NBER WORKING PAPER SERIES

JOB VACANCIES AND IMMIGRATION:
EVIDENCE FROM PRE- AND POST-MARIEL MIAMI

Jason Anastasopoulos
George J. Borjas
Gavin G. Cook
Michael Lachanski

Working Paper 24580
http://www.nber.org/papers/w24580

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2018

We are grateful to Ron Bird, Hugh Cassidy, Kirk Doran, Bill English, Richard Freeman, Kenneth Goldstein, Daniel Hamermesh, Gordon Hanson, Garett Jones, Daniel Leach, Joan Llull, Joan Monras, Marian Moszoro, John Nye, Solomon Polachek, Valerie Ramey, Dani Rodrik, Jeanne Shu, Jan Stuhler, Steve Trejo, and Jay Zagorsky for very helpful discussions and comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Job Vacancies and Immigration: Evidence from Pre- and Post-Mariel Miami
Jason Anastasopoulos, George J. Borjas, Gavin G. Cook, and Michael Lachanski
NBER Working Paper No. 24580
May 2018
JEL No. J6,J61,J63

ABSTRACT

How does immigration affect labor market opportunities in a receiving country? This paper contributes to the voluminous literature by reporting findings from a new (but very old) data set. Beginning in 1951, the Conference Board constructed a monthly job vacancy index by counting the number of help-wanted ads published in local newspapers in 51 metropolitan areas. We use the Help-Wanted Index (HWI) to document how immigration changes the number of job vacancies in the affected labor markets. Our analysis begins by revisiting the Mariel episode. The data reveal a marked decrease in Miami's HWI relative to many alternative control groups in the first 4 or 5 years after Mariel, followed by recovery afterwards. We find a similar initial decline in the number of job vacancies after two other supply shocks that hit Miami over the past few decades: the initial wave of Cuban refugees in the early 1960s, as well as the 1995 refugees who were initially detoured to Guantanamo Bay. We also look beyond Miami and estimate the generic spatial correlations that dominate the literature, correlating changes in the HWI with immigration across metropolitan areas. These correlations consistently indicate that more immigration is associated with fewer job vacancies. The trends in the HWI seem to most strongly reflect changing labor market conditions for low-skill workers (in terms of both wages and employment), and a companion textual analysis of help-wanted ads in Miami before and after the Mariel supply shock suggests a slight decline in the relative number of low-skill job vacancies.

Jason Anastasopoulos
Center for Information Technology Policy
309 Sherrerd Hall
Princeton University
ljanastas@princeton.edu

George J. Borjas
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138
and NBER
gborjas@harvard.edu

Gavin G. Cook
Department of Sociology
Princeton University
Princeton, NJ 08544
gcook@princeton.edu

Michael Lachanski
Woodrow Wilson School
Princeton University
Princeton, NJ 08544
mlachans@princeton.edu

Job Vacancies and Immigration: Evidence from Pre- and Post-Mariel Miami

Jason Anastasopoulos, George J. Borjas, Gavin G. Cook, and Michael Lachanski*

I. Introduction

How does immigration affect labor market opportunities in a receiving country? This is perhaps the central question in the economics of immigration, in terms of both its economic content and its political implications. Most of the fundamental problems in labor economics relate to how labor markets adjust to supply and demand shocks. An immigration-induced increase in labor supply creates opportunities to observe how firms and workers react and adjust to the changed environment. Put simply, we can exploit the supply shocks to understand what makes wages go up and down. Similarly, the debate over immigration policy is concerned with how immigration changes the size of the economic pie available to the receiving country, and, particularly, with how that pie is split. Who wins and who loses from immigration? The identification of winners and losers is likely to provide much insight into the political battle over immigration policy.

Not surprisingly, the centrality of the question inspired a voluminous amount of empirical research over almost four decades (Blau and Mackie, 2016). In the U.S. context, this literature has almost entirely used microdata, such as the decennial censuses or the Current Population Surveys (CPS), to document how wages change in those markets targeted by immigrants. Sometimes the markets are defined by geographic boundaries; sometimes the markets are defined by skill group (as in Borjas, 2003). But the basic strategy is the same. Immigrants tend to target some markets more than others. We then measure the impact of immigration by contrasting the evolution of wages in the markets hit by immigration with the evolution in the markets that immigrants shunned.

* Anastasopoulos: Woodrow Wilson School, Princeton University and School of Public and International Affairs, University of Georgia; Borjas: Harvard Kennedy School; Cook: Department of Sociology, Princeton University; Lachanski: Woodrow Wilson School, Princeton University, and U.S. Department of Commerce. We are grateful to Ron Bird, Hugh Cassidy, Kirk Doran, Bill English, Richard Freeman, Kenneth Goldstein, Daniel Hamermesh, Gordon Hanson, Garett Jones, Daniel Leach, Joan Llull, Joan Monras, Marian Moszoro, John Nye, Solomon Polachek, Valerie Ramey, Dani Rodrik, Jeanne Shu, Jan Stuhler, Steve Trejo, and Jay Zagorsky for very helpful discussions and comments.

Due to a host of technical issues (e.g., immigrants do not just happen to target some markets randomly; firms and workers diffuse the impact of local supply shocks by moving elsewhere; the available data often yields small samples of immigrants and natives for many cities and skill groups), the existing literature demonstrates that it is quite difficult to isolate the impact of immigration on wages. Even more problematic, it turns out that the evidence often depends on *researcher choices* about how to frame the empirical analysis.

The recent debate over the wage impact of the Mariel boatlift provides a classic example of how those choices can change the answer. Card's (1990) original work showed that the wage of the average worker in Miami was barely affected by the 8 percent increase in supply that Mariel represented. But Borjas (2017) shows that if one focused on the low-skill worker most likely to be affected by Mariel (as represented by the average prime-age, non-Hispanic man who is a high school dropout), the evidence suggested that the wage fell by at least 15 percent. In contrast, Peri and Yasenov (2015) illustrate that using an alternative definition for a low-skill worker (i.e., the average non-Cuban man or woman who is 16 to 61 years old and has not yet graduated from high school) implies that the wage was roughly constant in pre- and post-Mariel Miami.[1] By the time this debate runs its course, the presumed impact of Mariel will depend on the *reader's* choice of which of the studies best represents the true wage trend in Miami, and there will be a menu of choices for that reader to pick from.[2] This menu will allow some to argue that Mariel had a substantial impact on wages while simultaneously allowing others to argue that Mariel had no impact.

We contribute to the literature by reporting findings from a "new" data set, a data set that economists have used often since the 1960s, but, surprisingly, has never been exploited in the immigration context. Beginning in 1951, the Conference Board constructed an index of job vacancies in local labor markets by counting the number of help-wanted classified ads in newspapers in 51 metropolitan areas. Although the rise of online advertising obviously reduced the usefulness of this index beginning sometime around 2000, leading the Conference Board to

---

[1] See also Borjas (2016, 2017) and Clemens and Hunt (2017).

[2] In recent years, a large literature has examined the problem of researcher degrees of freedom (Simmons, Nelson, and Simonson, 2011). This literature shows how researcher decisions can bias inference (Gelman and Lokan, 2013), and numerous creative statistical procedures have been proposed to address the problem (Steegan et al., 2016). There is much less work addressing the impact of reader degrees of freedom in program evaluation, policy analysis, and journalism (although see Lim, 2017, for an example of the lattermost analysis).

discontinue collecting data in 2010, the Conference Board's Help-Wanted Index (HWI) now contains a historical series of the ebbs and flows of labor demand in local labor markets for the six decades between 1951 and 2010.[3]

The HWI has been used to study such diverse and important phenomena as the trend of wages and productivity in the stagnant 1970s (Medoff, 1982); the relevance of the sectoral shifts explanation of structural unemployment (Abraham and Katz, 1986); the relation between job vacancies and the unemployment rate (Cohen and Solow, 1967; Burch and Fabricant, 1968; and Abraham, 1987); the role of job search in a real business cycle framework (Andolfatto, 1996), and the cyclicality of job vacancies (Shimer, 2005). An important theme runs through these studies: The HWI provides valuable information about labor demand and is highly correlated with various measures of labor market conditions.

We use the HWI to document how immigration affects the number of job vacancies in local labor markets.[4] Our analysis begins by revisiting the Mariel episode. The data clearly reveal a marked decrease in Miami's HWI relative to many alternative control groups in the first 4 or 5 years after Mariel, followed by a full recovery afterwards. We then extend the analysis to two other supply shocks that also hit Miami over the past few decades: the initial wave of Cuban refugees in the early 1960s (prior to the abrupt ending of that flow with the Cuban Missile Crisis in 1962), as well as the 1994-1995 "Mariel Boatlift That Never Happened" (Angrist and Krueger, 1999). Both of these episodes reveal a similar pattern—an initial decline in Miami's HWI shortly after the supply shock.[5] Finally, we look beyond Miami and estimate the generic spatial correlations that dominate the literature, correlating changes in the HWI with immigration across the 51 metropolitan areas for which the index is available. These spatial correlations also

[3] The Conference Board started the Help-Wanted Online (HWOL) index in 2005, which counts online job ads by category.

[4] As far as we know, our paper is the first to investigate the quantitative relationship between immigration and job vacancies in the U.S. labor market. Withers and Pope (1985) find no statistically significant relationship between immigration and the Australian job-finding rate (see also Warren, 1982), while Pholphirul (2012) reports that immigration reduces short-term job vacancies in Thai manufacturing. Job vacancies are often present as a latent variable in both theoretical (Ortega, 2000) and empirical analyses (Davila and Saenz, 1990) of immigration's impact on labor market outcomes. Chassamboulli and Peri (2015) use the national HWI to calculate a long-run job-finding rate, which they then fix to calibrate a labor market matching model for evaluating various immigration policies.

[5] The HWI in Miami did not recover after the 1995 supply shock, perhaps because that shock was not followed by a hiatus in Cuban immigration.

indicate that less immigration is typically associated with increased employer effort to find workers by placing help-wanted ads in the local newspaper.[6]

The evidence raises a number of interesting issues. First, the persistent negative relation between immigration and job vacancies through five decades is peculiar given that the HWI presumably gives an aggregate measure of local labor market conditions, while immigration has been disproportionately low-skill. We show, however, that partly because of how classified ads were counted by the newspapers included in the Conference Board survey, the HWI is probably a better barometer for measuring labor market conditions at the bottom end of the skill distribution. The data, in fact, indicate that trends in the HWI are most strongly correlated with local labor market conditions for the least-educated workers: Both the wage and the employment rate of high school dropouts increase faster as the HWI rises. We also briefly look inside the black box of the HWI by examining the actual text of a small sample of help-wanted ads that appeared in Miami before and after Mariel. The textual analysis shows that the drop in the number of ads was somewhat larger for low-skill job vacancies.

Second, there is an incongruity between the unequivocal evidence for a negative immigration-job vacancies link and the confusion that permeates the literature that correlates immigration and wages. As noted earlier, some of the confusion can be traced to measurement issues related to how the change in the wage structure is best measured. Part of the incongruity, however, might arise because the trend in the number of job vacancies reflects the change in employer demand for the *marginal* worker. And these marginal responses are likely to occur quickly. In contrast, the wage changes that are the focus of the immigration literature are measuring the impact of supply shocks on the labor market conditions facing the *average* worker. The marginal-average distinction implies that wage changes will be harder to detect than the change in the number of job vacancies (and much more so if wages are sticky).[7] Our analysis suggests that there may be much to learn by looking for and studying data that provide

[6] Goodwin and Carlson (1981) show that wage controls from 1971 to 1974 were associated with an increase in classified job advertising. They present a static model in which wage ceilings induce firms to advertise more. Our work addresses the dual proposition that increased immigration reduces equilibrium classified job advertising.

[7] To complicate matters further, the average wage change is often measured over the decade between the decennial censuses. Local labor markets are hit by many other shocks in that span of time, some of which may be correlated with immigration, further clouding the identification of the wage impact of immigration.

alternative indicators of labor market conditions. The behavior of these additional metrics might help inform how the more standard manipulation of microdata from the Census or CPS should be conducted and interpreted.

Finally, the analysis of the HWI has one very valuable feature. It greatly reduces the number of degrees of freedom available to a researcher interested in estimating the labor market impact of immigration. Put bluntly, it removes the possibility of researcher intervention in deciding how to measure labor market conditions. The index was created concurrently with the supply shocks by independent organizations for a purpose totally unrelated to the immigration question that is at the core of this paper.[8] The historical trends in the HWI in the cities that received many or few immigrants are set in stone. Employers looked at conditions in the local labor market and decided to spend money (or not) on help-wanted ads. All we can do at this point is simply document and describe the cumulative pattern of employer decisions.

II. The Conference Board Help-Wanted Index

Unlike many OECD countries, the United States did not maintain any continuous official statistics on job openings until the year 2000, when the Bureau of Labor Statistics (BLS) introduced the Job Openings and Labor Turnover Survey (JOLTS). Prior to 2000, researchers interested in understanding the trends and determinants of job openings or vacancies relied primarily on the Conference Board's Help-Wanted Index (HWI).

Beginning in 1951, the Conference Board contacted 51 newspapers (listed in Appendix Table A1), each corresponding to a metropolitan area, and enumerated the classified ads placed each month in each paper. This number was adjusted for seasonality and day-of-the-week bias to create a monthly index for each metropolitan area. These metropolitan area indices were then aggregated to create an index for each geographic region, and for the nation as a whole.[9]

It is well known that the HWI correlates well with labor market conditions (Preston, 1977). As an example, Figure 1 illustrates the relation between the national HWI and the

---

[8] The HWI is also based on very large samples, involving the enumeration of thousands of ads monthly over six decades.

[9] Apart from the removal of the *Newark Evening News* (and the Newark metropolitan area) in 1971, and a swap of the *Dallas Times Herald* News for the *Dallas Morning News* in the early 1990s, the newspapers and cities surveyed did not change after 1970. Zagorsky (1998) combined previous surveys of help-wanted classifieds by the Metropolitan Life Insurance Company with the HWI to create a help wanted index that dates back to 1923.

unemployment rate, showing a consistent inverse relation between the two variables throughout much of the period.[10] However, Autor (2001, p. 27) noted that the HWI was "flat throughout the 1990s economic boom" and cited the migration of "vacancy listings… from newspapers to the Internet" as a possible explanation. In a similar vein, Kroft and Pope (2014) find that the growth of local online adds in Craigslist caused a reduction in the city's HWI. In response to the declining relevance of newspaper help-wanted sections, the Conference Board ceased the public release of the HWI in July 2008 and stopped internal data collection in October 2010. To avoid the reliability problems resulting from the growth of online advertising, we do not use the post-1999 HWI data in much of our analysis.

Despite the strong correlation between the unemployment rate and the HWI, there are several biases in the index that can influence the interpretation of observed trends. The first arises from the fact that the number of job vacancies *per ad* is procyclical. During booms, a single ad might advertise explicitly for two or more job openings. But the algorithm used by the Conference Board to construct the HWI counts this as only one advertised job (Preston, 1977). Figure 2*a* illustrates the bias through an ad published in the *Miami Herald* on March 2, 1975. This single posting advertised for "several openings" for test technicians.

