the sources and construction of control variables, used both to test robustness of the results found in Table 3.3 and to highlight mechanisms. Summary statistics for these variables are presented below in Table C.2. Note that we have not been given permission to publish summary statistics on $HHI_{j,1980}$ (described below).

C.4.1 GDP Per Capita: Avakaov (2015)

Avakov (2015) provides real GDP per capita estimates for the 159 land areas in our sample, including those that were not yet countries in 1980. These data allow us to assess robustness of our results to the inclusion of GDP per capita as a control, as well as how the interaction between migration networks and hurricanes change with sending country income.

C.4.2 World Bank World Development Indicators (WDI)

Beyond GDP per capita, we seek to assess robustness against a bevy of sending country characteristics that could mitigate the relationship between hurricanes, migrant networks, and migration to the U.S. The WDI aggregates many of these variables into one database, including remittances as a proportion of GDP and domestic credit as a proportion of GDP for 142 of the 159 countries in our sample. Because these variables are often missing for a given country in the year 1980, we employ a country-level average from 1970 to 1979 (throwing out missing observations) for these variables.

C.4.3 United Nations Population Division (UNDP): non-U.S. Immigrant Stocks

The UNDP estimates the stock of immigrants from a majority of our sending countries living in various destination countries starting in 1990. They construct this data by combining governmental estimates of immigration and emigration from each country.[11] These estimates allow us to test whether the primacy of the U.S. as a destination for a given source country affects our results. That is, if a source country is well-connected in multiple destination countries, the model presented in Section 3.1 implies that its hurricane-induced migrants would split their locational decisions between these countries.

[11] For example, the DHS data is used to generate immigrant stock estimates for the United States. The data can be found at https://esa.un.org/unmigration/.

Table C.2: Summary Statistics of Control Variables

			Percentile						
	Mean	Std. Dev.	10	25	50	75	90	N	Source
1980 Real GDP Per Capita	8,158	14,776	903	1,554	3,983	9,094	18,691	159	Avakov (2015)
log Real Meteorological Monetary Damages	1.44149	3.81300	0	0	0	0	9.11451	2,983	CRED
Meteorological Monetary Damages per 1980 GDP	0.00001	0.00019	0	0	0	0	<0.00001	2,975	CRED
Meteorological Disaster Deaths per 1980 Population	0.00001	0.00009	0	0	0	0	<0.00001	2,975	CRED
Meteorological Disaster Injuries per 1980 Population	0.00005	0.00191	0	0	0	0	<0.00001	2,975	CRED
Meteorological Disaster Affected Persons per 1980 Population	0.00732	0.05602	0	0	0	0	0.00062	2,975	CRED
g_{jt}: Real GDP per capita growth	0.00142	0.15438	-0.15265	-0.06430	0.01186	0.07772	0.14715	3,221	WDI
Remittances as a Prop. of 1980 GDP (1970-1980 Average)	3.54	9.79	0.04	0.22	0.84	2.93	6.49	74	WDI
Dom. Credit as Prop. of 1980 (1970-1980 Average)	21.82	15.75	6.10	12.94	18.90	28.26	40.18	104	WDI
Non-U.S. Stock of Immigrants as Prop. of 1980 Population	0.11464	0.18869	0.00959	0.01724	0.05316	0.12502	0.30538	158	UNDP
Land Area (sq. km)	591,653	1,431,563	360	5,130	108,430	581,540	1,280,000	159	*R* `maptools`
Distance from Capital City to D.C. (km)	9,051	4,150	2,936	5,837	9,968	12,391	13,906	159	*R* `maptools`

Notes: Historical real GDP data obtained from Avakov (2015). CRED data obtained from the Center for Research on Epidemiology of Disasters International Disaster Database. WDI data obtained from the World Bank. *R* `maptools` contains land area, and is also used to calculate Distance to Washington D.C.

C.4.4 Land Area and Distance to the U.S.

Proximity and the absence of undamaged land mass available within country can facilitate hurricane-induced migration to the U.S. In order to both understand the magnitude of these mechanisms and ensure they are not wholly driving our results, we construct two measures. The first the log of land area in squared kilometers and the second is the distance from each country's capital city to the U.S.—meant to mimic distance measures used in standard trade gravity models (e.g., Feenstra et al. (2001)). Each is constructed using data available in the `maptools` package in *R* (distance to Washington D.C. is calculated using this package after obtaining latitude and longitude coordinates of capital cities from Google Maps).[12] For a subset of countries without land area information available in this package, we employ land area information provided in the WDI.

C.4.5 Damages: Center for Research on Epidemiology of Disasters (CRED)

In order to verify that our independent hurricane index corresponds to immediate damages in potential sending countries on a level that could prompt immigration to the United States, we use data from EM-DAT: the Center for Research on Epidemiology of Disasters (CRED) International Disaster Database.[13] These estimates include monetary damages in nominal USD and the number of deaths, injuries, and total number of people affected by meteorological disasters in a given country and year. The sources of disaster impact data include national governments, UN agencies, non-governmental organizations, insurance companies, research institutes, and the media. In order to put the monetary damages in real terms (2010 USD), we employ the U.S. GDP deflator from the World Bank's World Development Indicators. The use of these data allow us to establish something akin to a "first stage" effect, that our objective hurricane index corresponds to monetary and human damages felt on the ground in potential sending countries. Additionally, we report damages as a proportion of 1980 real GDP. We obtain the denominator from Avakov (2015), who collects historic data for land masses small enough to cover our entire country sample.

C.4.6 Restricted-Access Census Bureau: 1980 Immigrant Concentration Index

In theory, we may expect that immigrant communities that are particularly concentrated in U.S. areas that are close to hurricane-hit countries—Miami, for example—are particularly suited to absorb hurricane-induced inflows. In order to test whether our stock interaction effect is solely driven by such concentrated communities, we construct a Herfindhal-style

[12]Source data for the `maptools` project is available from https://github.com/nasa/World-Wind-Java/tree/master/WorldW

[13]http://www.emdat.be

concentration index:

$$HHI_{j,1980} = \sum_c \left(\frac{S_{jc,1980}}{\sum_c S_{jc,1980}} \right)^2$$

where c represents a U.S. county and $S_{jc,1980}$ is the number of immigrants from country j living in county c in 1980. Note that the denominator is the same as $S_{j,1980}$ in this chapter's notation. The ability to construct this variable at the granular, county level comes from access to restricted-use Census Bureau data.

C.4.7 Populations: United Nations and U.S. Census Bureau International Data Base

Finally, in order to make country-year observations comparable, we use population data from the set of potential sending countries in our base year, 1980. For this, we used data publicly available data from the United Nations and the U.S. Census Bureau's International Data Base, which between them cover our entire sample. For most of the countries in our sample, estimates of the 1980 population were available from both sources, in which case we took a simple average. These 1980 population estimates are then used as denominators for our final migration inflow outcome variables and our 1980 stock estimates:

$$m_{jt} \equiv \frac{M_{jt}}{N_{j,1980}}$$
$$s_{j,1980} \equiv \frac{S_{j,1980}}{N_{j,1980}}$$

m_{jt} is our main outcome of interest from the data constructed using confidential data from the U.S. Census Bureau.

C.4.8 Predicting the 1980 Stock

We motivate the potential need for these predetermined control variables by using them to predict our interacting variable of interest: $s_{j,1980}$. Table C.3 presents the result from this exercise. Unsurprisingly, countries that are closer to the U.S. had higher proportional immigrant stocks in 1980. Somewhat surprisingly, larger countries, countries with more concentrated immigrant populations, and larger countries also featured higher immigrant stocks in 1980. Real GDP per capita, our best indicator for development, has a positive, but not statistically significant effect on 1980 proportional stocks.