A related enumeration problem arises with ads placed by private employment agencies. These ads often contained several job postings (as in Figure 2*b*). Some newspapers placed all private employment agency advertising in a section specifically demarcated for labor market intermediaries (and this section may not have been included in the Conference Board counts), while other newspapers made no distinction between private employment agency advertising and ads placed directly by individuals or firms (Walsh, Johnson, and Sugarman, 1975). Further, when reporting private employment agency ads to the Conference Board, some newspapers counted a posting by a private employment agency as a single help-wanted ad, but other newspapers did not (Abraham, 1987).[11]

---

[10] Friedman (1985, p. 63) examines trends in the HWI for the Phoenix labor market and concludes that the HWI is "a viable indicator of future employment activity in the Phoenix area, just as it is nationally."

[11] Cohen and Solow (1967, p. 108) noted these methodological issues, writing that "we know nothing, for example, about the number of jobs offered per advertisement." They also make the point, which we return to later in the paper, that "the index can not be decomposed by occupation," making it difficult to determine if the index is a better metric of labor market conditions for some skill groups than for others. An additional problem arises when (typically large) firms advertised the same position in newspapers in multiple markets, inflating the HWI relative to the actual number of unfilled vacancies (Zagorsky, 1998).

A final bias arises because the HWI only counts classified ads placed in the official help-wanted section of the newspaper.[12] As Figure 2*c* shows, again drawn from the pages of the *Miami Herald* on March 2, 1975, many high-skill jobs, especially those in finance, insurance and real estate (FIRE), were not advertised in the help-wanted section at all. They instead appeared in dedicated FIRE sections, juxtaposed with ads extolling "excellent land opportunities" or a "3,895 acre operating ranch," and those sections were not included in the computation of the index.[13]

These methodological issues imply that intercity differences in the *level* of the HWI do not provide a good metric for making comparisons in local labor market conditions. The Conference Board count of classified ads depended partly on newspaper advertising policies—how many job openings were posted in a specific ad and how that ad was counted; where the ad was physically placed in the newspaper.

Equally important, there are differences in the market power of the newspapers used by the Conference Board. In some cities, as in Miami (where the sampled paper was the *Miami Herald*), the paper used to construct the index was the key source of job classifieds in the area. In other locations, as in Minneapolis (where the sampled newspaper was the *Minneapolis Star Tribune*), there were other newspapers (the *St. Paul Dispatch-Pioneer*) that also contained many job classifieds (Courtney, 1991). The Conference Board, however, did not count help-wanted ads in the "secondary" newspapers. In sum, intercity differences in the level of the HWI may not be informative.[14]

We address this issue by rescaling the HWI so that the level of the index in each city equals 1 at some point in the pre-treatment period. In our context, this rescaling is inconsequential, except when illustrating the trends graphically. Our estimate of the impact of the supply shock on job vacancies uses a difference-in-differences estimator, so that the level of the index washes out in the calculation.[15]

---

[12] It is unclear if the newspapers that segregated private employment agency ads from other help-wanted ads were also those that did not count private agency ads in their reports to the Conference Board. Thus, we tentatively conceptualize the private agency ad bias as qualitatively different from the FIRE bias.

[13] Because the various biases work in both directions, quantitative attempts to "de-bias" the index conclude that very small correction factors are needed at the national level (Zagorsky, 1998).

[14] As an example of the cross-section variation in the HWI, the 5th percentile value of the index in 1980 is 40.2 and the 95th percentile value was more than four times larger, or 166.9.

[15] Medoff (1983) makes a similar point, arguing that although the absolute level of the index in any given region depends on the sample of newspapers surveyed by the Conference Board, cross-region

The HWI data has been used by many widely-cited studies, but it has never been applied to the immigration context. Abraham (1987) provides the most comprehensive discussion of the benefits and problems with the HWI and shows that the index (or some normalization of it) is correlated with the true number of job vacancies. For example, she compares the HWI with administrative data from Minnesota (one of only two states that collected information on the number of job openings at the time) and concludes: "the Minnesota data… suggest that the normalized help-wanted index is a reasonably good vacancy proxy" (Abraham, 1987, p. 213).

Although the explosion of online job postings reduced the relevance of newspaper-based indices, the HWI remains useful for historical research. For the time period that our research covers, the HWI is the gold standard for labor market data on job vacancies. We obtained the entire HWI time series for the 51 metropolitan areas directly from the Conference Board. Our analysis exploits the historical HWI data in a novel context. How does the tightness or slackness of a local labor market respond to immigration-induced supply shocks?

III. Job Vacancies and Supply Shocks in Miami

Figure 3 shows the number of Cuban immigrants migrating to the United States each year between 1955 and 2010. Most of these immigrants chose to settle in the Miami metropolitan area. The shaded areas in the figure help visualize that the Miami labor market area was hit by three unexpected, short-lived, and sizable supply shocks of Cuban immigrants during this period. The first was in the early 1960s, beginning soon after the Castro takeover and ending abruptly in 1962 with the Cuban missile crisis. The second is the very noticeable spike in 1980 associated with the Mariel boatlift. And the third occurred in 1994-1995 when a new boatlift of Cuban refugees was temporarily detoured to Guantanamo Bay by the Clinton administration. This detour did not last long, however, and most of the refugees ended up reaching the United States in 1995. The figure also shows that a sizable number of Cubans entered between 1965 and 1970, but this "shock" was less sudden and did not end abruptly within a year or two. We will discuss this particular wave of Cuban immigrants in more detail below.

---

differences in the percent change in the index can capture differences in the rate of growth of job vacancies. Courtney (1991) uses the share of advertising expenditures going to a particular newspaper in a metropolitan area to partially adjust for the cross-section variation.

Table 1 reports key characteristics about the size and skill composition of the three short-lived supply shocks. Each of the shocks increased the size of the workforce in Miami by a sizable amount. The initial wave of Cuban refugees increased the labor force by almost 17 percent; the *Marielitos* increased supply by 8.4 percent; and the mid-1990s shock increased supply by 3.9 percent.[16]

This section documents the response of Miami's HWI to these three distinct shocks. The short-run response in Miami, relative to comparable cities, was quite similar across the three shocks—a quick decline in the index of job vacancies within two or three years. In two of the three supply shocks, the drop was followed by a recovery.

A. Mariel

We begin by revisiting the supply shock that has had the most influence in both the academic literature and in the policy debate: the influx of Cuban refugees during the Mariel boatlift. The historical details of the Mariel episode are well known. On April 20, 1980, Fidel Castro declared that Cuban nationals wishing to move to the United States could leave freely from the port of Mariel. About 125,000 Cubans left before an agreement between the Carter administration and the Castro regime closed the escape valve in October 1980.

The top panel of Table 1 shows that the Mariel supply shock, which increased the size of the workforce in Miami by 8.4 percent, was composed of relatively low-skill workers: Nearly 60 percent of the refugees lacked a high school diploma and only 7.4 percent had a college degree. As a result, the supply shock disproportionately increased the size of the low-skill workforce. The number of workers in Miami without a high school diploma increased by 18.4 percent, but the number of college graduates rose by only 3.4 percent. The "effective" increase in the number of low-skill workers was probably larger than these statistics indicate, as Miami's employers "downgraded" the skills that the *Marielitos* brought. The typical refugee with a high skill diploma ended up working in an occupation where the average worker did not have a high school diploma.

---

[16] The calculation assumes that all Cuban refugees initially settled in Miami. The 1970 census indicates that 49.2 percent of the Cuban immigrants from the 1960-1964 cohort lived in Miami; the 1990 census indicates that 63.5 percent of the 1980-1981 cohort lived in Miami; and the 2000 Census indicates that 63.7 percent of the 1994-1995 cohort lived in Miami.

The labor market impact of the Mariel boatlift was first studied in Card's (1990) classic paper. Card's analysis of the Miami labor market, when compared to conditions in other labor markets that served as a control group, indicated that nothing much happened to Miami despite the large number of refugees. The average wage did not fall and the unemployment rate remained unchanged relative to what was happening in the cities that formed the control group.

The debate over the labor market impact of the *Marielitos* intensified in the past few years as a result of the Borjas (2017) reappraisal. The Borjas study, which used data from both the March CPS and the CPS Outgoing Rotation Group (ORG), specifically examined the wage trends experienced by the subset of workers in Miami who lacked a high school diploma. It turns out that looking at the wage trends of this very specific group reverses the stylized fact that little happened to the labor market in post-Mariel Miami. A flurry of subsequent papers argues that the Borjas (2017) evidence is sensitive to the definition of the "low-skill" workforce, to changes in the racial composition of the sample, and to the inevitable sampling error resulting from the small sample of workers that the CPS surveyed in the Miami metropolitan area.[17] Our analysis of the behavior of the HWI in Miami around the time of Mariel is impervious to these potential problems.

As we noted earlier, part of the cross-city variation in the level of the HWI arises because of the idiosyncratic nature in which classified ads were counted by different newspapers. We address this issue by rescaling the HWI so that the level of the index in each city equals 1 in the 1977-1979 period. Figure 4 begins the empirical analysis by illustrating the 1975-1989 trend in the HWI in Miami, in the South Atlantic region, and in the entire nation. We choose the 1975-1989 period because Cuban immigration to the United States was relatively low and stable in the pre-treatment period (hovering around 6 thousand annually between 1975 and 1978), and it was again relatively low and stable in the post-Mariel period between 1981 and 1989 (hovering around 10 thousand).

The trends in the raw HWI data are visually striking. The index for Miami declined very rapidly after 1980, reaching a nadir towards the end of 1982, and then began a slow recovery through the 1980s. By 1989, the value of the HWI index for Miami was again similar to the national index (although it was still lower than the index for the South Atlantic region).

[17] See Borjas (2016, 2017b), Borjas and Monras (2017), Peri and Yasenov (2015), and Clemens and Hunt (2017).

The relative trend in Miami's HWI index, in fact, resembles the wage trend for high school dropouts reported in Borjas (2017, Figure 2), with the dropout wage declining between 1980 and 1985 and then recovering by 1990. The similarity between the trends in Miami's low-skill wage and the HWI is conceptually interesting. It suggests that the HWI, which presumably measures "average" labor market conditions, does a particularly good job at capturing changing job opportunities for low-skill workers. We will return to this point in a subsequent section.

Of course, Miami's distinctive trend should be contrasted with what happened in "comparable" cities rather than with the regional or national aggregates in Figure 4.[18] To document the uniqueness of Miami's post-Mariel experience, we use a number of alternative controls. We start with the set of metropolitan areas selected by Card (1990) as a control group for the Miami of the early 1980s. This control group is composed of Atlanta, Los Angeles, Houston, and Tampa-St. Petersburg. Unfortunately, the HWI was never calculated for the Tampa-St. Petersburg metropolitan area, so our "Card control" only has the other three cities. The top panel of Figure 5 shows the trend in the HWI index for Miami and for the cities in the Card control group.[19] It is evident that Miami's HWI declined dramatically soon after Mariel relative to what was observed in the control cities.

We next use the control group of the four metropolitan areas that had similar growth rates for low-skill employment in the pre-Mariel period (as reported in Borjas, 2017). These four cities are Gary, Houston, Indianapolis, and Los Angeles. Figure 5 also shows that the post-1980 Miami experience was unusual when compared to this "low-skill control." The relative decline in the HWI index between 1980 and 1983 is about 30 percent (or, more precisely, -0.3 log points) if we use either the low-skill control group or the Card control group.

Finally, we use the synthetic control method (Abadie, Diamond, and Hainmueller, 2010). The method essentially searches across all 50 potential control cities and derives a weight that combines cities to create a new synthetic city. This synthetic city is the one that best resembled the pre-Mariel Miami labor market along some set of pre-specified conditions.

---

[18] We use the three-digit *metarea* variable (which defines a metropolitan statistical area) in the IPUMS files of the decennial censuses to gather information about the cities for which the HWI is available. The only exception occurs in the context of Gary, Indiana (which is officially defined as a "metropolitan division"), where we use the four-digit IPUMS code.

[19] The average index for the control group is a weighted average of the index across the cities, where the weight is the city's employment in 1980.

We calculate the synthetic control by using a large number of control variables, all calculated from the 1970 and 1980 IPUMS decennial census files. The control variables are: the education distribution of workers in the city in 1980 and the percent change in the number of workers in each education group between 1970 and 1980; the industrial distribution of workers in the city in 1980 and the percent growth in the number of workers employed in each of the industries between 1970 and 1980; the fraction of immigrants in the workforce in 1980 and the percent growth in the number of immigrants between 1970 and 1980; the fraction of workers who are male, the percent growth in the number of male workers between 1970 and 1980, and the percent growth in the number of female workers; the fraction of workers who are black (in 1980); and the fraction who are Hispanic. The cities that make up the synthetic control are: Charlotte (with a weight of 0.014), Jacksonville (0.023), Los Angeles (0.338), Memphis (0.141), New Orleans (0.053), and San Antonio (0.432). The top panel of Figure 5 also illustrates the trend in the HWI for the synthetic control, and it again shows that the Miami experience was unusual. The HWI in Miami declined by about 40 percent by 1983 relative to the synthetic control.

To calculate the impact of the Mariel supply shock on the HWI, we estimate a generic difference-in-differences regression model where the unit of analysis is a city-year-month cell:

$$\log H_{rtm} = \theta_r + \theta_t + \theta_m + \beta(\text{Miami} \times \text{Post-Mariel}) + \varepsilon, \tag{1}$$

where H_{rtm} is the HWI in city r, year t, and month m; θ_r is a vector of city fixed effects; θ_t is a vector of year fixed effects; θ_m is a vector of month fixed effects; “Miami” is a dummy variable identifying the Miami-Hialeah metropolitan area; and “Post-Mariel” indicates if cell (r, t, m) was observed after June 1980. The regression uses monthly data from January 1975 through December 1989.

The cities r included in the regression are Miami and the cities in a specific control group. For example, if the Miami experience is being compared to that of cities in the low-skill control group, there would be five cities in the data, and each of these cities would be observed 180 times between 1975 and 1989, for a total of 900 observations. The regression comparing Miami to the synthetic control is similar in spirit, but there are only two “cities” in this regression: Miami and the synthetic city, for a total of 360 observations. Because the impact of

Mariel might vary over time, the post-Mariel variable in equation (1) is a vector of fixed effects indicating whether the observation refers to the intervals June 1980-1982, 1983-1984, 1985-1986, or 1987-1989.

The top panel of Table 2 reports the estimated coefficients in the vector β (and robust standard errors) for various specifications of the regression model.[20] If we use the Card control group, the relative HWI in Miami declined by about 20 percent in 1981-1982 and by over 40 percent in 1983-1984. The second column replicates the regression using the low-skill control group and also shows sizable declines by the mid-1980s. The third column compares Miami to the synthetic control and shows a decline of about 25 percent by 1981-1982 and of over 40 percent by 1985.[21] Note that all regressions indicate a full recovery in Miami's HWI by the late 1980s (in fact, Miami's HWI is higher by the end of the decade). Finally, the last column of the table bypasses any attempt at selecting a control group, and simply compares Miami to the other 50 cities in the Conference Board survey. The regression then consists of 9180 observations (51 cities, each observed 180 times). The results again are unequivocal. The HWI for Miami declined steeply relative to all other cities through 1986, before starting a full recovery.[22]

In their well-cited work on the structural shifts hypothesis, Abraham and Katz (1986) normalize the national HWI index by dividing it by the size of the workforce. We define the "normalized" help-wanted index as:

$$(2) \qquad v_{rtm} = \frac{H_{rtm}}{E_{rtm}}$$

[20] We report robust standard errors because clustered standard errors are downward biased when there are few clusters (Cameron and Miller, 2015).

[21] The regression estimated using the synthetic control differs in a slight (and numerically trivial) way from the other regressions. As noted above, we rescaled the HWI index so that the index for each city equals 1 in the 1976-1978 period. This rescaling is irrelevant in the regressions that use the Card control group, the low-skill control group, or the entire set of 50 metropolitan areas, as the regressions include a vector of city fixed effects. The synthetic control regression aggregates all the control cities into a synthetic city and does not include individual fixed effects for each of those cities. The coefficients in the vector β are almost identical if we skip the step of rescaling the index to equal 1 in the pre-Mariel period. In particular, the coefficient for the 1980-1983 period is -0.273 (0.029); the coefficient for the 1984-1986 period is -0.511 (0.021), and the coefficient for the 1987-1989 period is 0.057 (0.030).

[22] Schwartz, Cohen, and Grimes (1986) note that the behavior of aggregate labor turnover data, such as national job vacancies, mask considerable regional heterogeneity. They conclude that national job vacancy statistics should not be used to analyze the sources of unemployment changes over the business cycle. Our findings shed light on one potential source for these observed heterogeneities by raising the question of whether recent labor supply shocks can increase a region's sensitivity to the business cycle.

where E_{rtm} gives non-agricultural employment in city r in year t and month m. The BLS has collected information on the size of the non-agricultural workforce for city-year-month cells since the early 1950s.[23]

The bottom panel of Figure 5 replicates the figure showing the trend in the normalized HWI for Miami and the alternative control groups. It is visually obvious that the trend in the normalized index strongly resembles the trend in the raw HWI. The bottom panel of Table 2 reports the regressions using the normalized index as the dependent variable. The regressions again document the very different experience of post-Mariel Miami relative to the control cities. If we use the synthetic control baseline, the normalized index declined by about 40 percent by the mid-1980s, and fully recovered by the late 1980s.[24]

B. The First Wave of Cuban Refugees

After several years of guerrilla warfare, Castro toppled the U.S.-supported government of President Fulgencio Batista on January 1, 1959. As Figure 3 showed, there was little migration of Cubans to the United States during the Batista years. Fewer than 8,000 Cuban nationals migrated to the United States in 1958. The number of immigrants remained low in 1959, during the first year of the revolutionary government, but began to increase rapidly soon after that, as the communist and totalitarian nature of the Castro regime came to the surface. In 1960, nearly 40,000 Cubans migrated, and over 50,000 Cubans migrated in both 1961 and 1962.[25]

---

[23] The BLS data at the city-year-month level for this period is available in a file maintained by the ICPSR, "Employment, Employment, Hours, and Earnings in States and Areas of the United States, 1940-1991" (ICPSR 9928).

[24] The obvious interpretation of the evidence is that a large immigrant supply shock leads to labor market slackness. There are two alternative (and related) interpretations that are of interest. The first is that immigration reduced employer dependence on help-wanted advertising via the establishment or deepening of existing Cuban social networks. Pholphirul (2012) provides suggestive evidence that the firms that used immigrants to fill vacancies in the past are much more likely than other firms to fill vacancies using immigrants in the future. The second is that, as a result of Miami's growing Cuban community, Miami's employers began to shift their advertising to Spanish-language outlets. In fact, the *Miami Herald* had begun to publish a sister newspaper in Spanish, *El Nuevo Herald*, in 1977. The Conference Board did not enumerate help-wanted ads in *El Nuevo Herald*, mechanically leading to a decline in the HWI in post-Mariel Miami. Both of these arguments fail to explain why Miami's HWI recovered after a few years unless the immigrant social networks break down quickly or employers gave up on advertising in Spanish language newspapers and returned to the *Miami Herald*. Regardless, the available data does not allow us to distinguish between these explanations.

[25] Hughes (1999) gives a detailed history of Cuban migration to the United States.

That first wave of Cuban refugees ended just as suddenly as it began. Most of those refugees arrived in the United States on a twice-daily commercial flight that Pan American Airways operated between Havana and Miami. On October 16, 1962, President Kennedy was informed that the Soviet Union had placed medium-range ballistic missiles in Cuba, setting in motion the Cuban Missile Crisis. Within a week, the Pan American flights were discontinued, and the Cuban exodus was abruptly halted, cutting the number of Cuban immigrants entering the United States by nearly two-thirds in 1963 and 1964.

The exodus of Cuban refugees resumed in December 1965, after the establishment of the “Freedom Flights” that helped reunite many families separated by the migration hiatus that followed the Cuban Missile Crisis. Many Cubans took advantage of the Freedom Flights. An average of 44,000 Cubans migrated to the United States each year between 1966 and 1969. The flow dwindled down to a small number (about 15,000 a year) by the early 1970s.

We initially focus on the labor market consequences of the 1960-1962 supply shock, which bears some resemblance to Mariel. First, it was unexpected. Second, although it did not happen in a matter of weeks (as was the case with Mariel), it was short-lived, lasting less than three years. Third, it ended just as suddenly as it began. The period 1954-1965, therefore, sets up another natural experiment that allows us to determine if the trends revealed by our analysis of the HWI data during the Mariel period also showed up in early 1960s Miami.

Panel B of Table 1 reports some basic facts about the size and skill composition of the first wave of Cuban immigrants. We use the 1960 census to calculate the size and the education distribution of Miami’s workforce prior to the influx of the refugees. There were almost 400,000 workers in Miami at the time, and almost half did not have a high school diploma. We then used the 1970 Census (the first large-scale microdata that identifies the Cuban immigrants who migrated in the early 1960s) to enumerate the Cuban workers who migrated to the United States between 1960 and 1964 as adults (so that they are at least 26 years old in 1970).[26] The influx of Cuban refugees increased the size of Miami’s workforce by 16.6 percent—almost twice as large as the 8.4 percent increase resulting from the Mariel boatlift. Notably, the early 1960s supply shock was remarkably large at the upper end of the education distribution. The first wave of

[26] The 1970 census does not identify the specific year of migration within the 1960-1964 interval, but the number arriving in 1963 and 1964 is relatively small. The data reported in Table 1 makes the small adjustment necessary to isolate the 1960-1962 immigrants using detailed year-of-migration information from the 2000 census.

Cuban refugees increased the number of college graduates in Miami's workforce by over 40 percent.

Note, however, that the typical Cuban immigrant in that early wave experienced a substantial amount of "down-skilling," with employers heavily discounting their credentials. We used the 1960 census to calculate the average years of education of native workers in each occupation, and then used this statistic to compute the years of education in the occupation employing the average native and the average Cuban immigrant. The bottom two rows of the middle panel of Table 1 show the extent to which the Miami labor market downgraded (at least initially) the educational attainment of the immigrants. The average Cuban who did not have a college diploma ended up working at a job where the average native was a high school dropout. And even the average Cuban college graduate ended up at a job that employed natives who had just one year more than a high school education.

The down-skilling suggests that the supply shock might have substantially increased the "effective" number of low-skill workers in Miami. For instance, if Cubans without a college diploma competed for jobs typically filled by natives who did not have a high school diploma, the effective supply of high school dropouts increased by 26.1 percent. In short, although many of the early Cuban refugees had educational credentials, the Miami labor market discounted those credentials and viewed them as low-skill workers. At least initially, therefore, the 1960-1962 supply shock bears some resemblance to the Mariel supply shock two decades later. They both increased the size of the low-skill workforce substantially.

Figure 6 illustrates the raw trends in the HWI for the 1954-1965 period, again comparing Miami with both the South Atlantic region and the nation as a whole.[27] It is visually obvious that the three indices had similar trends in the Batista years. Within a year after the Castro takeover in 1959, however, the HWI index for Miami began a relative decline and then began to recover (relative to the national index) after the flow of refugees stopped abruptly in October 1962. By 1965, the national index and the Miami index were again in relative equality (although the Miami index lagged behind the index for the South Atlantic region).

Rather than rely on creating arbitrary control groups (as the Card control group or the low-skill control group introduced earlier), we employ the synthetic control method to examine the trends in the 1954-1965 period (assuming the treatment occurred on January 1960). We use

[27] The HWI index for each city/region is normalized so that it equals 1 in 1956-1958.

exactly the same specification employed in the Mariel analysis reported earlier, except that all the control variables are now calculated using the 1950 and 1960 IPUMS census files. The cities that compose the synthetic control are: Charlotte (0.31), Jacksonville (0.085), Memphis (0.064), San Bernardino (0.373), Sacramento (0.023), and San Diego (0.144).[28]

The top panel of Figure 7 illustrates the relative trends in the HWI implied by the synthetic control method, and it again shows that the post-1960 Miami experience is unusual. The index for Miami began to decline (relative to the synthetic control) soon after the influx of Cuban refugees began in 1960 and had begun to recover by the mid-1960s.

The first column of Table 3 estimates the generic regression model in equation (1) using data from the 1954-1965 period. The coefficients indicate that Miami's HWI declined relative to the synthetic control between 1960 and 1963 and seems to have been recovering by the middle of the 1960s. The relative index dropped by 27 percent by 1963 and lagged by "only" 22 percent by 1965.

We also estimated the regression model using the normalized HWI defined in equation (2). However, data on the number of non-agricultural workforce at the city-year-month level is not available for all of the metropolitan areas in the 1954-1965 period. The BLS provides a complete time series in this period for only 14 of the cities in the HWI survey. Nevertheless, the bottom panel of Figure 6 shows very similar results when we examine trends in job vacancies in this small subset of metropolitan areas. It is visually obvious that the arrival of the first wave of Cuban refugees in the early 1960s is associated with a noticeable decline in the number of job vacancies in Miami. It is also evident that Miami had started to recover by 1965. Table 3 reports the estimates of the regression model and the coefficients suggest a drop of about 28 percent in the normalized HWI by 1963, with some recovery by 1965 (when Miami's index lagged behind the synthetic control by 13 percent).

As noted earlier, the immigration hiatus after the Cuban Missile Crisis did not last long. The Freedom Flights began in December 1965, initiating another large flow of Cuban refugees that lasted until the early 1970s. It is of interest to extend the period of analysis beyond 1965 to determine how the HWI in Miami reacted to the restart of the Cuban immigrant influx and to its eventual slowdown, which then led to a hiatus in the 1970s that lasted through Mariel.

---

[28] The HWI data for some of the metropolitan areas did not begin to be collected until after 1954. We use the subsample of the 44 metropolitan areas that have a complete time series for the period.

We use the synthetic control analysis reported above to extend the post-treatment period through 1978 (shortly before Mariel). The two panels of Figure 8 illustrate the long-term trends in job vacancies. It is evident that the recovery that seemed to have been taking place in 1965 did not, in fact, take hold. As soon as the influx of Cuban immigrants restarted in 1966, the HWI for Miami again began to diverge from that of the synthetic control. This divergence lasted until the end of the Freedom Flights, but the recovery in the early 1970s was short-lived. The deep recession of 1973-1975 led to a severe decline in Miami's HWI relative to the synthetic control, and Miami did not recover after that.[29]

Column 2 of Table 3 reports the associated regression coefficients. The coefficients suggest that the mid-1960s recovery was stalled by the initiation of the Freedom Flights. Miami's HWI was 22 percent below that of the synthetic control in 1965 and remained at roughly that level through 1971. The situation improved after the Freedom Flights ended. By 1972-1984, the HWI in Miami was only 9.8 percent below that of the synthetic control. Note, however, that the 1973-1975 recession and its aftermath had a particularly adverse impact on job vacancies in Miami.

The trends illustrated in Figure 8 raise a very interesting, but never analyzed, question about the relation between the business cycle and the ability of local labor markets to absorb and recover from supply shocks. If the Miami experience were generalizable, it hints at the possibility that cyclical downturns have a particularly severe effect on labor markets recently hit by supply shocks.[30]

It is easy to show, however, that the diverging trends between Miami and the synthetic control cannot be solely attributable to differential impacts of the business cycle. The regressions reported in column 3 of Table 3 add an interaction between the Miami dummy variable and the national unemployment rate in the year-month cell. The relevant coefficients still show a decline in Miami's HWI soon after the initial wave of Cuban refugees. They also show that the recovery

[29] The large discrepancy in the HWI between Miami and the synthetic control in the late 1970s raises an interesting conceptual issue. It seems that relative to the cities that Miami resembled in the mid-1950s, the Miami of the late 1970s is far worse off. The Mariel analysis reported earlier uses a synthetic control composed of a *different* set of cities that resembled Miami in the late 1970s. It is evident that the cumulative impact of both supply and demand shocks change the starting point for the difference-in-differences framework in ways that deserve much greater attention.

[30] We discuss the role of cyclical fluctuations in more detail in the next section. In particular, we isolate the long-term trend in the HWI after removing transitory cyclical fluctuations that might have affected local labor markets during the period.

was stalled by the Freedom Flights and that it was not until after the Freedom Flights ended (by 1972-1974) that the Miami disadvantage eased considerably.

In sum, labor market conditions in Miami in the 1960s responded in a fashion that is strikingly similar to what was observed during the Mariel years, at least as measured by the HWI. The supply shock was followed by a softening of labor market conditions, with some recovery occurring after the migration flow stopped.

C. The "Mariel Boatlift That Did Not Happen"

As Figure 3 showed, there was another spike in Cuban immigration in 1995. The number of immigrants rose from about 18,000 in 1993 to over 50,000 in 1995, before quickly falling to below 20,000 in 1996. This spike coincides with the period examined by Angrist and Krueger (1999, p. 1328):

> In the summer of 1994, tens of thousands of Cubans boarded boats destined for Miami in an attempt to emigrate to the United States in a second Mariel Boatlift that promised to be almost as large as the first one…Wishing to avoid the political fallout that accompanied the earlier boatlift, the Clinton Administration interceded and ordered the Navy to divert the would-be immigrants to a base in Guantanamo Bay. *Only a small fraction of the Cuban émigrés ever reached the shores of Miami*. Hence, we call this event, "The Mariel Boatlift That Did Not Happen." [emphasis added]

At the time that Angrist and Krueger wrote about this episode, they had no way of knowing what the 2000 Census would eventually uncover. That the Mariel boatlift that did not happen indeed happened; it was just delayed by a year. The refugees diverted to Guantanamo made it to the United States after President Clinton reversed course in May 1995 and permitted their entry.

We examine the labor market consequences of this particular supply shock by focusing on the period between 1990 and 1999. It is important to note, however, that this supply shock differs in two crucial ways from the first wave of Cuban immigration and from the Mariel episode. First, although it involved a sizable number of immigrants (with over 75,000 Cubans migrating in 1994-1995), Miami was a much larger city by the mid-1990s. As a result, the relative increase in labor supply was much smaller—only about 3.9 percent. Similarly, although the mid-1990s Cuban influx was disproportionately low-skill, the number of high school dropouts in Miami's workforce rose by only 5.5 percent.

Second, and perhaps more important, the aftermath of the 1994-1995 supply shock was not followed by a hiatus or even a short period of stability in the number of Cubans migrating to the United States. Before 1994, the immigration flow from Cuba hovered around 15,000 persons per year. After the 1995 spike, Cuban immigration began a steady rise that continued through 2010. There were 16.7 thousand Cuban immigrants in 1997, and this number almost doubled by 1999 when 29.7 thousand Cubans entered the country. The steady increase in Cuban immigration after the "Mariel Boatlift That Did Not Happen" implies that it is far from an ideal natural experiment, perhaps making it more difficult to detect the recovery suggested by our analysis of the two other supply shocks that hit Miami.