Table C.3: Predicting $s_{j,1980}$, the 1980 Proportional Stock

	$s_{j,1980}$
1980 Immigrant Concentration Index (divided by one million)	-0.0360*
	0.0183
Log 1980 Real GDP Per Capita	0.0015
	0.0017
Log 1980 Population	-0.0068***
	0.0022
Remittances as a Prop. of GDP (average in 1970's)	-0.0123
	0.0237
Domestic Credit as a Prop. of GDP (average in 1970's)	0.0079
	0.0063
Land Area (millions of Sq. KM)	0.0020*
	0.0011
Distance from Capital City to D.C. (millions of KM)	-2.3294***
	0.6525
1990 Proportional Stock in non-U.S. countries	0.0451
	0.0316
Indicator: Missing Remittances as Prop. of GDP	-0.002
	0.0042
Indicator: Missing Domestic Credit as a Prop. of GDP	0.0094**
	0.0045
Indicator: Missing p_stock1990 in non-US countries	-0.0098*
	0.0055
Countries	159
R^2	0.4776

Notes: Each column refers to a different OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). Outcome variables obtained from sources described in Section C.4. "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center.* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

C.5 Robustness of Stock-by-Citizenship Results

This section combines Equations (3.4.3) with (3.4.6) to test the robustness of the results presented in Table 3.5. That is, it estimates

$$m_{jt} = \pi_0 + \pi_1 H_{jt} + \pi_2 (H_{jt} \times s_{j,1980}^{\text{citizen}}) + \pi_3 (H_{jt} \times s_{j,1980}^{\text{non-cit}}) + \pi_c (H_{jt} \times c_j) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt} \quad \text{(C.1)}$$

This estimating equation modifies Equation (3.4.3) by adding an additional set of interaction terms with time-invariant control variables. Online Appendix Section C.4 (above) details the construction of each of these variables. Table C.4 displays the results of estimating Equation (C.1) with each individual control variable as well as with the complete set. The estimated coefficients $\hat{\pi}_2$ and $\hat{\pi}_2$ remain stable. The p-vales from the test that $\pi_2 = \pi_3$ are shown in the bottom row. They show that there appears to be a robust effect of the citizen stock of immigrants itself, as opposed to the many factors it may additionally proxy for. When the "kitchen sink" set of controls is included, this result is only strengthened.

Table C.4: Robustness

	Outcome for all columns: Migrants(t) as a Prop. of 1980 Population									
Hurricane Index(t)	-0.0005	-0.0014*	0.0181	0.0080	-0.0013	-0.0019	-0.0005	-0.0016	0.0011	0.0665***
	(0.0009)	(0.0009)	(0.0173)	(0.0050)	(0.0021)	(0.0027)	(0.0010)	(0.0019)	(0.0028)	(0.0239)
Hurricane Index(t) ×	0.4044*	0.4191*	0.4396*	0.3615*	0.4173*	0.3377*	0.4040*	0.4042*	0.5172*	0.6559***
1980 Proportional Citizen Immigrant Stock	(0.2245)	(0.2266)	(0.2552)	(0.2030)	(0.2329)	(0.2023)	(0.2243)	(0.2266)	(0.2976)	(0.2502)
Hurricane Index(t) ×	-0.1444	-0.1491	-0.1631	-0.1311	-0.1388	-0.0773	-0.1449	-0.1300	-0.2225	-0.2736
1980 Proportional Non-Citizen Immigrant Stock	(0.1661)	(0.1663)	(0.1854)	(0.1541)	(0.1685)	(0.1593)	(0.1663)	(0.1653)	(0.2132)	(0.1941)
Hurricane Index(t) ×		0.0046***								-0.0088
Immigrant Concentration Index		(0.0017)								(0.0058)
Hurricane Index(t) ×			-0.0021							0.0023
log(1980 Real GDP Per Capita)			(0.0020)							(0.0566)
Hurricane Index(t) ×				-0.0664*						-0.0664
log(1980 Population)				(0.0400)						(0.0566)
Hurricane Index(t) ×					-0.1111					-0.3682***
[1970s Remittances as Prop. of GDP]					(0.0775)					(0.1149)
Hurricane Index(t) ×					0.0011					-0.0030
𝟙[Missing: Remittances]					(0.0023)					(0.0048)
Hurricane Index(t) ×						-0.0024				0.0032
[1970s Dom. Credit as Prop. of GDP]						(0.0058)				(0.0076)
Hurricane Index(t) ×						0.0022				0.0122**
𝟙[Missing: Dom. Credit]						(0.0030)				(0.0055)
Hurricane Index(t) ×							-0.0015			-0.0082
[Land Area (mil. sq. km)]							(0.0040)			(0.0110)
Hurricane Index(t) ×								0.1144		0.1667
[Distance to U.S. (mil. km)]								(0.1670)		(0.1729)
Hurricane Index(t) ×									-0.0085	-0.0273***
[1990 Prop. non-U.S. Emigrant Stock]									(0.0094)	(0.0091)
Hurricane Index(t) ×									-0.0021	-0.0124***
𝟙[Missing: non-U.S. Emigrant Stock]									(0.0035)	(0.0046)
Country-Years	3,900	3,900	3,900	3,900	3,900	3,900	3,900	3,900	3,900	3,900
Countries	159	159	159	159	159	159	159	159	159	159
p-value: Equal Interaction Effect of Citizen and Non-Citizen Proportional Stock	0.1540	0.1430	0.1660	0.1630	0.1590	0.2400	0.1540	0.1650	0.1460	0.0365

Notes: Each column refers to a different OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equation (C.1). 1970s Domestic Credit as Prop. of GDP and 1970s Remittances as a Prop. of GDP divide averages of non-missing data of Domestic Credit and Remittances from 1970 through 1979 by 1980 GDP. "Migrants," "1980 Proportional Citizen Immigrant Stock," and "1980 Proportinal Non-Citizen Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

C.6 Placebo Tests

In order to verify that the results presented above are not just the result of spurious statistical noise, we test the following model:

$$m_{jt} = p_0 + p_1 H_{j,t+1} + p_2(H_{j,t+1} \times s_{j,1980}) + \eta_j + \delta_t + \phi_j t + \varepsilon_{jt}$$

We should not expect hurricanes in the future to affect current migration if they are unexpected, exogenous events, as the theoretical considerations laid out in Section 3.1 assume. Table C.5 presents the result of this test, and demonstrates that we cannot reject the hypotheses that $p_1 = 0$ or $p_2 = 0$. This buttresses the notion that H_{jt} is causing migration through the negative income and asset shock channels that we propose.

Table C.5: The Effect of Future Hurricanes on Migration—Placebo Test, 1980-2004

	As a Prop. of 1980 Population	
	Migrants(t)	Migrants(t)
Hurricane Index($t+1$)	0.0017	0.0028
	(0.0015)	(0.0020)
Hurricane Index($t+1$) $\times$		-0.0266
1980 Proportional Immigrant Stock		(0.0281)
Country-Years	3,900	3,900
R^2	0.4273	0.4277
Countries	159	159

Notes: Each column refers to a different OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

C.7 TPS Responses to Hurricane Mitch

The following table displays the sensitivity of our main results to dropping Honduras and Nicaragua, countries that were granted TPS status in response to Hurricane Mitch.

Table C.6: Hurricane Mitch Robustness

	Full Sample		Dropping Mitch-Affected	
Panel A: Census, 1980-2004				
Hurricane Index(t)	0.0040**	-0.001	0.0040**	-0.0011
	(0.0020)	(0.0010)	(0.0020)	(0.0010)
Hurricane Index(t) ×		0.1163**		0.1170**
1980 Proportional Immigrant Stock		(0.0451)		(0.0452)
Country-Years	3,900	3,900	3,800	3,800
R^2	0.4319	0.4409	0.4247	0.4341
Countries	159	159	157	157
Panel B: DHS non-immigrant, 1983-2004				
	Full Sample		Dropping Mitch-Affected	
Hurricane Index($t, t-1$)	0.2193***	-0.0627	0.2197***	-0.0608
	(0.0788)	(0.0689)	(0.0788)	(0.0699)
Hurricane Index($t, t-1$) ×		5.7883**		5.7541**
1980 Proportional Immigrant Stock		2.3536		2.3647
Country-Years	2,200	2,200	2,200	2,200
R^2	0.4485	0.4495	0.4489	0.4498
Countries	156	156	154	154
Panel C: DHS LPR, 1982-2004				
	Full Sample		Dropping Mitch-Affected	
Hurricane Index($t, t-1$)	0.0023	-0.0035	0.0024	-0.0034
	(0.0040)	(0.0039)	(0.0040)	(0.0039)
Hurricane Index($t, t-1$) ×		0.1266***		0.1254***
1980 Proportional Immigrant Stock		(0.0402)		(0.0408)
Country-Years	2,600	2,600	2,500	2,500
R^2	0.2954	0.2966	0.2957	0.2969
Countries	156	156	154	154

Notes: Outcome for each specification is the estimated migrant inflows to the U.S. from a given country in year t as a proportion of that country's 1980 population. Each column within a panel refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). Outcomes in Panels B and C obtained from electronic copies of the *Yearbook of Immigration Statistics* (1996-2004) and *Statistical Yearboook of the Immigration and Naturalization Service* (prior to 1996). "Hurricane Index($t, t-1$)" refers to the average of a hurricane index for a given country across years t and $t-1$. "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. LPR: legal permanent resident; "non-imm:" non-immigrant. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

C.8 Analysis With Publicly-Available Data

The following table displays the results of estimating Equations (3.4.1) and (3.4.2) using publicly-available data from the Census Bureau. The large differences in coefficient and standard error estimates display the importance of using restricted-access Census data for the main analyses presented in this chapter.

Table C.7: The Effect of Hurricanes on Migration, Public Data, 1980-2004

Outcome, Estimated from Public Data:	As a Prop. of 1980 Population	
	Migrants(t)	Migrants(t)
	(1)	(2)
Hurricane Index(t)	0.0016	0.0003
	(0.0015)	(0.0019)
Hurricane Index(t)$\times$		0.0267
Public-Data 1980 Proportional Immigrant Stock		(0.0343)
Country-Years	2,215	2,215
R^2	0.3917	0.3921
Countries	97	97

Notes: Each column refers to a different OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equations (3.4.1) and (3.4.2). "Public-Data 1980 Proportional Immigrant Stock" refers to the immigrant stock from a given country living in the U.S. in 1980 as a proportion of that country's 1980 population, estimated from IPUMS-USA (Ruggles et al., 2019b). "Migrants" refers to the estimated immigrant inflows to the U.S. from a given country in year t as a proportion of that country's 1980 population, estimated from IPUMS-USA as well. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

C.9 The Role of Small Countries

This section complements Section C.8 in demonstrating that the primary results in the chapter—contained in Table 3.3—are driven by smaller countries in our sample. These are countries for whom 1980 stocks and migration inflows can only be measured accurately using RDC or DHS data. However, this section also shows that these primary results are not driven by a very small number of these countries.

Table C.8 demonstrates how the result from Column 2 in Table 3.3 changes with population weighting and within sub-samples of countries defined by quartiles of 1980 population. When no lags of the hurricane index are included (Panel A), as in Equation (3.4.2), weighting by 1980 population eliminates the effect. In this specification, India and China receive almost one-half of the total weight in the regression, and the effects in these countries do not appear to be large. However, when we weight by log 1980 population instead, the effect is essentially identical to that found in Table 3.3 (reproduced in the Column 1 here for convenience). Columns 4-7 show that the effect is driven by countries in the bottom quartile by population. We also produce results that add an additional lag in the hurricane index to the model estimated in Equation (3.4.2). These results (in Panel B), largely mirror those in Panel A, but also show a lagged interaction effect response in the second population quartile.

To provide further insight into the role of small countries, we drop progressively larger sets of countries (starting with the smallest in terms of 1980 population) from our analysis. Table C.9 displays these results. The Census Bureau's rounding rules do not permit us to disclose the exact number of countries we drop in each column, but this exercise was done in a systematic way, with the number of dropped countries increasing from left to right in the table (the number of dropped countries has been rounded to the nearest ten in the table, and cannot be specified below 15). The point estimate on the migrant stock interaction term becomes smaller when more and more of the small countries are dropped from the sample. This pattern is consistent with results in Appendix Table A7: the small countries provide key identifying variation. But the result is not contingent on the presence in the sample of only a handful of countries, only disappearing when the smallest 20 countries (approximately) are dropped from the sample (Column 4 of the table).

Finally, Table C.10 sorts the 159 countries in our sample by 1980 population. It demonstrates that the smallest countries in the sample tend to have significant "leverage" in our main regression specifications, with high mean values in the hurricane index. Among the smaller countries, there is variation in the population share of prior migrants, providing additional variation for identifying heterogeneity in the impact of hurricanes.

Table C.8: Alternate Weighting and Unweighted Effect by 1980 Population Quartile

Panel A: No lags in Hurricane Index	Outcome for all columns: Migrants(t) as a Prop. of 1980 Population						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hurricane Index(t)	-0.0010	-0.0002	-0.0010	-0.0012	0.000	-0.0010	-0.0001
	(0.0010)	(0.00020)	(0.00080)	(0.00210)	(0.00060)	(0.00450)	(0.0003)
Hurricane Index(t) $\times$	0.1163**	0.0092	0.1087**	0.1240**	-0.0027	-0.0623	-0.0228
1980 Proportional Immigrant Stock	(0.0451)	(0.0162)	(0.0422)	(0.0520)	(0.0071)	(0.1808)	(0.0536)
Weight	None	1980 Population	Log 1980 Population	None	None	None	None
Quartile	All	All	All	1st	2nd	3rd	4th
Country-Years	3,900	3,900	3,900	1,000	1,000	1,000	1,000
Panel B: One lag in Hurricane Index	Outcome for all columns: Migrants(t) as a Prop. of 1980 Population						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hurricane Index(t)	-0.0011	-0.0003	-0.0011	-0.0012	-0.0001	-0.0013	0.0000
	(0.0010)	(0.0003)	(0.0008)	(0.0022)	(0.0005)	(0.0049)	(0.0004)
Hurricane Index($t-1$)	-0.0009	-0.0006	-0.0009	-0.0005	-0.0016***	0.0007	0.0007
	(0.0008)	(0.0006)	(0.0008)	(0.0019)	(0.0006)	(0.0054)	(0.0006)
Hurricane Index(t) $\times$	0.1175***	0.0117	0.1103***	0.1242**	0.0029	-0.0812	-0.0380
1980 Proportional Immigrant Stock	(0.0450)	(0.0157)	(0.0419)	(0.0520)	(0.0067)	(0.1819)	(0.0652)
Hurricane Index($t-1$) $\times$	0.0150	0.0171	0.0180	-0.0021	0.0871***	-0.1792	-0.1550
1980 Proportional Immigrant Stock	(0.0135)	(0.0548)	(0.0157)	(0.0217)	(0.0080)	(0.1725)	(0.1129)
Weight	None	1980 Population	Log 1980 Population	None	None	None	None
Quartile	All	All	All	1st	2nd	3rd	4th
Country-Years	3,900	3,900	3,900	1,000	1,000	1,000	1,000

Notes: Each column within a panel refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equation (3.4.2). "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table C.9: Dropping Small Countries

Panel A: No lags in Hurricane Index	Outcome for all columns: Migrants(t) as a Prop. of 1980 Population							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hurricane Index(t)	-0.0010	-0.0015*	-0.0007	-0.0003	-0.0002	-0.0001	-0.0005	-0.0005
	(0.0010)	(0.0009)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0004)	(0.0004)
Hurricane Index(t) ×	0.1163**	0.1283***	0.0719***	0.0390	0.0477	0.0592	0.0421*	0.0395
1980 Proportional Immigrant Stock	(0.0451)	(0.0468)	(0.0260)	(0.0319)	(0.0353)	(0.0381)	(0.0254)	(0.0297)
Country-Years	3,900	3,800	3,600	3,500	3,400	3,300	3,200	3,000
Dropped Countries	0	¡15	¡15	20	20	30	30	40
Panel B: One lag in Hurricane Index	Outcome for all columns: Migrants(t) as a Prop. of 1980 Population							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hurricane Index(t)	-0.0011	-0.0016*	-0.0008	-0.0004	-0.0003	-0.0002	-0.0007	-0.0007
	(0.0010)	(0.0009)	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0004)	(0.0005)
Hurricane Index($t-1$)	-0.0009	-0.0009	-0.0008	-0.0008	-0.0010	-0.0008	-0.0011	-0.0012
	(0.0008)	(0.0008)	(0.0009)	(0.0011)	(0.0010)	(0.0009)	(0.0009)	(0.0008)
Hurricane Index(t) ×	0.1175	0.1298***	0.0754***	0.0427	0.0506	0.0617*	0.0442**	0.0417*
1980 Proportional Immigrant Stock	(0.0450)	(0.0467)	(0.0257)	(0.0312)	(0.0340)	(0.0364)	(0.0214)	(0.0242)
Hurricane Index($t-1$) ×	0.0150	0.0170	0.0242	0.0378	0.0217	0.0197	0.0167	0.0176
1980 Proportional Immigrant Stock	(0.0135)	(0.0142)	(0.0210)	(0.0362)	(0.0390)	(0.0390)	(0.0411)	(0.0493)
Country-Years	3,900	3,800	3,600	3,500	3,400	3,300	3,200	3,000
Dropped Countries	0	¡15	¡15	20	20	30	30	40