Figure 9 shows the raw trend in Miami's HWI during the 1990s, contrasting it with the trend in both the South Atlantic region and in the national index.[31] A very notable feature of the figure is the huge upward spike in the Miami index in the last half of 1992. The value of the HWI for Miami rose from 47 to 91 in the four months between August and November, 1992, an increase of 93.6 percent.[32] This spike coincides exactly with the aftermath of Hurricane Andrew. Andrew was a Category 5 hurricane that made landfall in Homestead, Florida, just south of the city of Miami, on August 24, 1992. At the time, it was the strongest hurricane to ever make landfall in the United States, causing $45 billion in damage (in 2018 dollars).

The behavior of Miami's HWI in the aftermath of Andrew dramatically shows how a local labor market tightens substantially after a major environmental disaster that requires a lot of rebuilding (Belasen and Polachek, 2009). After the quick spike, the HWI began to decline slowly until late 1993. It then remained relatively stable through 1995, at which point the index began a steady decline at the same time that the national and regional economies were booming. Note that the raw data do not suggest any type of recovery in the Miami index in the post-1995 period.

We replicated the empirical analysis by applying the synthetic control method to the 1994-1995 supply shock. To avoid the obvious contamination created by the short-term effect of

---

[31] We normalize the index for each region by setting the index at a value of 1 for the period 1991-August 1992.

[32] A doubling of a city's index over a four-month period (or a halving of the index) is extremely rare. We calculated all possible four-month changes in the index in all 51 metropolitan areas over the entire 1951-1999 period. Out of the 29,408 four-month changes, the one observed in Miami in late 1992 is exceptionally large, at the 99.99 percentile. To put the 93.6 percent increase in the index in perspective, the spread from the 1st percentile to the 99th percentile of all potential four-month changes goes from a decline of 25.2 percent to an increase of 29.0 percent.

Andrew, we time the treatment as of August 1992, so that the synthetic control method will create a synthetic city that resembles Miami prior to both the hurricane and the supply shock. The set of controls is exactly the same as that used earlier in our analysis of the other shocks, with the only change being that we used the 1980 and 1990 IPUMS census files to construct the control variables. The synthetic control is composed of: Los Angeles (0.448), New Orleans (0.116), New York (0.178), Providence (0.022), and San Bernardino (0.237).

The top panel of Figure 10 shows the trend in Miami's HWI relative to the index in the synthetic control. The comparison reveals the same insights as the raw data. There was a remarkable spike due to Hurricane Andrew with the index becoming relatively stable throughout the 1994 calendar year, but then the index began to decline with the arrival of the refugees from Guantanamo Bay. Notably, Miami's HWI continued to decline throughout the last half of the 1990s.

The top panel of Table 4 reports the coefficients from the difference-in-differences regression model. We expand the specification of the model to allow for the separate identification of the impact of Hurricane Andrew. In particular, we use the interregnum between Andrew and the supply shock (i.e., the period between January 1994 and June 1995) as the baseline. As the first column of the table shows, Andrew increased the HWI by about 37 percent, and the index fell (relative to the baseline) by 15 to 20 percent between 1995 and 1999. Notably, there is no evidence that the Miami index recovered after the supply shock. As we suggested earlier, the continued decline in Miami's index after the 1995 supply shock can perhaps be attributed to the steady increase in Cuban immigration after 1996.

There is one other event in the post-1995 period that likely influenced the calculation and interpretation of the HWI: The appearance and explosive growth of online job postings in the Internet. It is possible that the relative decline of the post-1995 Miami index resulted from a faster rate of adoption in online help-wanted ads in Miami, with a corresponding decline in printed ads.

The CPS conducted two supplemental surveys in December 1998 and August 2000 that report whether a household owns a personal computer and whether anyone in the household uses the Internet. The pooled data from these two surveys indicate that Miami lagged behind the average city in the HWI sample in both computer ownership and Internet use. In particular, 54 percent of adults across all metropolitan areas lived in a household that owned a personal computer in the late 1990s, and 38 percent lived in a household where someone used the Internet. In contrast, 42 percent of adults in Miami lived in a household that owned a computer, and 33 percent lived in a household where someone used the Internet.

We can assess the sensitivity of the regression coefficients to the intercity differences in computer use (and presumably increased use of online job postings) by estimating the difference-in-differences regression model using the entire sample of the 51 metropolitan areas. The second column of Table 4 reports the basic regression. The coefficients again reveal the strong positive impact of Andrew on the HWI, as well as the decline in the index in the late 1990s. The last two columns of the table measure the sensitivity of these trends to differential computer use across cities by interacting the year fixed effects (beginning with July 1994) with a variable measuring either the percent of the households in the city that own a computer or the percent of the households where someone uses the internet. The inclusion of these interactions barely affects the estimated impact of the supply shock on Miami's HWI.

We can also use data from the BLS on non-agricultural employment at the city-year-month level to calculate the normalized HWI defined in equation (2). The bottom panel of Figure 10 shows the trends in the normalized index in Miami and in the synthetic control. It is visually obvious that our evidence is similar regardless of which index we examine. Similarly, the middle panel of Table 4 reports the regression coefficients of interest. We find that Hurricane Andrew increased the normalized HWI by nearly 40 percent (relative to the baseline), and that the mid-1990s supply shock reduced it by 15 to 25 percent.

IV. Testing the Robustness of the Evidence

This section presents a number of sensitivity tests to help assess if our empirical finding of a negative association between immigration and job vacancies is robust.

A. Randomizing the Control Group

The inference that supply shocks reduce job vacancies may hinge crucially on the construction of the counterfactual. One way of determining how much the choice of a control group matters is to bypass entirely the decision of picking control variables to create a synthetic city that is comparable to Miami, and instead estimate the wage impact for *every* potential control group. We illustrate this approach by estimating the short-run effect of the three supply shocks relative to each of the 230,300 possible four-city control groups that can be assembled from the HWI data.

The top panel of Figure 11 illustrates the approach by estimating the short-run impact of the Mariel supply shock on Miami's HWI relative to all potential four-city control groups. In particular, we use the data from the 1979-1984 period, set the treatment date as of June 1980, and estimate the generic regression difference-in-differences model in equation (1). The relevant coefficient is the interaction of an indicator variable identifying the Miami metropolitan area and whether the observation refers to a time period between June 1980 and 1984. The density function in the figure shows the distribution of all the 230,300 estimated short-run effects.

The mean of the coefficients was -0.232, with a standard deviation of 0.106. Moreover, only 1.7 percent of the coefficients were zero or positive. The figure also shows the location of the comparable impact estimated from a synthetic control regression. It seems the synthetic control is "picking" a short-run impact that is large relative to the mean.

The middle panel of the Figure replicates the analysis for the 1960-1962 supply shock. The regressions use the period 1958-1963 and set the treatment date as of January 1960. The density again shows that the mean impact is negative and numerically important (at -0.183), and that none of the estimated effects in the 230,300 regressions led to a coefficient that was greater than or equal to zero. Finally, the bottom panel shows the density of impacts for the 1994-1995 supply shock, based on regressions using the period 1994 through 1997, and setting the treatment date as of June 1995. The density again reveals that the mean impact is negative and sizable. And, again, none of the estimated short-run effects is positive.

B. The Job-Finding Rate and the Beveridge Curve

Since 1976, the BLS reports the monthly unemployment rate for the metropolitan areas examined in our analysis.[33] These data allow us to examine how two of the supply shocks that hit Miami might affect what is known as the "job-finding rate" in the macro literature.

Assume that there is a matching function $M(U, V)$ giving the number of newly hired workers, where U is the number of workers unemployed and V is the number of job vacancies (Pissarides, 1985; Shimer, 2005). If the matching function has constant returns to scale, the job-

[33] The pre-1990 unemployment rate data are available in the monthly *Employment and Earnings* series published by the Bureau of Labor Statistics. The post-1990 data are available online at the BLS website.

finding rate (i.e., the probability *M/U* that an unemployed worker finds a job) can be written as a function of the ratio *V*/*U*. Empirically, we can then proxy the job-finding rate by the ratio of the HWI to the unemployment rate at the city-year-month level.

The two panels of Figure 12 illustrate the behavior of this proxy in Miami and the synthetic control for the two post-1980 supply shocks. The top panel shows a rapid decline in Miami's job-finding rate soon after Mariel, with a full recovery by the late 1980s.[34] The bottom panel shows that the job-finding rate experienced a huge increase immediately after Hurricane Andrew, and then fell to its previous level within a year. After the 1995 supply shock, however, the job-finding rate in Miami began to lag behind that of the synthetic control. The economic boom of the late 1990s increased the job-finding rate in the synthetic control, while Miami's job-finding rate remained roughly constant.[35] In sum, the joint trends in job vacancies and unemployment suggest that unemployed workers find it more difficult to find a job after a sizable supply shock.

The discussion, in fact, suggests that there may be a shift in the Beveridge curve after a supply shock. The downward-sloping Beveridge curve describes the relationship between job vacancies and unemployment (Elsby, Michaels, and Ratner, 2015). An inward shift in the curve is typically interpreted as a more efficient labor market—there is less unemployment for a given level of job vacancies. At the same time, however, note that an inward shift also means that there are fewer job vacancies for a given level of unemployment, making it harder for the unemployed to find a job. Our analysis of the job-finding rate suggests that the Beveridge curve might well have shifted as a result of the changed labor market opportunities resulting from immigration.

We illustrate the potential shift by focusing on the Mariel episode.[36] The top panel of Figure 13 illustrates the data scatter that forms the national Beveridge curve (using the national

[34] The 1976-1989 unemployment series is not complete for 29 of the 51 cities in the HWI survey. In some cases, the missing data affects only a few random months during the 14-year period. In other cases, however, the problem is severe; the unemployment rate for Detroit, for example, was not reported for 40 months. Our analysis of the job-finding rate uses data for 34 metropolitan areas. We excluded all cities where the unemployment rate is missing for more than 3 months and linearly interpolated the unemployment rate in those cities where there are three or fewer missing months.

[35] For the sake of brevity, we do not report the companion regressions. The measured impact of the two supply shocks on the job-finding rate in Miami is statistically significant.

[36] The unemployment rate at the metropolitan area level is not available in the 1960s and the vacancy-unemployment relation in Miami for the early 1990s is likely contaminated by the impact of

HWI and the national unemployment rate) before and after 1980. The national Beveridge curve seems to have shifted out after 1980, probably as a result of the deep recession in the early 1980s. The middle panel shows the Beveridge curve for the synthetic control, calculated by taking a weighted average of the unemployment and vacancy rates across metropolitan areas (using the weights implied by the synthetic control analysis of the job-finding rate). As with the national data, the Beveridge curve for the synthetic control also shifted out. Finally, the bottom panel shows the comparable data for Miami, which generates exactly the opposite pattern, an inward shift in the Beveridge curve. As noted above, this shift suggests that the supply shock may have made Miami's labor market more efficient (i.e., less unemployment for a given number of vacancies), but also that it became harder for the unemployed to find a job in post-Mariel Miami.[37]

We can estimate the magnitude of the shift in Miami's Beveridge curve by pooling the time series of the HWI and unemployment data across metropolitan areas. Specifically, we pool the data for the 34 metropolitan areas for which we have information on the local unemployment rate for the 1976-1984 period. The estimated relative shift in Miami's Beveridge curve is:

$$\log H_{rtm} = \underset{(0.002)}{-0.062}\, u_{rtm} - \underset{(0.030)}{0.276}\, (\text{Miami} \times \text{Post-Mariel}) + \theta_r + \theta_t + \theta_m, \tag{3}$$

where u_{rtm} gives the unemployment rate in cell (r, t, m); and the "post-Mariel" period goes from June 1980 through December 1984. The Mariel supply shock shifted down Miami's Beveridge curve by about 28 percent relative to what was happening in other cities at the time.[38]

C. Cyclical Fluctuations and the HWI

Hurricane Andrew. The Mariel episode thus provides the cleanest way of establishing whether (and how) a supply shock shifted the Beveridge curve.

[37] Warren (1982) discusses the link between immigration and the Beveridge curve in the Australian context. He finds little evidence that changes in Australian immigration policy shifted the Beveridge curve.

[38] We also estimated the regression model comparing Miami with the synthetic control. The interaction coefficient was -0.301 (0.028). If we extend the sample period through 1989, we find that the Beveridge curve for Miami shifted back up again by the end of the 1980s; the interaction coefficient for 1988-1989 was -0.071 with a standard error of 0.045.

To make the analysis transparent, the evidence presented in this paper has relied on the index constructed by the Conference Board, with minimal manipulation of the raw data. It is obvious that the HWI has a cyclical trend, and that this cyclical trend might vary across metropolitan areas (perhaps *because* of immigration-induced supply shocks). It is of interest to illustrate how our results are affected if we remove the short-term cyclical fluctuations from the index.

One simple way of doing so is by stacking the HWI data across all time periods and all 51 metropolitan areas and then running a regression of the index on a vector of interacted year-month fixed effects. The residual from this regression gives the value of the HWI for a city-year-month cell after removing any part of the index that was common across all cities in that particular time period (including the cyclical movement). Put differently, the within-city trend in this residual rids the data of the impact of "transitory" cyclical fluctuations that affect all geographic units equally, leaving behind the "permanent" trend in the help-wanted index for a particular metropolitan area.

The left column of Figure 14 shows the behavior of the adjusted index for each of the three supply shocks we have analyzed.[39] Consider initially the top panel of the figure, which shows the impact of Mariel. The statistical exercise obviously removes quite a bit of the cyclical variation in the HWI for the synthetic control. It is also obvious, however, that the exercise reveals yet again how different the post-1980 Miami experience was from what was observed in the cities that Miami resembled prior to Mariel. By the mid-1980s, Miami's adjusted index had fallen by about 40 percent relative to the index in the synthetic control.

The bottom panels of the figure show the same pattern for the two other supply shocks. We find that the adjusted index fell soon after the initial wave of Cuban refugees; there is a hint of a recovery in the mid-1960s after that initial flow ended, but that recovery was halted by the Freedom Flights; and it wasn't until 1974 that Miami's index was again in parity with that of the synthetic control. Similarly, we find the adjusted index in Miami jumping sharply as a result of

---

[39] The synthetic control results in Figure 14 are calculated using a two-step approach. We first remove the impact of the short-run cyclical fluctuation from the HWI time series. We then use the adjusted HWI data to find the synthetic control and generate the trends summarized in the figure. The synthetic control method, therefore, compares Miami with cities that had similar permanent trends prior to the supply shock.

Hurricane Andrew, and then declining steadily after the “Mariel Boatlift That Never Happened” happened.

An alternative way of removing the cyclical fluctuations, popular in macroeconomics, is to use the Hodrick-Prescott (HP) filter.[40] In a panel data context, this filter decomposes the HWI time series for each metropolitan area into a long-run trend and a short-term cyclical component. Conceptually, the long-run trend is somewhat comparable to the residual from the fixed effects regression discussed above. But there is one important difference between the long-run trends calculated by the two approaches. Because the HP filter is applied individually to the time series in each metropolitan area, it ignores the possibility that the intensity of the cyclical fluctuations in a particular city (e.g., Miami) might themselves have been affected by supply shocks. Netting out cyclical fluctuations, therefore, might remove some of the effect of supply shocks on the locality. A finding that the long-run trend implied by the HP filter is shifted by supply shocks would provide strong evidence for the hypothesis that such shocks have a sizable and more persistent impact on local labor market conditions.