Notes: Each column within a panel refers to an OLS specification with a constant term, country fixed effects, year fixed effects, and country-specific time trends along with the variables displayed. Standard errors clustered at the country level. See Equation (3.4.2). "Migrants" and "1980 Proportional Immigrant Stock" constructed using restricted-access data from the Census Bureau's Research Data Center. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table C.10: Sample Countries Sorted by Population

Country	1980 Population	Mean Hurricane Index 1980–2004	1980 Proportional Immigrant Stock [a]
Tokelau	1,553	0.0001	0.0515
Falkland Islands (Malvinas)	1,856	0	0.2155
Niue	3,402	0.0214	—[b]
St. Helena	5,899	0	—[b]
Anguilla	6,607	0.0199	0.0605
Turks and Caicos Islands	7,495	0.0114	0.0507
Nauru	7,599	0	0.0105
British Virgin Islands	11,001	0.0160	0.6836
Wallis and Futuna	11,016	0.0003	—[b]
Montserrat	11,845	0.0200	0.0912
Cayman Islands	16,623	0.0264	0.0578
Cook Islands	17,817	0.0003	0.0090
St. Kitts and Nevis	43,388	0.0210	0.0438
Kiribati	56,023	0	0.0025
Bermuda	56,067	0.0170	0.1413
Aruba	59,999	0.0001	0.0403
Seychelles	64,817	0	0.0074
French Guiana	67,801	0	0.0006
Antigua & Barbuda	69,424	0.0299	0.0562
Dominica	74,600	0.0034	0.0386
Micronesia	75,024	0.0003	—[b]
Grenada	89,584	0.0072	0.0804
Tonga	92,407	0.0054	0.0732
Sao Tome & Principe	94,512	0	—[b]
St. Vincent & the Grenadines	99,323	0.0063	0.0358
Vanuatu	116,213	0.0224	—[b]
St. Lucia	120,231	0.0087	0.0126
Western Sahara	137,458	0	0.0017
New Caledonia	140,633	0.0350	0.0010
Belize	144,284	0.0037	0.0992
French Polynesia	151,299	0.0015	0.0054
Maldives	153,593	0	0.0033
Samoa	157,298	0.0109	—[b]
Netherlands Antilles	172,296	0.0003	0.0269
Brunei	189,135	0	0.0041
Bahamas	210,210	0.0199	0.0675
Qatar	226,422	0	0.0046
Solomon Islands	230,691	<0.0001	—[b]
Equatorial Guinea	238,299	0	0.0004

Barbados	250,375	0.0089	0.1141
Macau	251,005	0.0122	0.0112
Cape Verde	299,019	0	0.0336
Martinique	325,459	0.0012	0.0030
Comoros	326,678	0.0006	0.0006
Guadeloupe	328,678	0.0381	0.0025
Djibouti	343,060	0	0.0005
Bahrain	353,735	0	0.0020
Suriname	359,850	0	0.0038
Bhutan	429,390	0	0.0053
Reunion	509,259	0.0179	0.0008
Timor-Leste	568,946	<0.0001	0.0004
Swaziland	607,418	0	0.0012
Gambia	628,415	0	0.0006
Fiji	634,881	0.0241	0.0131
Cyprus	657,838	0	0.0132
Gabon	720,141	0	<.0001
Guyana	768,140	0	0.0667
Guinea-Bissau	803,589	0	0.0002
Botswana	949,005	0	0.0006
Mauritius	964,869	0.0163	0.0007
United Arab Emirates	1,007,555	0	0.0006
Namibia	1,035,391	0	0.0002
Trinidad & Tobago	1,087,911	0.0001	0.0616
Oman	1,169,927	<0.0001	0.0003
Lesotho	1,332,988	0	0.0001
Kuwait	1,370,632	0	0.0033
Mauritania	1,539,525	0	0.0003
Mongolia	1,672,445	0	0.0001
Congo	1,735,761	0	0.0001
Liberia	1,874,816	0	0.0017
Panama	1,974,814	0	0.0306
Jamaica	2,180,542	0.0208	0.0908
Jordan	2,235,174	0	0.0093
Central African Republic	2,311,433	0	<0.0001
Costa Rica	2,323,776	<0.0001	0.0128
Singapore	2,414,214	0.0001	0.0021
Togo	2,673,175	0	0.0002
Lebanon	2,753,241	0	0.0194
Uruguay	2,923,111	0	0.0047
Nicaragua	3,026,750	0.0012	0.0145
Papua New Guinea	3,030,944	<0.0001	0.0002
Libya	3,069,342	0	0.0022

New Zealand	3,147,183	0.0001	0.0039
Paraguay	3,185,226	0	0.0010
Sierra Leone	3,257,631	0	0.0006
Laos	3,272,042	0.0007	0.0157
Honduras	3,519,165	0.0005	0.0106
Benin	3,588,043	0	0.0001
Israel	3,732,547	0	—[b]
Burundi	4,212,187	0	0.0001
Guinea	4,471,424	0	0.0002
Chad	4,517,575	0	<0.0001
El Salvador	4,615,483	<0.0001	0.0205
Hong Kong	5,058,392	0.0172	0.0157
Rwanda	5,140,312	0	<0.0001
Bolivia	5,405,100	0	0.0025
Haiti	5,587,661	0.0036	0.0165
Senegal	5,590,117	0	0.0001
Zambia	5,693,800	0	0.0003
Dominican Republic	5,761,285	0.0068	0.0288
Somalia	5,941,631	<0.0001	0.0001
Niger	5,963,859	0	0.0004
Malawi	6,247,395	0	0.0001
Tunisia	6,375,640	0	0.0005
Burkina Faso	6,570,515	0	<0.0001
Cambodia	6,793,898	0.0001	0.0027
Mali	6,801,635	0	<0.0001
Guatemala	6,825,347	0.0004	0.0093
Zimbabwe	7,229,519	0.0001	0.0005
Angola	7,421,478	0	0.0001
Ecuador	7,914,966	0	0.0112
Ivory Coast	8,429,406	0	0.0001
Yemen	8,519,761	<0.0001	0.0004
Madagascar	8,718,880	0.0053	0.0001
Cameroon	8,844,030	0	0.0002
Syria	8,848,002	0	0.0025
Cuba	9,741,318	0.0071	0.0633
Saudi Arabia	9,932,392	<0.0001	0.0016
Ghana	10,977,531	0	0.0007
Chile	11,095,449	0	0.0033
Mozambique	12,122,316	0.0008	<0.0001
Uganda	12,284,744	0	0.0003
Iraq	13,443,098	0	0.0023
Malaysia	13,646,914	<0.0001	0.0008
Afghanistan	14,112,360	0	0.0003

Nepal	14,498,764	0	0.0001
Sudan	14,600,904	0	0.0002
Australia	14,708,323	0.0001	0.0025
Venezuela	14,932,161	<0.0001	0.0021
Sri Lanka	15,044,572	0.0001	0.0004
Kenya	16,299,302	0	0.0004
Peru	17,311,920	0	0.0033
Taiwan	17,848,320	0.0564	0.0042
Tanzania	18,670,128	0	0.0002
Algeria	19,140,632	0	0.0002
Morocco	19,642,988	0	0.0005
Canada	24,511,056	0.0003	0.0344
Colombia	26,782,940	<0.0001	0.0055
Zaire	27,684,130	0	<0.0001
Argentina	28,244,966	0	0.0024
South Africa	29,164,364	0	0.0006
Burma	33,905,584	0.0014	0.0003
Ethiopia	35,638,836	0	0.0002
South Korea	37,787,544	0.0073	0.0077
Iran	39,299,124	0	0.0031
Egypt	43,783,092	0	0.0010
Turkey	44,476,880	0	0.0012
Thailand	47,197,448	0.0001	0.0012
Philippines	47,843,828	0.0263	0.0107
Vietnam	54,306,296	0.0037	0.0044
Mexico	69,350,248	0.0010	0.0316
Nigeria	74,263,440	0	0.0003
Bangladesh	81,826,248	0.0036	0.0001
Pakistan	82,601,704	0.0002	0.0004
Japan	115,912,104	0.0239	0.0019
Brazil	121,402,072	0	0.0003
Indonesia	147,907,968	<0.0001	0.0002
India	691,926,656	0.0007	0.0003
China	985,918,656	0.0013	0.0003

[a]Estimated using publicly available data from IPUMS-USA (Ruggles et al., 2019b).

[b]Country's 1980 immigrant stock not avaiable in IPUMS-USA.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abadie, A., Diamond, A., and Hainmueller, J. (2010). 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.

Abarcar, P. (2017). The return motivations of legal permanent migrants: Evidence from exchange rate shocks and immigrants in australia. *Journal of Economic Behavior & Organization*, 144:62 – 77.

Abel, G. J. and Sander, N. (2014). Quantifying global international migration flows. *Science*, 343(6178):1520–1522.

Abowd, J. M. and Ashenfelter, O. C. (1981). Anticipated Unemployment, Temporary Layoffs, and Compensating Wage Differentials. In *Studies in Labor Markets*, NBER Chapters, pages 141–170. National Bureau of Economic Research, Inc.

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

Abramitzky, R., Boustan, L. P., and Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *The American Economic Review*, 102(5):1832–1856.

Adao, R., Kolesar, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Akee, R. (2010). Who leaves? deciphering immigrant self-selection from a developing country. *Economic Development and Cultural Change*, 58(2):323–344.

Altonji, J. G., Kahn, L. B., and Speer, J. D. (2016). Cashier or consultant? entry labor market conditions, field of study, and career success. *Journal of Labor Economics*, 34(S1):361–401.

Ambler, K., Aycinena, D., and Yang, D. (2015). Channeling remittances to education: A field experiment among migrants from el salvador. *Amer. Econ. J.: App. Econ.*, 7(2):207–232.

Amuedo-Dorantes, C. and Mundra, K. (2007). Social networks and their impact on the earnings of mexican migrants. *Demography*, 44(4):849–863.

Anastasopoulos, J., Borjas, G. J., Cook, G. G., and Lachanski, M. (2018). Job vacancies and immigration: Evidence from pre- and post-mariel miami. Working Paper 24580, National Bureau of Economic Research.

Angelucci, M. (2015). Migration and financial constraints: Evidence from mexico. *Review of Economics and Statistics*, 97(1):224–228.

Angrist, J. D. and Kugler, A. D. (2003). Protective or counter-productive? labour market institutions and the effect of immigration on eu natives. *The Economic Journal*, 113(488):F302–F331.

Ardington, C., Case, A., and Hosegood, V. (2009). Labor supply responses to large social transfers: Longitudinal evidence from south africa. *American Economic Journal: Applied Economics*, 1(1):22–48.

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review*, 103(5):1553–97.

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

Avakov, A. V. (2015). *Two Thousand Years of Economic Statistics, Years 1–2012*. Algora Publishing.

Azariadis, C. (1975). Implicit contracts and underemployment equilibria. *Journal of Political Economy*, 83(6):1183–1202.

Baez, J., Caruso, G., Mueller, V., and Niu, C. (2017). Droughts augment youth migration in northern latin america and the caribbean. *Climatic Change*, 140(3):423–435.

Baldwin, J. R. (1995). Turnover and productivity growth. In Baldwin, J. R., editor, *The Dynamics of Industrial Competition*, chapter Turnover and Productivity Growth, pages 208–38. Cambridge University Press.

Baldwin, J. R., Jarmin, R. S., and Tang, J. (2002). The Trend to Smaller Producers in Manufacturing in Canada and the U.S. Working Papers 02-06, Center for Economic Studies, U.S. Census Bureau.

Bandiera, O., Rasul, I., and Viarengo, M. (2013). The making of modern america: Migratory flows in the age of mass migration. *Journal of Development Economics*, 102:23–47.

Bartelsman, E. J. and Doms, M. (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature*, 38(3):569–594.

Bauer, T., Dietz, B., Zimmermann, K. F., and Zwintz, E. (2005). German migration: Development, assimilation, and labour market effects. *European Migration: What Do We Know?*, pages 197–261.

Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the us. *The Review of Economic Studies*, 79(1):128–161.

Beaudry, P. and DiNardo, J. (1991). The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data. *Journal of Political Economy*, 99(4):665–688.

Beaudry, P., Green, D. A., and Sand, B. M. (2018). In search of labor demand. *American Economic Review*, 108(9):2714–57.

Beerli, A. and Peri, G. (2015). The labor market effects of opening the border: Evidence from switzerland. Working Paper 21319, National Bureau of Economic Research.

Beine, M., Docquier, F., and Rapoport, H. (2001). Brain drain and economic growth: theory and evidence. *Journal of Development Economics*, 64(1):275–289.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain drain and human capital formation in developing countries: winners and losers. *The Economic Journal*, 118(528):631–652.

Belasen, A. R. and Polachek, S. W. (2009). How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida. *Journal of Human Resources*, 44(1).

Belot, M. V. and Hatton, T. J. (2012). Immigrant selection in the oecd. *The Scandinavian Journal of Economics*, 114(4):1105–1128.

Bernard, A. B. and Jensen, J. B. (1999). Exceptional exporter performance: cause, effect, or both? *Journal of International Economics*, 47(1):1–25.

Bertoli, S., Fernández-Huertas Moraga, J., and Keita, S. (2016). The elasticity of the migrant labour supply: Evidence from temporary filipino migrants. *The Journal of Development Studies*, pages 1–13.

Bertoli, S., Moraga, J. F.-H., and Ortega, F. (2013). Crossing the border: Self-selection, earnings and individual migration decisions. *Journal of Development Economics*, 101:75–91.

Bertoli, S. and Rapoport, H. (2015). Heaven's swing door: Endogenous skills, migration networks, and the effectiveness of quality-selective immigration policies. *The Scandinavian Journal of Economics*, 117(2):565–591.

Blanchard, O. and Giavazzi, F. (2003). Macroeconomic Effects of Regulation and Deregulation in Goods and Labor Markets*. *The Quarterly Journal of Economics*, 118(3):879–907.

Blaum, J., Lelarge, C., and Peters, M. (2018). The gains from input trade with heterogeneous importers. *American Economic Journal: Macroeconomics*, 10(4):77–127.

Blumenstock, J. E., Chi, G., and Tan, X. (2018). Migration and the value of social networks. Working paper, Centre for Economic Policy Research.

Blumenstock, J. E., Eagle, N., and Fafchamps, M. (2016). Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics*, 120:157 – 181.

Bohn, S. and Lofstrom, M. (2012). Employment effects of state legislation against the hiring of unauthorized immigrant workers. Technical report, The Institute for the Study of Labor (IZA).

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

Bohra-Mishra, P., Oppenheimer, M., and Hsiang, S. M. (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences*, 111(27):9780–9785.

Borjas, G. J. (1986). The self-employment experience of immigrants. *Journal of Human Resources*, 21(4):485–506.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *The American Economic Review*, 77(4):pp. 531–553.

Borjas, G. J. (1991). Immigration and self-selection. In *Immigration, Trade, and the Labor Market*, pages 29–76. University of Chicago Press.

Borjas, G. J. (1999). The economic analysis of immigration. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 28, pages 1697–1760. Elsevier.

Borjas, G. J. (2014). *Immigration Economics*. Harvard University Press.

Borjas, G. J. (2017). The wage impact of the marielitos: A reappraisal. *ILR Review*, 70(5):1077–1110.

Borusyak, K., Hull, P., and Jaravel, X. (2018). Quasi-experimental shift-share research designs. Working Paper 24997, National Bureau of Economic Research.

Boustan, L. P., Kahn, M. E., Rhode, P. W., and Yanguas, M. L. (2017). The effect of natural disasters on economic activity in us counties: A century of data. Working Paper 23410, National Bureau of Economic Research.

Brown, C. (1980). Equalizing differences in the labor market. *Quarterly Journal of Economics*, 94(1):113–134.

Brücker, H., Capuano, S., and Marfouk, A. (2013). Education, gender and international migration: insights from a panel-dataset 1980-2010. *Methodology Report*.

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, 82(5):1671–1748.