The right column of Figure 14 illustrates the movements in this long-run trend when we run the HWI data through the HP filter. Each of the figures confirms the basic story that the raw data are telling. The long-run trend in Miami’s HWI was dislodged (relative to what happened in the synthetic control) after each of the supply shocks. The drop is particularly obvious in the context of Mariel and in the aftermath of the first wave of Cuban immigrants. And although the HP filter rids Miami’s index from the transitory effects of Hurricane Andrew, the data still show the growing divergence between Miami and comparable cities after the 1994-1995 supply shock and the steady increase in Cuban immigration subsequently.

The fact that the impact of Hurricane Andrew is visible in the right column of Figure 14 but not in the left provides an insight into how the two approaches differ in the way they remove transitory fluctuations. The fixed effects method removes any transitory fluctuation that was common to all cities; Hurricane Andrew obviously was not. The HP filter removes every transitory fluctuation in every city, regardless of whether that fluctuation is unique to one city or common to all. The HP filter reveals that Andrew’s impact, although specific to Miami, was

---

[40] The HP filter requires that the researcher specify the value of a smoothing parameter. The value typically recommended for quarterly data is 1600 and the value recommended for monthly data (which we use) is 129,600, which equals $1600p^4$, where p is the number of periods in a quarter (Rahn and Uhlig, 2002). Hamilton (2018) gives a detailed discussion of the problems with the HP filter.

transitory and did not displace Miami's long-run trend in the job vacancy index. In contrast, the supply shocks, although also specific to Miami, had a sufficiently large and longer-lasting impact, shifting the long-run trend.

V. Skills and the Help-Wanted Index

The evidence showing that the HWI index in Miami responded in similar ways to the three distinct (and sudden) supply shocks that hit Miami between 1960 and 2000 raises an interesting question: *Whose* labor market conditions does the HWI actually reflect?

The HWI is presumably providing information about the aggregate tightness (or slackness) of the local labor market. Note, however, that the education distribution of the Cuban immigrants in each of the shocks, as well as well as the downgrading of their educational skills, suggests that it was the *low-skill* labor market that was most affected by the supply shocks. The response in the HWI to the three supply shocks suggests that perhaps the index is doing a particularly good job at capturing labor market trends in the lower end of the skill distribution. As noted earlier, this tendency may be a mechanical consequence of how classified ads actually appeared in the newspapers sampled by the Conference Board, and in how those ads were counted. In particular, many help-wanted ads for high-skill workers did not appear in the classified pages at all. They sometimes appeared in specialized sections elsewhere in the paper and were not included in the construction of the index.

We use the CPS data to show that changes in the HWI are, in fact, strongly correlated with labor market conditions for the least educated workers in terms of both wages and employment. Consider the following regression model:

$$y_{rst} = \theta_r + \theta_s + \theta_t + \beta_0 \log H_{rt} + \beta_1 (\log H_{rt} \times \theta_s) + \varepsilon, \tag{4}$$

where y_{rst} is a labor market outcome for city *r*, education group *s*, and calendar year *t*; θ_r is a vector of city fixed effects; θ_s is a vector of education fixed effects; θ_t is a vector of calendar year fixed effects; and H_{rt} is the HWI index for city *r* in year *t*. By including the city fixed effects, the coefficient vector (β_0, β_1) is essentially estimating the correlation between a within-city change in the HWI and the corresponding change in labor market outcome y, and how that correlation varies across education groups.

We estimate equation (4) using four education groups: high school dropouts (who have less than 12 years of education), high school graduates (exactly 12 years), some college (13-15 years), and college graduates (at least 16 years). We examine two labor market outcomes: wages and employment. The average wage or employment propensity for cell (r, s, t) is calculated from residuals to individual-level regressions estimated in the CPS data. We use a simple regression model to calculate the age- and gender-adjusted outcome in cell (r, s, t). Specifically, we estimate the following individual-level regression separately in each CPS cross-section:

$$y_{irst} = \alpha_t + \boldsymbol{x_i}\,\gamma_t + \varepsilon, \tag{5}$$

where y_{irst} is the outcome for worker i in city r with education s at time t; and $\boldsymbol{x_i}$ is a vector of fixed effects giving the person's age and gender.[41] We estimate the regressions using both the March CPS (for the calendar years 1972-1999), and the ORG (for the years 1979-1999).[42] The dependent variable in the wage regressions is either the log weekly wage (in the March CPS) or the log hourly wage rate (in the ORG). The employment regressions use a linear probability model where the dependent variable equals one if the person worked at some point in the calendar year (in the March CPS) or during the reference week (in the ORG). The average residual from the regression for cell (r, s, t) gives the age- and gender-adjusted mean outcome in the cell.[43]

Table 5 reports the coefficients in the vector β for alternative specifications of the regression model. Consider initially the regression coefficients in the first row of the top panel of the table. These coefficients measure the correlation between wage trends in the March CPS and

[41] The employment regressions use the sample of persons aged 18-64, while the wage regressions use the sample of workers aged 25-59 (to minimize the sensitivity of the calculated weekly wage to school enrollment or early retirement). The age fixed effects indicate if the person is aged 24 or less, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, or 60-64.

[42] The 1973 March CPS is the first that identifies a relatively large number of metropolitan areas (33), including Miami.

[43] The calculation of the average residual wage in the cell weighs each individual observation by the product of the person's earnings weight times the number of weeks worked in the year (in the March CPS) or the usual number of hours worked weekly (in the ORG).

the HWI. The interactions of the HWI with the education fixed effects indicate that the wage of less-educated workers is more strongly correlated with the HWI than the wage of workers with more education. In fact, the correlation declines monotonically with education. A 10 percent increase in the index is associated with a 1.2 percent increase in the wage of high-school dropouts, a 0.7 percent increase in the wage of high school graduates, a 0.4 percent increase in the wage of workers with some college, and no change in the wage of college graduates. Note that we find a similar pattern in the regression coefficients using the ORG.

The second row of each panel shows the correlation between employment trends and the HWI. The March CPS shows that employment propensities and the HWI are most strongly correlated for the least-educated workers. A 10 percent increase in the index is associated with a 1.1 percentage point increase in the employment probability of high school dropouts (relative to college graduates), a 0.6 increase in the probability for high school graduates, and a 0.4 increase in the probability for workers with some college. The ORG also shows a similar monotonic decline, although the drop is not as steep.

We can combine the estimates of the correlation between the wage and the HWI reported in Table 5 with our earlier estimates of the link between supply shocks and the HWI to back out the implied wage elasticity—the wage decline resulting from an immigration-induced increase in supply. To simplify the discussion, we focus on the experience of low-skill workers after the Mariel supply shock. As we saw, the sudden 20 percent increase in the number of workers without a high school diploma lowered the job vacancy index in Miami by about 40 percent in the early 1980s. This implies that the elasticity measuring the response of the index to an increase in supply $\left(\text{or } \frac{d \log H}{d \log L}\right)$ is –2. At the same time, the March CPS evidence reported in Table 5 suggests that the elasticity giving the change in the wage for a one-percent increase in the HWI $\left(\text{or } \frac{d \log w}{d \log H}\right)$ is about –0.12. The implied wage elasticity is given by the product of these two numbers, or –0.24.

This elasticity sits comfortably in the ballpark of what is usually reported in the literature (Blau and Mackie, 2016, Chapter 5), where the spatial correlations are often estimated to be "small" (often less than –0.2), while the national-level correlations tend to be larger (between –0.3 and –0.6). The –0.24 estimate is also remarkably close to the short-run elasticity implied by the simplest theoretical model that uses an aggregate (and linear homogeneous) Cobb-Douglas production function to examine the impact of immigration in a competitive market. The wage elasticity implied by that model is the (negative of) capital's share of income, which typically hovers around 0.3.

The Content of Help-Wanted Ads

Although our analysis of the HWI documents a short-term decline in the number of job vacancies after a supply shock, the trends in the index do not provide any insight into exactly which types of job vacancies are "vanishing" from the local labor market. To determine the jobs that are most affected, we need to directly explore the *textual* content of the help-wanted ads. To answer such questions in the past would have required a great deal of effort to read and categorize the advertisements. Modern machine learning techniques, however, give us the ability to extract the relevant labor market content of help-wanted classifieds directly from the text.

We illustrate the promise of this approach by conducting a preliminary content analysis of a small sample of classified ads at the time of Mariel using a supervised machine learning approach with gradient boosted trees. Specifically, we examine the text of the 1,454 classifieds that appeared in the *Miami Herald* on two randomly chosen Mondays in the month of December before and after Mariel. The *Herald* published 1,054 help-wanted classifieds on December 18, 1978, and 400 ads on December 20, 1982.[44] We examine the text of these classifieds to document: (a) the types of ads that typically appeared in the help-wanted section of that newspaper; and (b) whether there was a relative decline in classifieds for low-skill workers in Miami. Our evidence suggests that low-skill positions comprised almost three-quarters of the advertised jobs at the time, and that there was a slight decline in the fraction of ads advertising low-skill vacancies after Mariel.

Our supervised machine learning analysis involves three steps: (1) human coding of a subsample (also known as the "training data") of help-wanted ads into low- and high-skill vacancies; (2) using the human-coded data to train, fine-tune, and test the algorithm that will be used to classify ads into low- and high-skill vacancies; and (3) assigning all ads outside of the training data to one of the two skill groups.

---

[44] We first picked a random date 2 to 3 years after Mariel. Once that date was determined, we worked backwards to pick a similarly situated date before Mariel. The randomly chosen date in the post-Mariel period coincides with the beginning of the recovery from the 1982 recession.

All machine learning algorithms are inherently data-driven and require high-quality data to accurately perform classification tasks. Hence, we first need to acquire high-quality, human-coded classified ads to teach the algorithm how to classify a particular ad into a specific skill group. We selected a random sample of 20 percent (300) of the 1,454 classifieds in our sample. The text of these classifieds, with the dates stripped, were then presented, independently, to two human coders who identified them as belonging to either a low-skill or a high-skill occupation. The coders were instructed to classify jobs as low-skill if the advertised job was in the list of the 50 largest occupations that employed workers without a high school diploma in Miami at the time.[45] The coders agreed on the classification of 89 percent of the advertised vacancies. The conflicts were resolved by re-presenting the coders with the text of the ads that they disagreed on. In the end, the coders identified 66 percent of the classifieds as low-skill vacancies.[46]

The random sample of human-coded classifieds was then used to train a gradient-boosted trees classifier to distinguish between ads using only the words in the ads. Gradient-boosted trees are a popular type of decision tree algorithm which, like their random forests predecessor, grows multiple trees from random subsets of the training data and use a "majority vote" of the trees to generate the final class label. They have become popular in the economics literature and have been used for text classification because of their transparency and their ability to generate highly accurate predictions in various contexts (Athey, Tibshirani and Wage, 2016; Chalfin et al., 2016; Chen and Guestrin, 2016; and Kleinberg et. al., 2018). Gradient-boosted trees tend to exhibit significantly improved classification performance over vanilla random forests because they have several hyperparameters that can be fine-tuned using cross-validation methods.

Training the algorithm to identify low- and high-skill ads in the training data involved the following steps: (1) text pre-processing; (2) conversion of text into a document-term matrix; (3) algorithm training and fine-tuning via cross-validation; and (4) performance assessment on the test data.

---

[45] The list was calculated using the 1980 decennial census. The 3 largest occupations in the list are "textile sewing machine operators," "truck, delivery, and tractor drivers," and "janitors." The 3 smallest occupations are "helpers, constructions," "bus drivers," and "garage and service station related occupations." The 50 occupations accounted for 77.5 percent of all employment among high school dropouts.

[46] In contrast to our evidence from the *Miami Herald*, direct content analysis of help-wanted classifieds conducted in Walsh, Johnson, and Sugarman (1975), and summarized in Abraham (1987) and Zagorsky (1998), suggests that a single day's cross-section of job advertising over-represents high-skill "white-collar" vacancies relative to the actual number of such vacancies available.

The text pre-processing stage involves standardizing the text of the ads so that only the words (or parts of words) with the highest amount of useful information are retained (Denny and Spirling, 2018; Gentzkow, Kelley, and Taddy, 2017; and Grimmer and Stewart, 2013).[47] The processed text is then converted into a document-term matrix. The 300 rows in this matrix contain each classified ad in the training data, and the 1,355 columns contain the number of "cleaned" words left after text pre-processing. The entry in each cell of the matrix is the number of times the word appeared in the help-wanted ad.

The training process then involves randomly selecting a training and test set. We opt for a 75/25 train/test split which is the default settings on most machine learning software packages (Pedregosa et. al. 2011). Model training involves prediction of the low- or high-skill classification for each ad in the training data using only the words contained in the document term matrix. This is accomplished through growing multiple trees via an iterative loss minimization process using an objective function, $O(\theta)$, which is comprised of a logistic regression loss function $\mathrm{L}(\theta)$ of the tree parameters θ and a regularization term, $\Gamma(f_k)$, which is a function of the number of k trees grown where each tree is represented by a function $f_k \in F$ in the function space F of all possible trees:

$$(6) \qquad O(\theta) = \mathrm{L}(\theta) + \sum_{k=1}^{K} \Gamma(f_k) = \sum_{i=1}^{T} l(c_i, c_i^p) + \sum_{k=1}^{K} \Gamma(f_k).$$

The goal of training is to minimize $O(\theta)$ by simultaneously accounting for the difference between the true and predicted skill classification for the classified ads (c_i and c_i^p) and the regularization term $\sum_{k=1}^{K} \Gamma(f_k)$ which prevents overfitting of the model.

An important part of the training process involved hyper-parameter tuning using 10-fold cross-validation on the training data to select the model with the minimum average cross-validated test error as defined by the objective function in equation (6). Appendix Figure A1 gives examples of two of the trees grown for the final model; Appendix Figure A2 plots the

---

[47] The pre-processing stage has to be tailored for the specific type of text used in the analysis. The "pipeline" through which we passed our text involved: (1) "tokenization," or separation of the words in the classifieds into unigrams (single words); (2) removal of stop words (such as "the") that add little information; (3) removal of all special characters, numbers and punctuation; and (4) "stemming," which is a natural language processing method that removes the suffixes from terms (e.g., "secretaries"; becomes "secretari").

average cross-validated training and test error estimated during the training process; and Appendix Table A2 presents the confusion table.

Table 6 summarizes the performance of the classifier in the test data. The classifier has a decent accuracy rate of 72 percent. It is particularly good at identifying low-skill ads. The sensitivity rate (or the conditional probability that the classifier assigns a low-skill ad correctly) is 80.4 percent, while the specificity rate (or the conditional probability that the classifier assigns a high-skill ad correctly) is 54.0 percent. The table also reports the results from applying the classifier to the remaining 1,154 ads in our sample. This exercise reveals a slight decline in the relative number of low-skill ads published in the *Miami Herald* between 1978 and 1982, although the decline in our small sample of classifieds is not statistically significant.[48] In particular, 777 of the 1,054 ads in 1978 (or 73.7 percent) were for low-skill vacancies. In 1982, 285 of the 400 ads (or 71.3 percent) were for low-skill vacancies.[49]

One of the benefits of tree-based classifiers is that they provide some understanding of how the classifier makes its decisions. In our context, the tree provides information about which *words* the classifier used to best distinguish between low- and high-skill vacancies. Figure 15 plots the terms (ranked from most to least important) that help distinguish classifieds using the information gain across trees. Interestingly, terms such as "firm", "salari" and "benefit," which are more likely to appear in high-skill ads, help to best distinguish between the classifieds.