Burstein, A., Hanson, G., Tian, L., and Vogel, J. (2017). Tradability and the labor-market impact of immigration: Theory and evidence from the u.s. Working Paper 23330, National Bureau of Economic Research.

Bustos, P. (2011). Trade liberalization, exports, and technology upgrading: Evidence on the impact of mercosur on argentinian firms. *American Economic Review*, 101(1):304–40.

Cadena, B. C. and Kovak, B. K. (2016). Immigrants equilibrate local labor markets: Evidence from the great recession. *American Economic Journal: Applied Economics*.

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

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

Card, D. and DiNardo, J. (2000). Do immigrant inflows lead to native outflows? *American Economic Review*, 90(2):360–367.

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

Carrington, W. J., Detragiache, E., and Vishwanath, T. (1996). Migration with endogenous moving costs. *The American Economic Review*, pages 909–930.

Centre for Research on the Epidemiology of Disasters (2016). The human cost of weather related disasters: 1995-2015.

Chen, J. J., Kosec, K., and Mueller, V. (2019). Temporary and permanent migrant selection: Theory and evidence of ability-search cost dynamics. *Review of Development Economics*, 23(4):1477–1519.

Chernina, E., Dower, P. C., and Markevich, A. (2014). Property rights, land liquidity, and internal migration. *Journal of Development Economics*, 110:191 – 215. Land and Property Rights.

Chiquiar, D. and Hanson, G. H. (2005). International migration, self-selection, and the distribution of wages: Evidence from mexico and the united states. *Journal of Political Economy*, 113(2):239–281.

Chiswick, B. and Hatton, T. (2003). International migration and the integration of labor markets. In Bordo, M., Taylor, A., and Williamson, J., editors, *Globalization in Historical Perspective*. University of Chicago Press.

Chiswick, B. R. (1999). Are immigrants favorably self-selected? *The American Economic Review*, 89(2):181–185.

Christopoulou, R. and Vermeulen, P. (2012). Markups in the euro area and the us over the period 1981–2004: a comparison of 50 sectors. *Empirical Economics*, 42(1):53–77.

Clark, X., Hatton, T. J., and Williamson, J. G. (2007). Explaining u.s. immigration, 1971-1998. *The Review of Economics and Statistics*, 89(2):pp. 359–373.

Clemens, M. A. (2017). Violence, development, and migration waves: Evidence from central american child migrant apprehensions. Technical report, Center for Global Development. Working Paper 459.

Clemens, M. A. and Hunt, J. (2017). The labor market effects of refugee waves: Reconciling conflicting results. Working Paper 23433, National Bureau of Economic Research.

Clemens, M. A., Lewis, E. G., and Postel, H. M. (2017). Immigration restrictions as active labor market policy: Evidence from the mexican bracero exclusion. Working Paper 23125, National Bureau of Economic Research.

Clemens, M. A. and Tiongson, E. R. (2017). Split decisions: Household finance when a policy discontinuity allocates overseas work. *The Review of Economics and Statistics*, 99(3):531–543.

Collins, W. J. (1997). When the tide turned: Immigration and the delay of the great black migration. *The Journal of Economic History*.

Collins, W. J. and Wanamaker, M. H. (2015). The great migration in black and white: New evidence on the selection and sorting of southern migrants. *The Journal of Economic History*.

Comola, M. and Mendola, M. (2015). Formation of migrant networks. *The Scandinavian Journal of Economics*.

Cortes, P. (2008). The effect of low-skilled immigration on u.s. prices: Evidence from cpi data. *Journal of Political Economy*, 116(3):381–422.

D'Amuri, F., Ottaviano, G., and Peri, G. (2010). The labor market impact of immigration in western germany in the 1990s. *European Economic Review*, 54(4):550–570.

Davis, S. J. and von Wachter, T. (2011). Recessions and the costs of job loss. *Brookings Papers on Economic Activity*, pages 1–72.

De Weerdt, J. and Hirvonen, K. (2016). Risk sharing and internal migration. *Economic Development and Cultural Change*.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24.

Department of Homeland Security (1996–2004). Yearbook of immigration statistics.

Department of Homeland Security (2006). 2004 yearbook of immigration statistics.

di Giovanni, J., Levchenko, A. A., and Ortega, F. (2014). A global view of cross-border migration. *Journal of the European Economic Association*, 13(1):168–202.

Dilley, M., Chen, R. S., and Deichmann, U. (2005). *Natural Disaster Hotspots: A Global Risk Analysis*. With Jonathan Agwe et al. Disaster Risk Management Series, vol. 5. Washington, D.C.: World Bank.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297–308.

Docquier, F., Peri, G., and Ruyssen, I. (2014). The cross-country determinants of potential and actual migration. *International Migration Review*, 48(s1):S37–S99.

Dolfin, S. and Genicot, G. (2010). What do networks do? the role of networks on migration and coyote use. *Review of Development Economics*, 14(2):343–359.

Doms, M., Lewis, E., and Robb, A. (2010). Local labor force education, new business characteristics, and firm performance. *Journal of Urban Economics*, 67(1):61 – 77. Special Issue: Cities and Entrepreneurship.

Dube, A. and Zipperer, B. (2015). Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies. IZA Discussion Papers 8944, Institute for the Study of Labor (IZA).

Duleep, H., Jaeger, D. A., and Regets, M. (2012). How Immigration May Affect U.S. Native Entrepreneurship: Theoretical Building Blocks and Preliminary Results. IZA Discussion Papers 6677, Institute of Labor Economics (IZA).

Dustmann, C. and Glitz, A. (2015). How Do Industries and Firms Respond to Changes in Local Labor Supply? *Journal of Labor Economics*, 33(3):711–750.

Egger, H. and Kreickemeier, U. (2009). Firm heterogeneity and the labor market effects of trade liberalization. *International Economic Review*, 50(1):187–216.

Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*.

Fairlie, R. W. and Lofstrom, M. (2015). Chapter 17 - immigration and entrepreneurship. In Chiswick, B. R. and Miller, P. W., editors, *Handbook of the Economics of International Migration*, volume 1 of *Handbook of the Economics of International Migration*, pages 877 – 911. North-Holland.

Fajardo, G., Gutierrez, E., and Larreguy, H. (2017). Taking One for the Team: Shocks at Destination and Households' Supply of Migrants. Socioeconomic research working papers 1082, CAF Development Bank Of Latinamerica.

Feenstra, R. C., Markusen, J. R., and Rose, A. K. (2001). Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics*, 34(2):430–447.

Feliciano, C. (2005). Educational selectivity in us immigration: How do immigrants compare to those left behind? *Demography*, 42(1):131–152.

Ferman, B., Pinto, C. C. d. X., and Possebom, V. A. (2016). Cherry picking with synthetic controls. Textos para discussão 420, FGV/EESP - Escola de Economia de São Paulo, Getulio Vargas Foundation (Brazil).

Fort, T. C. and Klimek, S. D. (2018). The effects of industry classification changes on us employment composition. Working Papers 18-28, Center for Economic Studies, U.S. Census Bureau.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Franklin, S. and Labonne, J. (2019). Economic Shocks and Labor Market Flexibility. *Journal of Human Resources*, 54(1):171–199.

Freeman, R. B. (1978). Job satisfaction as an economic variable. *American Economic Review*, 68(2):135–141.

Garcia-Perez, M. (2009). *Firm Owners and Workers: An Analysis of Immigrants and Ethnic Concentration*. PhD thesis, University of Maryland.

Gathmann, C. (2008). Effects of enforcement on illegal markets: Evidence from migrant smuggling along the southwestern border. *Journal of Public Economics*, 92(10):1926–1941.

Gibson, J., McKenzie, D., and Rohorua, H. (2014). Development impacts of seasonal and temporary migration: A review of evidence from the pacific and southeast asia. *Asia & the Pacific Policy Studies*, 1(1):18–32.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik instruments: What, when, why, and how. Working Paper 24408, National Bureau of Economic Research.

Gonzalez, L. and Ortega, F. (2011). How do very open economies adjust to large immigration flows? Evidence from Spanish regions. *Labour Economics*, 18(1):57–70.

Greenwood, M. J. (1985). Human migration: Theory, models, and empirical studies. *Journal of Regional Science*, 25(4):521–544.

Gröger, A. and Zylberberg, Y. (2016). Internal labor migration as a shock coping strategy: Evidence from a typhoon. *American Economic Journal: Applied Economics*, 8(2):123–153.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1):42–57.