It is important to reemphasize that we have only analyzed a small sample of classified ads. Nevertheless, this type of textual analysis seems like an important first step in any attempt to understand what an index of job vacancies based on help-wanted ads measures, and which types of jobs may be most affected by low-skill immigration.

[48] Although it is not difficult to conduct the statistical analysis on a much larger sample of help-wanted ads, there is still the formidable task of converting images of the classified pages from archived newspapers into a text format that can be easily parsed and manipulated. We are in the process of increasing the size of the textual sample of classifieds in Miami and of constructing comparable samples in other cities.

[49] We also examined the occupational distribution in the large sample of help-wanted classifieds compiled by Atalay et al (2018). They assigned a census occupation code to ads appearing in the *New York Times*, the *Wall Street Journal*, and the *Boston Globe*. We calculated the fraction of ads in those three newspapers that are in the list of the 50 largest occupations employing high school dropouts in Miami in 1980. The share of ads that advertised such low-skill vacancies is far lower than in the *Miami Herald*, and this share barely changed between 1978 (27.9 percent) and 1982 (27.7 percent).

VI. Beyond Miami

Up to this point, our analysis focused on documenting the response of the HWI to (presumably) exogenous supply shocks that hit the city of Miami between 1960 and 2000. Immigration also affected many other cities in those decades, as the annual number of immigrants entering the United States increased from about 250,000 in the 1950s to over 1 million legal and illegal immigrants by 2000. It is well known that these immigrants clustered in a relatively small number of metropolitan areas, although no other city experienced the sudden and large supply shocks that hit Miami in 1960-1962, 1980, and 1994-1995.

Beginning with Grossman (1982), a large literature in labor economics exploits the geographic distribution of immigrants to estimate the labor market impact of immigration. This literature essentially calculates "spatial correlations," a correlation between some economic outcome in the city—typically the average wage of some group of workers—and the number of immigrants settling in that city. Although these studies often report that the spatial correlation is relatively weak (suggesting that native outcomes are not strongly affected by immigration), it is well known that the correlation is contaminated by the endogenous settlement of immigrants in high-wage cities. This sorting makes it difficult to detect the potential adverse effect of supply shocks on the wage of competing workers. Moreover, it seems that the widely used instrument of (some variation of) lagged immigration does not solve the endogeneity problem and understates the adverse wage impact of supply shocks (Jaeger, Ruist, and Stuhler, 2018).

Given the dominance of this methodological approach in the literature (and the rarity of "experimental" supply shocks like the ones that hit Miami), it is of interest to estimate the analogous spatial correlation between the HWI and immigration. Ideally, we would have monthly or yearly data on the number of immigrants arriving in each of the cities surveyed by the Conference Board over the four-decade period. Unfortunately, it was not until 1994 that the CPS began to collect information on a person's immigration status. Moreover, the reliability of the HWI index deteriorated in the late 1990s with the explosion of the Internet and on-line job postings. The CPS data, therefore, would only let us analyze a very short time series, making it unlikely that any credible spatial correlation could be detected.[50]

[50] The monthly data from the Job Openings and Labor Turnover Survey (JOLTS) conducted by the BLS since 2000 could, in principle, be used to estimate the spatial correlations in the post-2000 period. Unfortunately, the data are not available at the metropolitan area level. In fact, the JOLTS vacancy rate series

We pursue the alternative of using the decennial censuses to calculate the immigration-induced supply shock in each of the cities. We then correlate these supply shocks with the decadal change in the city's HWI. The regression model is given by:

$$(7) \qquad \log\frac{H_{r,\tau}}{H_{r,\tau-1}} = \gamma\frac{\textit{Immigration into city } r}{\textit{Baseline number of natives in city } r} + \theta_\tau + \eta,$$

where $H_{r,\tau}$ gives the HWI for city r in census year τ (τ = 1960,…,2000). The HWI index for census year τ is defined as the average HWI observed in the three-year interval around τ. For example, the average HWI for Rochester in census year 1980 is the average HWI reported monthly for Rochester between 1979 and 1981. The two variables used to measure the supply shock (the number of immigrants, and the number of natives) give population counts for persons aged 18-64 in a particular city. There seems to be a lot of confusion in the literature as to how exactly the ratio defining the supply shock in equation (7) should be defined (Borjas, 2014; Borjas and Monras, 2017; and Card and Peri, 2016). We will use alternatives measure of the native baseline to demonstrate the robustness of the estimated spatial correlation.

It is important to emphasize that the same endogeneity problem that plagues estimates of the spatial correlation between the city's wage and immigration plagues the regression in equation (6). Immigrants are more likely to settle in cities where labor demand is strong, and employers are actively (and spending money) looking for workers. This endogeneity builds in a positive correlation between the change in the HWI and the number of immigrants settling in that city during the decade.

Table 7 reports estimates of the coefficient γ using a number of alternative specifications. Consider initially the regression reported in the first column of Panel A. The supply shock in this panel is defined as the ratio of the number of immigrants who migrated to the city between census year τ-1 and τ (or $M(\tau, \tau\text{-}1)$) to the number of natives residing in the city at time τ-1 (or $N(\tau\text{-}1)$).[51] This specification ignores the fact that there may have been a supply response as natives either moved in or out of immigrant-receiving cities. The first row of the table uses a

are only publicly released for four broad census regions (the Northeast, the South, the Midwest, and the West), greatly reducing the possibility that the data can be used to estimate credible spatial correlations.

[51] The number of immigrants who migrated to the United States in the interval (τ−1, τ) and settled in city r is obtained from the decennial census in year τ.

supply shock that aggregates across all skill groups, so that the supply shock simply gives the percent increase in the city's adult population (relative to the initial native population). The estimated spatial correlation is negative and significant, suggesting that an immigration-induced 10 percent increase in supply is associated with a reduction in the HWI of about 9 percent.

The second row in Panel A uses *education-specific* measures of the supply shock, and each of the four education-specific measures is introduced into the regression one at a time. For example, the coefficient for high school graduates comes from a regression where the supply shock is defined as the number of high school graduates who immigrated to city r in the interval (τ - 1, τ) divided by the number of native high school graduates residing in that city in census year τ - 1. When these education-specific supply shocks are entered individually, each of the coefficients is negative and significant. As the third row of the panel shows, however, the simultaneous inclusion of all four education-specific supply shocks leads to a regression where none of the supply shocks is significant. This insignificance is likely attributable to the multicollinearity among the supply variables. The correlation coefficient between the total supply shock and each of the immigration-specific supply shocks ranges from 0.86 to 0.97. Immigrant-receiving cities tend to receive all types of immigrants, both low- and high-skill.

The last row of Panel A addresses the endogeneity issue by using the "shift-share" instrument that is commonly employed in the literature. In rough terms, the instrument uses the geographic settlement of earlier immigrant waves belonging to particular national origin groups to allocate new immigration across the cities. The validity of the instrument obviously hinges on whether the local economic conditions that attracted immigrants to a particular city persist for some time after arrival. We use the instrument for the total immigrant share for each city calculated by Jaeger, Ruist, and Stuhler (2018).[52] The IV estimate of the impact is negative and significant, and larger in absolute value than the OLS estimate reported in the first row of the panel. Interestingly, the estimated elasticity $\left(\frac{d \log H}{d \log L}\right)$ is about −2, which is similar to the elasticity implied by the difference-in-differences analysis of the Mariel episode reported earlier.

The other panels of the table use alternative definitions of the supply shock in equation (4). In Panel B, the supply shock is the ratio $M(\tau, \tau\text{-}1)/N(\tau)$. It differs from the definition in Panel

[52] We are very grateful to David Jaeger, Joakim Ruist, and Jan Stuhler for generously sharing their data. We use the version of their instrument that uses the settlement of immigrants as of 1960 to predict the geographic sorting of new arrivals in each decade.

A because it does not lag the native baseline (thereby allowing for a potential native supply response). Finally, in Panel C, the supply shock is the difference $\left(\frac{M_{c,\tau}}{N_{c,\tau}} - \frac{M_{c,\tau-1}}{N_{c,\tau-1}}\right)$.[53] Regardless of how we define the supply shock, Table 7 shows a negative correlation between the change in the HWI and aggregate immigration, and a negative correlations between the change in the HWI and each education-specific measure of the supply shock.

In short, our estimate of the spatial correlation between the HWI and immigration confirms the conclusion implied by the experimental supply shocks that hit Miami: Help-wanted classified ads, which presumably measure the availability of jobs in a local labor market, tend to fall in those markets receiving large numbers of immigrants.

VII. Summary

This paper addresses what is perhaps the central economic question in the immigration debate: How do supply shocks affect the labor market in receiving countries? We contribute to the literature by exploiting a data set that has a long history in economics but has never been employed in the immigration context before, the Conference Board Help-Wanted Index (HWI).

Beginning in 1951, the Conference Board constructed an index of job vacancies in local labor markets by counting the number of help-wanted classified ads in newspapers in 51 metropolitan areas. It is well known that the HWI provides valuable information about trends in local labor demand and is highly correlated with various measures of labor market conditions.

We use the HWI to document how immigration affects job vacancies in labor markets affected by supply shocks. We exploit both natural experiments created by random and sudden labor supply shocks in the city of Miami, as well as estimate spatial correlations that measure how local trends in job vacancies are related to immigration-induced supply shifts. Our findings include:

1. The labor market in Miami responded strongly to the Mariel boatlift in 1980. The HWI dropped relative to the trend observed in many alternative control groups between 1980 and 1983. Miami's HWI recovered fully by 1990.

[53] If there were only two cross-sections, the regression relating the first-difference in the HWI and the first-difference in the ratio M/N is equivalent to a panel regression where the level of the HWI in city r and census year τ is regressed on sets of city and year fixed effects and on the ratio M/N.

2. Miami's labor market also reacted to the initial wave of Cuban refugees that arrived between 1960 and 1962 (before the Cuban missile crisis abruptly stopped the flow). The HWI in Miami dropped relative to the trend observed in control cities, before beginning to recover by the mid-1960s. This recovery, however, was stalled by the arrival of a new wave of Cuban immigrants in the last half of the 1960s.
3. Miami saw a marked increase in the HWI as a result of Hurricane Andrew in mid-1992. But soon after Miami's labor market recovered from the aftermath of Andrew, the city was hit yet again by another unexpected supply shock of Cuban immigrants in 1994-1995. Miami's relative HWI again dropped after this supply shock. The index, however, did not recover by the end of the 1990s, perhaps because there was a steady (and increasing) flow of Cuban immigrants after the 1994-1995 shock.
4. There is a negative cross-city correlation between the change in the HWI and the number of immigrants entering the local labor market. The measured spatial correlation is negative and significant despite the obvious endogeneity bias created by the non-random settlement of immigrants in cities where there are job openings.
5. Many of the immigration-induced supply shocks we examined disproportionately increased the size of the low-skill workforce. Although the HWI presumably provides some measure of "average" local labor market conditions, the persistent negative correlation between the HWI and immigration suggests that the index might be a particularly good barometer for labor market conditions at the bottom end of the skill distribution. Our analysis indeed indicates that the HWI seems to be more strongly correlated with wage and employment trends for the least educated workers. And a textual analysis of a small sample of classifieds in the *Miami Herald* before and after Mariel also suggests a relative decline in the number of ads for low-skill job vacancies.

In sum, our evidence consistently indicates that immigration-induced supply shocks are typically followed by a short-run period of slackness in the local labor market, as measured by the number of advertised job openings. The labor market, however, tends to recover after a few years. In the absence of any additional supply shocks, the local labor market returns to its pre-immigration equilibrium within 5 to 10 years.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association* 105 (June 2010): 493-505.

Abraham, Katharine G. 1987. "Help-Wanted Advertising, Job Vacancies, and Unemployment," *Brookings Papers on Economic Activity* (1): 207-243.

Abraham, Katharine G. and Lawrence F. Katz. 1986. "Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?" *Journal of Political Economy*, Vol. 94, No. 3, Part 1.

Andolfatto, David. 1996. "Business Cycles and Labor-Market Search," *American Economic Review*, 86(1).

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. 2018. "The Evolving U.S. Occupational Structure," Working Paper. University of Wisconsin, Madison.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2016. "Generalized Random Forests." *arXiv preprint arXiv:1610.01271*.

Autor, David H. 2001. "Wiring the Labor Market," *Journal of Economics Perspectives* 15 (Winter 2001): 25-40.

Belasen, Ariel R. and Solomon W. Polachek, "How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida," *Journal of Human Resources* 44 (Winter 2009): 251-276.

Blei, David M. 2012. "Probabilistic topic models." *Communications of the ACM* 55 (4): 77-84.

Blau, Francine D. and Christopher Mackie, editors. 2016. *The Economic and Fiscal Consequences of Immigration*. Washington, DC: National Academies Press, 2016.

Borjas, George J. 2003. "The Labor Demand Curve *Is* Downward Sloping: Reexamining the Impact of Immigration on the Labor Market,' *Quarterly Journal of Economics* 118 (November 2003): 1335-1374.

Borjas, George J. 2016. "The Wage Impact of the Marielitos: Additional Evidence," NBER Working Paper No. 21850, January 2016.

Borjas, George J. 2017. "The Wage Impact of the Marielitos: A Reappraisal," *Industrial and Labor Relations Review*, 70(5): 1077-1110.

Borjas, George J. 2017. "Still More on Mariel: The Role of Race," NBER Working Paper No. 23504, June 2017.

Borjas, George J. and Joan Monras. 2017. “The Labor Market Consequences of Refugee Supply Shocks,” *Economic Policy* 32: 361-413.

Burch, Susan W. and Fabricant, Ruth A. 1968. “A Further Comment on the Behavior of Help-Wanted Advertising,” *Review of Economics and Statistics*, Vol. 50, No. 2.

Card, David. 1990. “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review* 43 (January 1990): 245-257.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "Productivity and selection of human capital with machine learning." *American Economic Review* 106, no. 5 (May 2016): 124-27.

Chassamboulli, Andri and Giovanni Peri. 2015. “The Labor Market Effects of Reducing Undocumented Immigrants,” *Review of Economic Dynamics* 18 (October 2015): 792-821.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 13-17, 2016: 785-794.

Clemens, Michael A., and Jennifer Hunt. “The Labor Market Effects of Refugee Waves: Reconciling Conflicting Results,” NBER Working Paper No. 23433, May 2017.

Cohen, Malcolm S. and Robert M. Solow. 1967. “The Behavior of Help-Wanted Advertising”, *Review of Economics and Statistics*, 49(1).

Courtney, Hugh. 1991. “The Beveridge Curve and Okun's Law: A Re-Examination of Fundamental Macroeconomic Relationships in the US,” unpublished Ph.D., dissertation, MIT.

Davila, Alberto and Saenz, Rogelio. 1990. “The Effect of Maquiladora Employment on the Monthly Flow of Mexican Undocumented Immigration to the U.S., 1978–1982,” *International Migration Review*, 24 (February 1990): 96–107.

DeBoer, Larry and Seeborg, Michael. 1984. “The Female-Male Unemployment Differential: Effects of Changes in Industry Employment,” *Monthly Labor Review*, 107.

Denny, Matthew, and Arthur Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, And What to Do About It," *Political Analysis*: 1-22.

Elsby, Michael W., Ryan Michaels and David Ratner. 2015. “The Beveridge Curve: A Survey,” *Journal of Economic Literature* 53 (September 2015): 571-630.

Friedman, Susan Krug. 1982. “The Help-Wanted Advertising Index: A Leading Indicator of the Local Economy,” *Business Economics* 17 (May 1982): 61-64.

Gelman, Andrew and Eric Loken. 2013. “The Garden of Forking Paths: Why multiple comparisons can be a problem, even when there is no ‘fishing expedition’ or ‘p-hacking’ and the research hypothesis was posited ahead of time,” Columbia University, Working Paper, November 14, 2013.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. 2017. “Text as Data,” NBER Working Paper No. 23276, March 2017.

Goodwin, William B. and Carlson, John A. 1981. “Job-Advertising and Wage Control Spillovers,” *Journal of Human Resources* 16 (Winter 1981): 80–93.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (Summer 2013): 267-297.

Grossman, Jean Baldwin. 1982. “The Substitutability of Natives and Immigrants in Production,” *Review of Economics and Statistics* 64 (November 1982): 596-603.

Hughes, Joyce. 1999. “Flight from Cuba,” *California Western Law Revie*w: Vol. 36: No. 1, Article 4, pp. 39-75.

Jaeger, David A., Joakim Ruist, and Jan Stuhler. 2018. “Shift-Share Instruments and the Impact of Immigration,” NBER Working Paper No. 24285.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions," *Quarterly Journal of Economics.* 133 (February 2018): 237-293.

Kroft, Kory and Devin G. Pope, 2014. “Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist,” *Journal of Labor Economics* 32, pp. 259-314.

Lim, Chloe. 2017. “Checking How Fact-Checkers Check,” Stanford University, Working Paper, May 16, 2017.

Medoff, James L. 1983. “U.S. Labor Markets: Imbalance, Wage Growth, and Productivity in the 1970s,” *Brookings Papers on Economic Activity*, pp. 87-120.

Ortega, Javier. 2000. “Pareto-Improving Immigration in an Economy with Equilibrium Unemployment,” *Economic Journal* 110 (January 2000): 92–112.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (October 2011): 2825-2830.

Peri, Giovanni and Yasenov, Vasil. 2015. "The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift," No. w21801, National Bureau of Economic Research.

Pholphirul, Piriya. 2013. "Immigration, Job Vacancies, and Employment Dynamics: Evidence from Thai Manufacturers," *Journal of Asian Economics* 24 (February 2013): 1-16.

Pissarides, Christopher A. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages," *American Economic Review* 75 (September 1985): 676-690.

Preston, Noreen L. 1977. *The Help-Wanted Index: Technical Description and Behavioral Trends*. Report no. 716. New York: Conference Board, 1977.

Schwartz, Arthur R., Malcolm S. Cohen, and Donald R. Grimes. 1986. "Structural/Frictional vs. Deficient Demand Unemployment: Comment," *American Economic Review* 76 (March 1986): 268-272.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review* 95 (March 2005): 25-49.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science* 22(11): 1359-1366.

Steegan, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. "Increasing Transparency Through a Multiverse Analysis," *Perspectives on Psychological Analysis* 11 (September 2016): 702-712.

Walsh, John, Miriam Johnson, and Sugarman, Marged. 1975. *Help-Wanted: Case Studies of Classified Ads*. Salt Lake City: Olympus, 1975.

Warren, Donald. 1982. "Immigration and the Natural Rate of Unemployment in Australia," *Journal of Macroeconomics* 4 (Fall 1982): 449-457.

Withers, Glenn and David Pope. 1985. "Immigration and Unemployment," *Economic Record* 61 (June 1985): 554-564.

Zagorsky, Jay L. 1993. "Understanding the Economic Factors Affecting Help-Wanted Advertising," *Journal of Advertising* 22 (September 1993): 75-84.

Zagorsky, Jay L. 1998. "Job Vacancies in the United States: 1923 to 1994," *Review of Economics and Statistics*, 80(2).

Figure 1. The Help-Wanted Index (HWI) and the unemployment rate

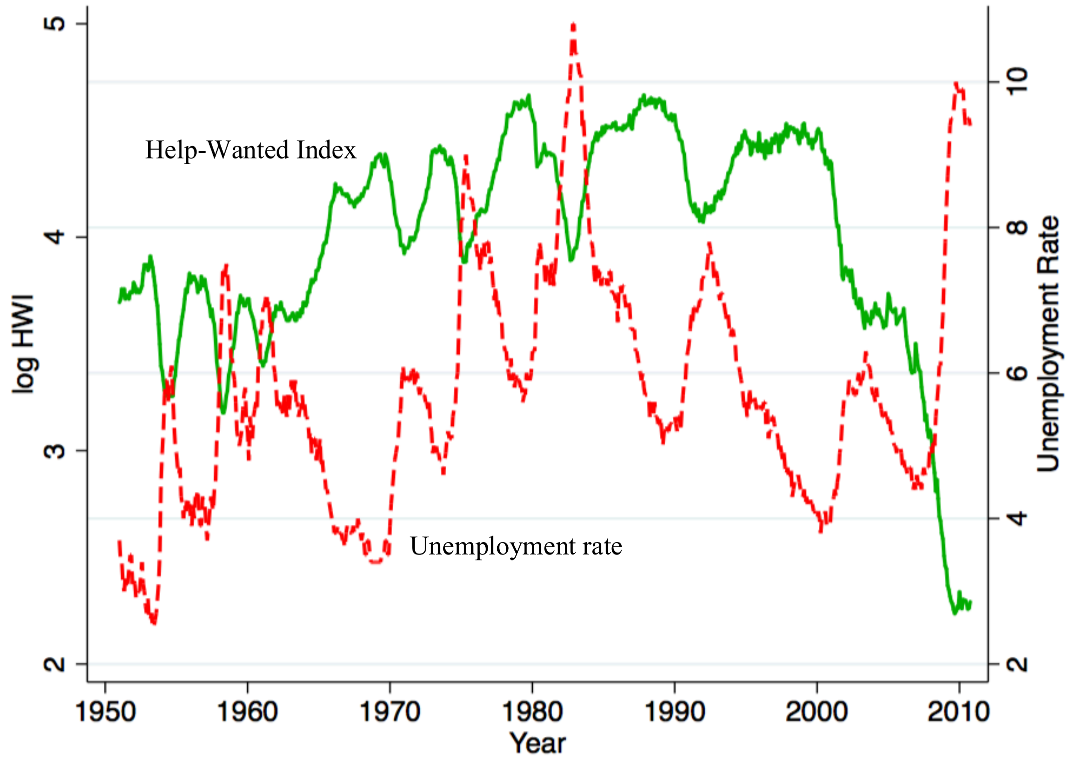


Figure 2. Enumeration biases in the Help-Wanted Index

a. A single ad with multiple job listings

TEST TECHNICIANS

MANUFACTURER of 2-way radios has several openings for Test Technicians for the daytime shift. Must have experience FCC license helpful. Call for interview 445-2671

COMCO

AVE

445-

Equal Opportunity Employer

b. An ad posted by an employment agency

SOUTH MIAMI EMPLOYMENT

PERSONNEL

FEE paid. A plush opening to work as a placement advisor. Excellent chance for a person who has a sales personality and is interested in helping people. Top money rewards your efforts.

NW ASSISTANT MANAGER

INTERNATIONAL firm needs sharp person to train, some college, supervisory and auto parts helpful. Equal opportunity. To $800

FIELD SERVICING

FEE paid. Miami area. 2 year degree electronics or military equivalency, experienced medical equipment. To $12,000

MANAGER

FEE paid. Broward county. Must have savings and loan experience. To $16,000

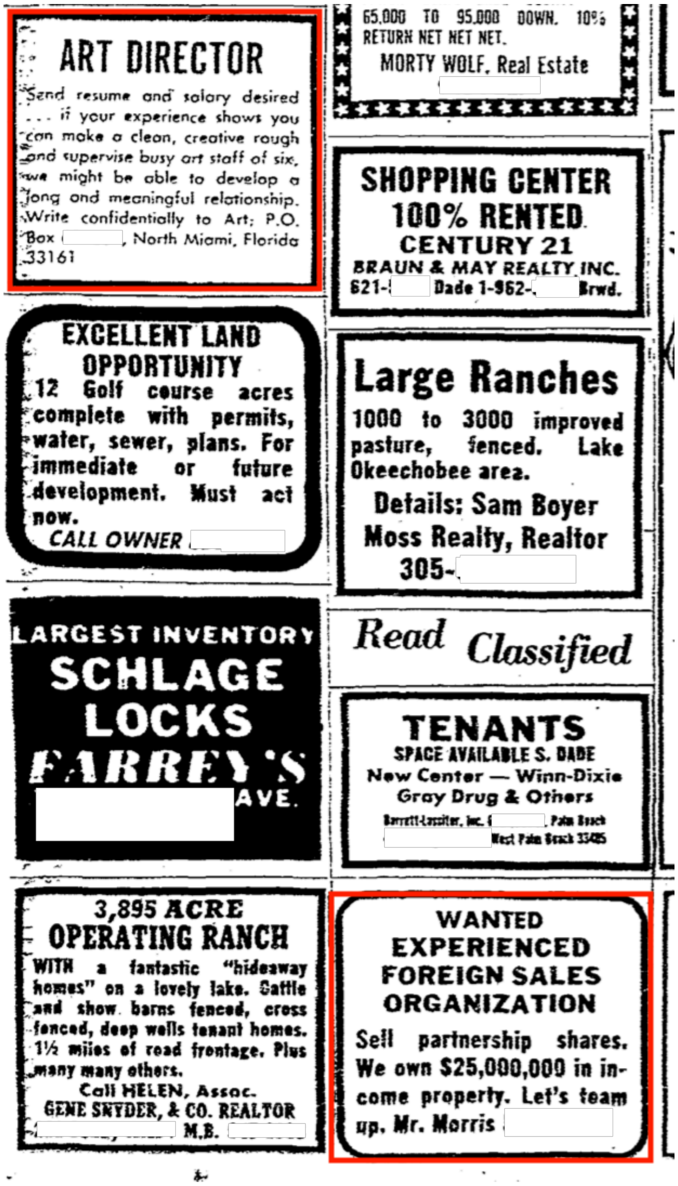
OFFICE MANAGER

SUPERVISORY experience. Worldwide firm will teach their operation to right person. Capable of responsibility. Top skills. Aptitude for figures. Split fee. To $200

RESUME SERVICE

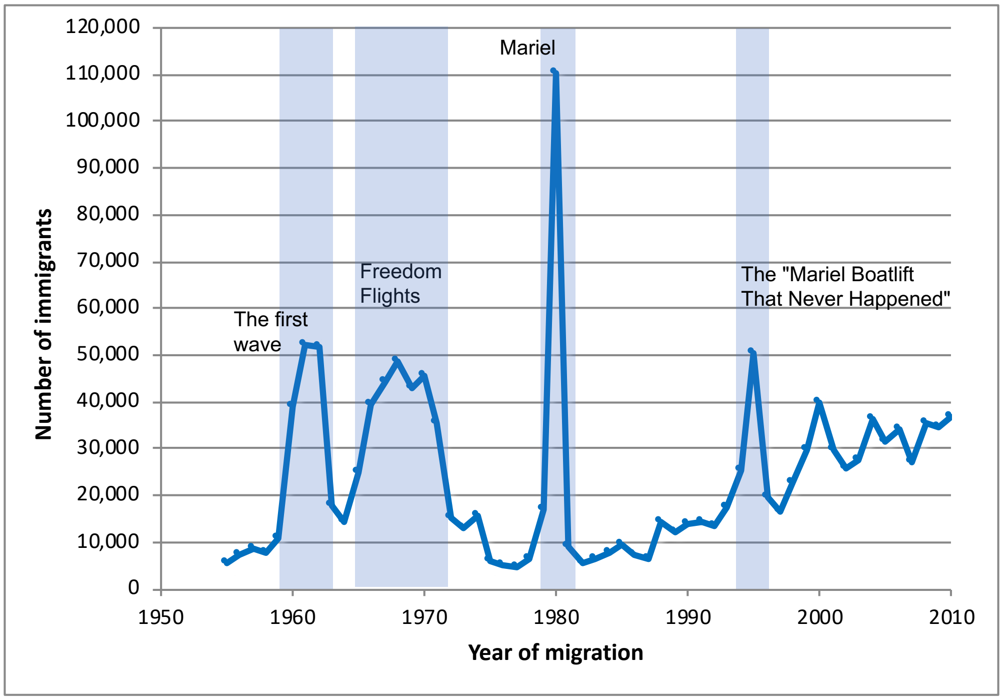
Many positions in all areas

c. Ads posted in the FIRE section of the newspaper



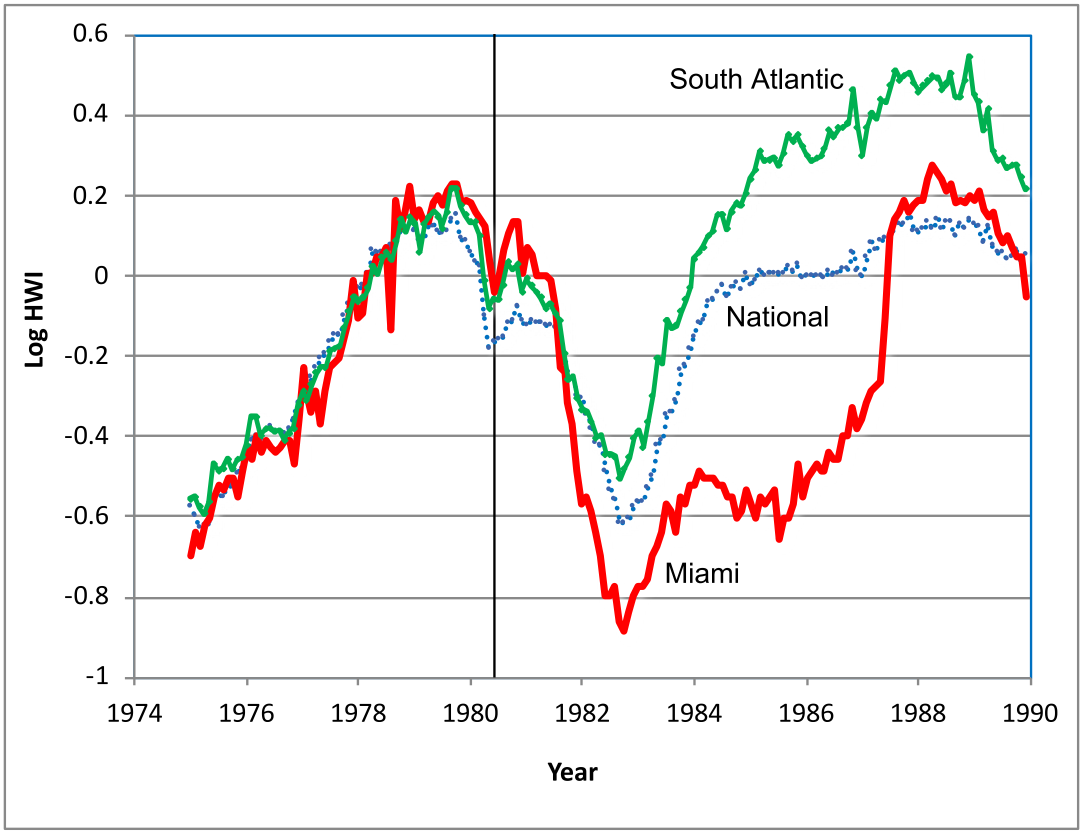
Notes: All ads appeared in the March 2, 1975 edition of the *Miami Herald*.