Halliday, T. (2006). Migration, risk, and liquidity constraints in el salvador. *Economic Development and Cultural Change*, 54(4):893–925.

Haltiwanger, J. (2015). Job creation, job destruction, and productivity growth: The role of young businesses. *Annual Review of Economics*, 7(1):341–358.

Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). Who creates jobs? small versus large versus young. *The Review of Economics and Statistics*, 95(2):347–361.

Hanson, G., Liu, C., and McIntosh, C. (2017). The rise and fall of u.s. low-skilled immigration. Working Paper 23753, National Bureau of Economic Research.

Hanson, G. H. and McIntosh, C. (2012). Birth rates and border crossings: Latin american migration to the us, canada, spain and the uk. *The Economic Journal*, 122(561):707–726.

Hatton, T. J. and Williamson, J. G. (1993). After the famine: Emigration from ireland, 1850-1913. *The Journal of Economic History*.

Hatton, T. J. and Williamson, J. G. (1994). International migration and world development: a historical perspective. In *Economic aspects of international migration*, pages 3–56. Springer.

Hong, G. and McLaren, J. (2015). Are immigrants a shot in the arm for the local economy? Working Paper 21123, National Bureau of Economic Research.

Hopenhayn, H., Neira, J., and Singhania, R. (2018). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. Working Paper 25382, National Bureau of Economic Research.

Hornbeck, R. (2012). The enduring impact of the american dust bowl: Short- and long-run adjustments to environmental catastrophe. *American Economic Review*, 102(4):1477–1507.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the caribbean and central america. *Proceedings of the National Academy of Sciences*, 107(35).

Hsiang, S. M. and Jina, A. S. (2014). The causal effect of environmental catastrophe on long-run economic growth: Evidence from 6,700 cyclones. Working Paper 20352, National Bureau of Economic Research.

Hunt, J. (2011). Which immigrants are most innovative and entrepreneurial? distinctions by entry visa. *Journal of Labor Economics*, 29(3):417–457.

Hunt, J. and Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2(2):31–56.

Ibarraran, P. and Lubotsky, D. (2007). Mexican immigration and self-selection: New evidence from the 2000 mexican census. In *Mexican Immigration to the United States*, pages 159–192. University of Chicago Press.

Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5):2048–82.

Imbert, C. and Papp, J. (2018). Costs and benefits of seasonal migration: Evidence from india. Working Paper 1, University of Warwick.

Immigration and Naturalization Service (1982–1995). Statistical yearbook of the immigration and naturalization service. USCIS General Collection.

Jack, W. and Suri, T. (2013). Risk sharing and transaction costs: Evidence from kenya's mobile money revolution. *Amer. Econ. Rev.*, 104(1):183–223.

Jacobson, L. S., LaLonde, R. J., and Sullivan, D. G. (1993). Earnings losses of displaced workers. *American Economic Review*, 83(4):685–709.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. Working Paper 24285, National Bureau of Economic Research.

Jarmin, R. S. and Miranda, J. (2002). The Longitudinal Business Database. Working Papers 02-17, Center for Economic Studies, U.S. Census Bureau.

Jarosch, G. (2015). Searching for job security and the consequences of job loss. Technical report, University of Chicago.

Jasso, G. and Rosenzweig, M. (1982). Estimating the emigration rates of legal immigrants using administrative and survey data: The 1971 cohort of immigrants to the united states. *Demography*, 19(3):279–290.

Jasso, G. and Rosenzweig, M. R. (1989). Sponsors, sponsorship rates and the immigration multiplier. *International Migration Review*, pages 856–888.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy*, 114(3):538–575.

Kahn, L. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316.

Kandel, W. A. (2016). U.s. family-based immigration policy. Technical report, Congressional Research Service.

Karahan, F., Pugsley, B., and Sahin, A. (2019). Demographic origins of the startup deficit. Working Paper 25874, National Bureau of Economic Research.

Kaul, A., Klobner, S., Pfeifer, G., and Schieler, M. (2015). Synthetic control methods: Never use all pre-intervention outcomes together with covariates. Mpra paper, University Library of Munich, Germany.

Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*.

Kerr, S. P., Kerr, W., Ozden, C., and Parsons, C. (2016). High-skilled migration and agglomeration. Working Paper 22926, National Bureau of Economic Research.

Kerr, S. P. and Kerr, W. R. (2016). Immigrant entrepreneurship. Working Paper 22385, National Bureau of Economic Research.

Kerr, S. P. and Kerr, W. R. (2018). Immigrant entrepreneurship in america: Evidence from the survey of business owners 2007 & 2012. Working Paper 24494, National Bureau of Economic Research.

Kerr, W. R. (2013). U.s. high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence. Working Paper 19377, National Bureau of Economic Research.

Kleemans, M. and Magruder, J. (2018). Labor market changes in response to immigration: Evidence from internal migration driven by weather shocks. *The Economic Journal*, 128(613):2032–2065.

Kossin, J. P., Hall, T., Knutson, T., Kunkel, K. E., Trapp, R. J., Waliser, D. E., and Wehner, M. F. (2017). Extreme storms. In Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Dokken, D. J., Stewart, B. C., and Maycock, T. K., editors, *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, chapter 9, pages 256–276. U.S. Global Change Research Program, Washington, D.C.

Krolikowski, P. (2017). Job ladders and earnings of displaced workers. *American Economic Journal: Macroeconomics*, 9(2):1–31.

Kuppe, A. M., Lorig, B., Stohr, A., and Schwarz, H. (2013). Training regulations and how they come about. Technical report, Federal Institute for Vocational Education and Training, Bonn.

Leung, D., Meh, C., and Terajima, Y. (2008). Firm Size and Productivity. Staff Working Papers 08-45, Bank of Canada.

Lewis, E. G. (2005). Immigration, skill mix, and the choice of technique. Working Papers 05-8, Federal Reserve Bank of Philadelphia.

Lewis, E. G. (2011). Immigrant-native substitutability: The role of language ability. Working Paper 17609, National Bureau of Economic Research.

Lewis, E. G. (2012). Immigration and production technology. Working Paper 18310, National Bureau of Economic Research.

Llull, J. (2017). The effect of immigration on wages: exploiting exogenous variation at the national level. *Journal of Human Resources*, pages 0315–7032R2.

Lucas, R. E. (1977). Hedonic wage equations and psychic wages in the returns to schooling. *American Economic Review*, 67(4):549–558.

Lucas, R. E. and Stark, O. (1985). Motivations to remit: Evidence from botswana. *Journal of Political Economy*, 93(5):901–918.

Maclean, J. (2013). The health effects of leaving school in a bad economy. *Journal of Health Economics*, 32(5):951–964.

Malchow-Moller, N., Munch, J. R., and Skaksen, J. R. (2012). Do immigrants affect firm-specific wages?*. *The Scandinavian Journal of Economics*, 114(4):1267–1295.

Marchiori, L., Maystadt, J.-F., and Schumacher, I. (2012). The impact of weather anomalies on migration in sub-saharan africa. *Journal of Environmental Economics and Management*.

Massey, D. S. (1988). Economic development and international migration in comparative perspective. *The Population and Development Review*, pages 383–413.

McKenzie, D. and Rapoport, H. (2010). Self-selection patterns in mexico-us migration: the role of migration networks. *The Review of Economics and Statistics*, 92(4):811–821.

McKenzie, D., Stillman, S., and Gibson, J. (2010). How important is selection? experimental vs. non-experimental measures of the income gains from migration. *Journal of the European Economic Association*, 8(4):913–945.

McKenzie, D., Theoharides, C., and Yang, D. (2014). Distortions in the international migrant labor market: evidence from filipino migration and wage responses to destination country economic shocks. *American Economic Journal: Applied Economics*, 6(2):49–75.

McLaughlin, K. J. and Bils, M. (2001). Interindustry mobility and the cyclical upgrading of labor. *Journal of Labor Economics*, 19(1):94–135.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Minale, L. (2018). Agricultural productivity shocks, labour reallocation and rural-urban migration in china. *Journal of Economic Geography*, 18(4):795–821.

Mitaritonna, C., Orefice, G., and Peri, G. (2017). Immigrants and firms' outcomes: Evidence from france. *European Economic Review*, 96:62 – 82.

Monras, J. (2015). Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis. Sciences Po Economics Discussion Papers 2015-04, Sciences Po Departement of Economics.

Morduch, J. (1995). Income smoothing and consumption smoothing. *Journal of Economic Perspectives*, 9(3):103–114.

Morten, M. (2019). Temporary migration and endogenous risk sharing in village india. *Journal of Political Economy*, 127(1):1–46.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the u. s. labor market. *The Quarterly Journal of Economics*, 118(2):pp. 549–599.

Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics*, 13(4):653–677.

Noy, I. (2009). The macroeconomic consequences of disasters. *Journal of Development Economics*, 88(2):221–231.

Okun, A. (1973). Upward mobility in a high-pressure economy. Technical Report 1, The Brookings Institution.

Olney, W. W. (2013). Immigration and firm expansion. *Journal of Regional Science*, 53(1):142–157.

Oreopoulos, P., von Wachter, T., and Heisz, A. (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1).

Orrenius, P. (1999). The role of family networks, coyote prices and the rural economy in migration from western mexico: 1965-1994. Working Papers 9910, Federal Reserve Bank of Dallas.

Orrenius, P. M. and Zavodny, M. (2005). Self-selection among undocumented immigrants from mexico. *Journal of Development Economics*, 78(1):215–240.

Ortega, F. and Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, 1(1):47–74.

Ottaviano, G. I. and Peri, G. (2008). Immigration and national wages: Clarifying the theory and the empirics. Working Paper 14188, National Bureau of Economic Research.

Ottaviano, G. I. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10.

Oyer, P. (2006). Initial labor market conditions and long-term outcomes for economists. *Journal of Economic Perspectives*, 20(3):143–160.

Oyer, P. (2008). The making of an investment banker: Stock market shocks, career choice, and lifetime income. *Journal of Finance*, 63(6):2601–2628.

Passel, J. S. and Suro, R. (2005). Rise, peak, and decline: Trends in u.s. immigration 1992-1994. Technical report, Pew Hispanic Center.

Pedersen, P. J., Pytlikova, M., and Smith, N. (2008). Selection and network effects–migration flows into oecd countries 1990–2000. *European Economic Review*, 52(7):1160–1186.

Pelham, B. and Torres, G. (2008). One-quarter of the world's population may wish to migrate. Technical report, Gallup.

Peri, G. (2012). The effect of immigration on productivity: Evidence from u.s. states. *The Review of Economics and Statistics*, 94(1):348–358.

Peri, G., Shih, K., and Sparber, C. (2015). Stem workers, h-1b visas, and productivity in us cities. *Journal of Labor Economics*, 33(S1):S225–S255.

Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3).

Peri, G. and Yasenov, V. (2015). The labor market effects of a refugee wave: Applying the synthetic control method to the mariel boatlift. Working Paper 21801, National Bureau of Economic Research.

Piguet, E., Pecoud, A., and Guchteneire, P. (2011). *Migration and Climate Change*. Cambridge University Press.

Porter, M. (2003). The economic performance of regions. *Regional Studies*, 37(6-7):549–578.

Redstone, I. and Massey, D. S. (2004). Coming to stay: An analysis of the u.s. census question on immigrants' year of arrival. *Demography*, 41(4):pp. 721–738.

Rinz, K. (2019). Did Timing Matter? Life Cycle Differences in Effects of Exposure to the Great Recession. Working Papers 19-25, Center for Economic Studies, U.S. Census Bureau.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Rosen, S. (1985). Implicit contracts: A survey. *Journal of Economic Literature*, 23(3):1144–1175.

Rosen, S. (1987). The theory of equalizing differences. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, volume 1 of *Handbook of Labor Economics*, chapter 12, pages 641–692. Elsevier.

Rosenzweig, M. and Stark, O. (1989). Consumption smoothing, migration, and marriage: Evidence from rural india. *Journal of Political Economy*, 97(4):905–926.

Rosenzweig, M. and Wolpin, K. (1993). Credit market constraints, consumption smoothing, and the accumulation of durable assets in low-income countries: Investments in bullocks in india. *J. Polit. Econ.*, 101(2):223–244.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146.

Ruffner, J. and Siegenthaler, M. (2016). From labor to cash flow? the abolition of immigration restrictions and the performance of swiss firms. KOF Working papers 16-424, KOF Swiss Economic Institute, ETH Zurich.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2019a). Ipums usa: Version 9.0 [dataset].

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2019b). Ipums usa version 9.0 [dataset]. Minneapolis, MN: IPUMS, 2019.

Schoar, A. (2010). The divide between subsistence and transformational entrepreneurship. In *Innovation Policy and the Economy, Volume 10*, pages 57–81. University of Chicago Press.

Schwandt, H. and von Wachter, T. (2016). Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets. In *Youth Labor Markets*, NBER Chapters. National Bureau of Economic Research, Inc.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, 70(5, Part 2):80–93.

Sorkin, I. (2018). Ranking firms using revealed preference. *The Quarterly Journal of Economics*, 133(3):1331–1393.

Stark, O. (1991). *The migration of labor*. Basil Blackwell.

Stecklov, G., Winters, P., Stampini, M., and Davis, B. (2005). Do conditional cash transfers influence migration? a study using experimental data from the mexican progresa program. *Demography*, 42(4):769–790.

Strobl, E. (2011). The economic growth impact of hurricanes: Evidence from u.s. coastal counties. *Review of Economics and Statistics*, 93(2):575–589.

Taylor, J. E. (1987). Undocumented mexico–us migration and the returns to households in rural mexico. *American Journal of Agricultural Economics*, 69(3):626–638.

Theoharides, C. (2018). Manila to malaysia, quezon to qatar: International migration and the effects on origin-country human capital. *Journal of Human Resources*, 53(4).

United Nations (2017). World population prospects: The 2017 revision, voume i: Comprehensive tables.

U.S. Census Bureau (2). International data base.

van Ark, B. and Monnikhof, E. (1996). Size distribution of output and employment: A data set for manufacturing industries in five oecd countries, 1960s-1990. OECD Economics Department Working Papers 166, OECD Publishing.

Vespa, J. and Armstrong, D. M. (2018). Ps 25-1144: Demographic turning points for the united states: Population projections for 2020 to 2060. Technical report, U.S. Census Bureau.

vom Berge, P., Konig, M., and Seth, S. (2013). Sample of integrated labour market biographies 1975-2010 (siab 7510), fdz data report. Technical report, Research Data Center (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.

von Wachter, T. and Bender, S. (2006). In the right place at the wrong time - the role of firms and luck in young workers' careers. *American Economic Review*, 96(5).

Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., Knutson, T. R., Kossin, J. P., Lee, T.-c., Sobel, A., and Sugi, M. (2016). Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*.

Waugh, M. E. (2017). Firm dynamics and immigration: The case of high-skilled immigration. Working Paper 23387, National Bureau of Economic Research.

Winters, P., De Janvry, A., and Sadoulet, E. (2001). Family and community networks in mexico-us migration. *Journal of Human Resources*, pages 159–184.

World Bank Group (2016). Migration and remittances: Recent developments and outlook. World Bank Group Migration and Development Brief 26.

Wozniak, A. K. (2010). Are college graduates more responsive to distant labor market opportunities? *Journal of Human Resources*, 45(4).

Yang, D. (2006). Why do migrants return to poor countries? evidence from philippine migrants' responses to exchange rate shocks. *The Review of Economics and Statistics*, 88(4):715–735.

Yang, D. (2008a). Coping with disaster: The impact of hurricanes on international financial flows, 1970-2002. *B.E. Journal of Economic Analysis and Policy (Advances)*, 8(1).

Yang, D. (2008b). International migration, remittances and household investment: Evidence from philippine migrants' exchange rate shocks. *The Economic Journal*, 118(528):pp. 591–630.

Yang, D. (2008c). Risk, migration, and rural financial markets: Evidence from earthquakes in el salvador. *Social Research*, pages 955–992.

Yang, D. and Choi, H. (2007). Are remittances insurance? evidence from rainfall shocks in the philippines. *World Bank Econ. Rev.*, 21(2):219–248.

Yang, D. and Martinez, C. (2006). Remittances and poverty in migrants' home areas: Evidence from the philippines. *International Migration, Remittances and the Brain Drain Washington DC: World Bank*, pages 81–121.

Zavodny, M. (1997). Welfare and the locational choices of new immigrants. *Economic Review-Federal Reserve Bank of Dallas*, page 2.