Figure 3. Cuban immigration to the United States, 1955-2010



Source: Adapted from Borjas (2017), p. 1080.

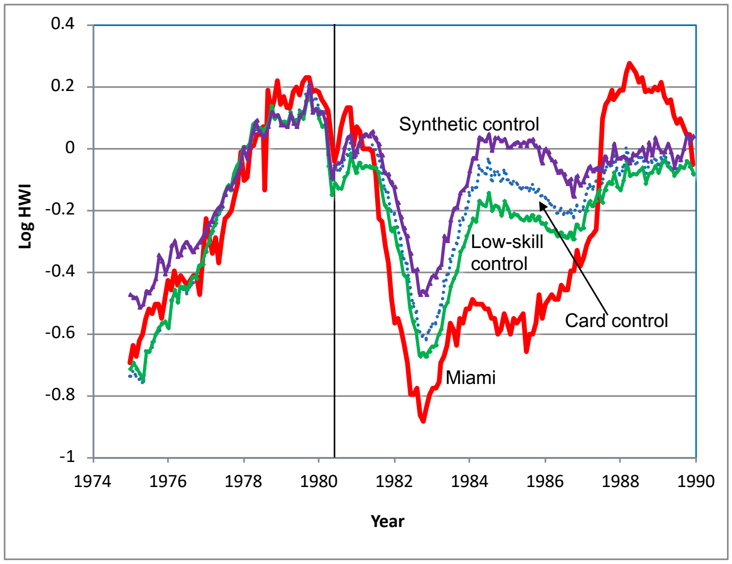
Figure 4. The Help-Wanted Index in Miami, 1975-1989



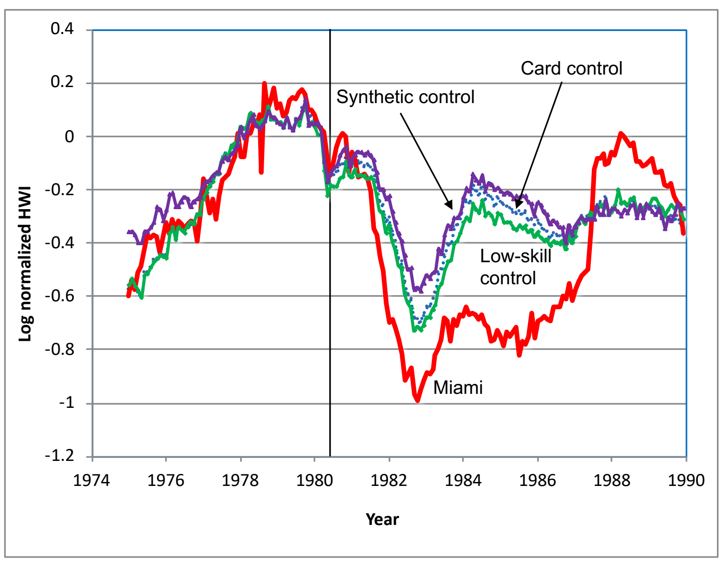
Notes: The HWI for each city/region is rescaled to equal 1 in 1977-1979. The treatment line is drawn as of June 1980.

Figure 5. Job vacancies in Miami relative to control cities, 1975-1989

A. The Help-Wanted Index

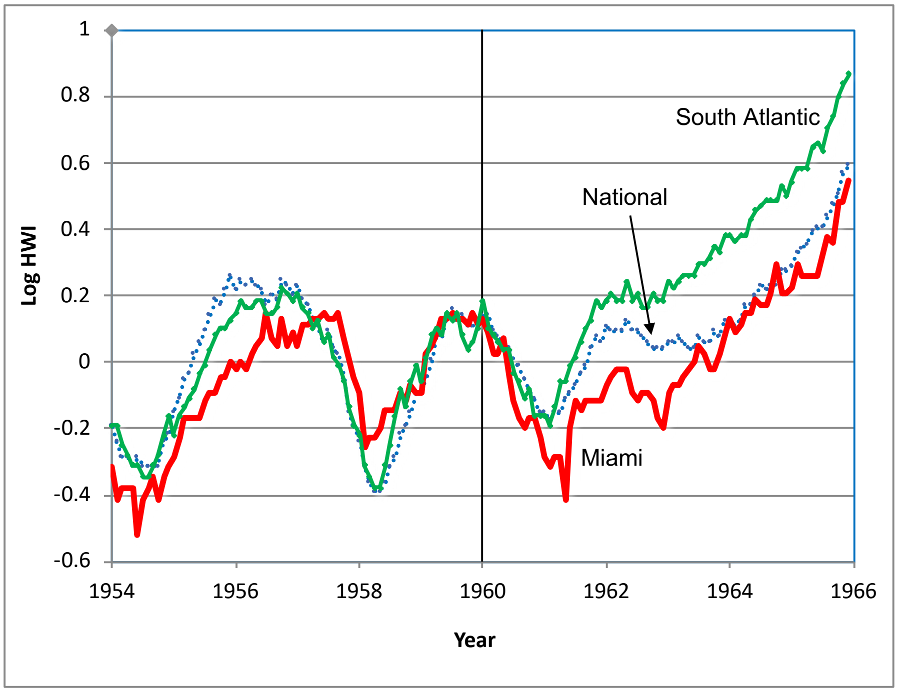


B. The normalized HWI



Notes: The normalized HWI is defined as the HWI divided by non-agricultural employment in the city-year-month cell. Both the HWI and the normalized index for each city are rescaled to equal 1 in 1977-1979. The treatment line is drawn as of June 1980.

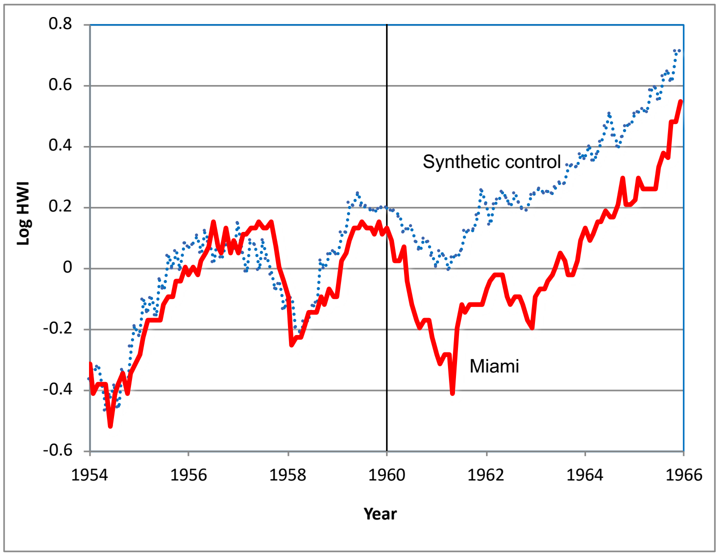
Figure 6. The Help-Wanted Index in Miami, 1954-1965



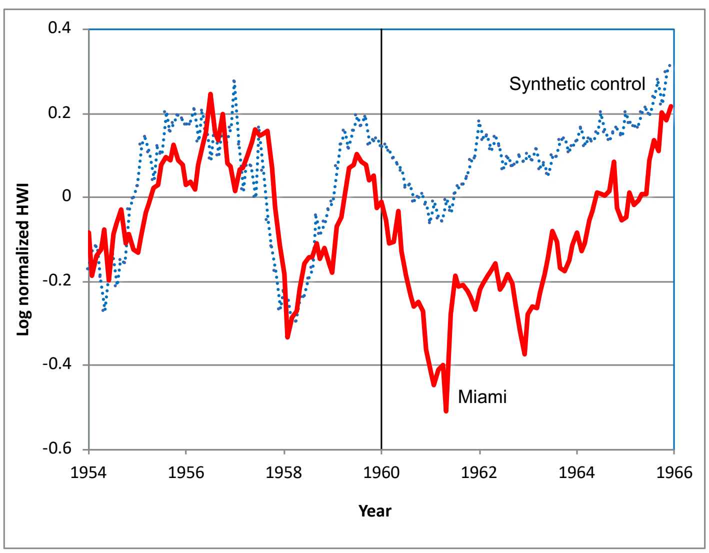
Notes: The HWI for each city/region is rescaled to equal 1 in 1956-1958. The treatment line is drawn as of January 1960.

Figure 7. Job vacancies in Miami relative to the synthetic control, 1954-1965

A. The Help-Wanted Index



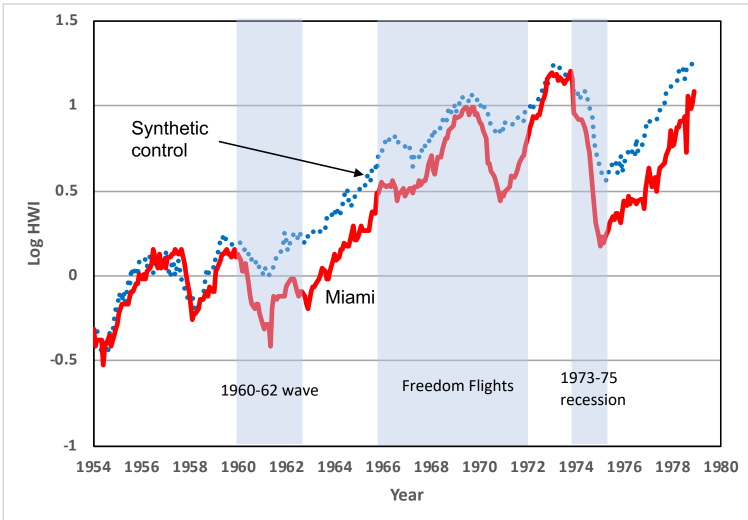
B. The normalized HWI



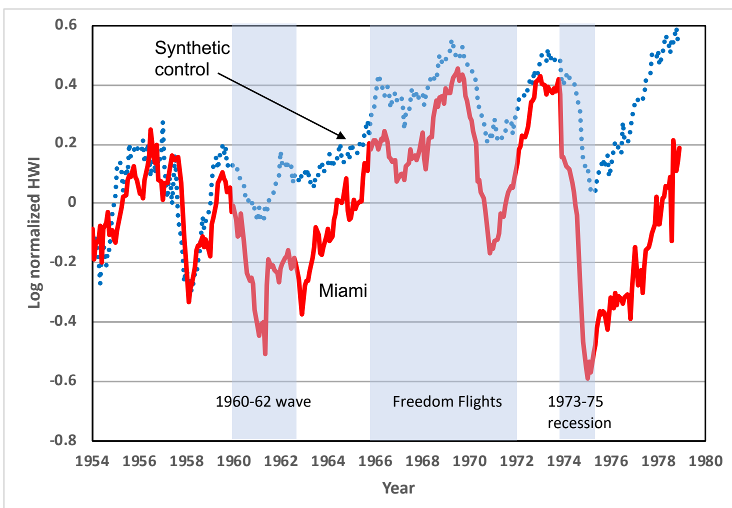
Notes: The normalized HWI is defined as the HWI divided by non-agricultural employment in the city-year-month cell. Both the HWI and the normalized index for each city are rescaled to equal 1 in 1956-1958. The treatment line is drawn as of January 1960.

Figure 8. Job vacancies in Miami relative to the synthetic control, 1954-1978

A. The Help-Wanted Index

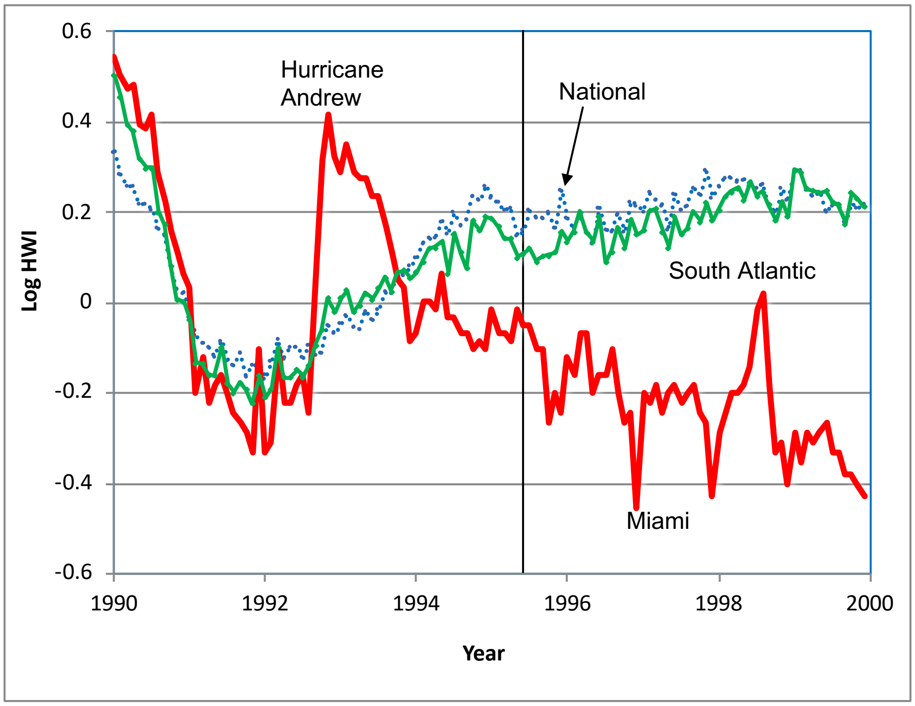


B. The normalized HWI



Notes: The normalized HWI is defined as the HWI divided by non-agricultural employment in the city-year-month cell. Both the HWI and the normalized index for each city are rescaled to equal 1 in 1956-1958.

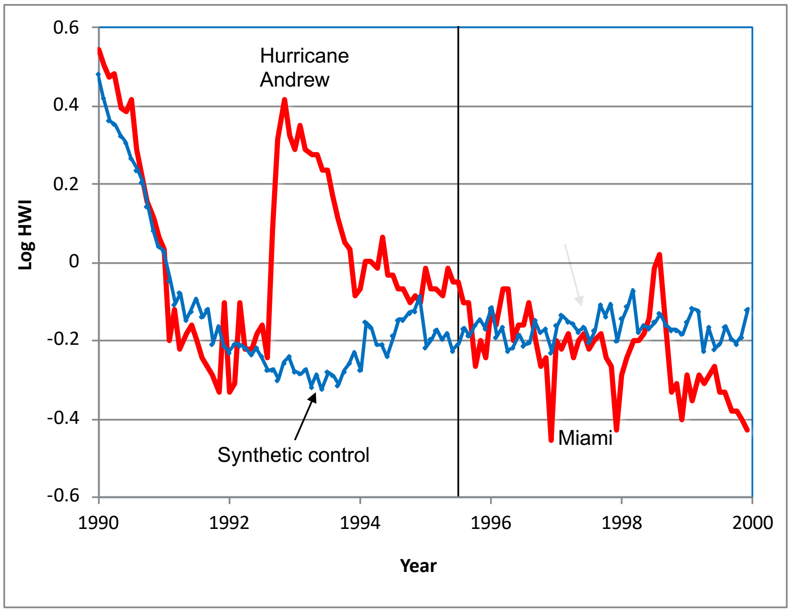
Figure 9. The Help-Wanted Index in Miami, 1990-1999



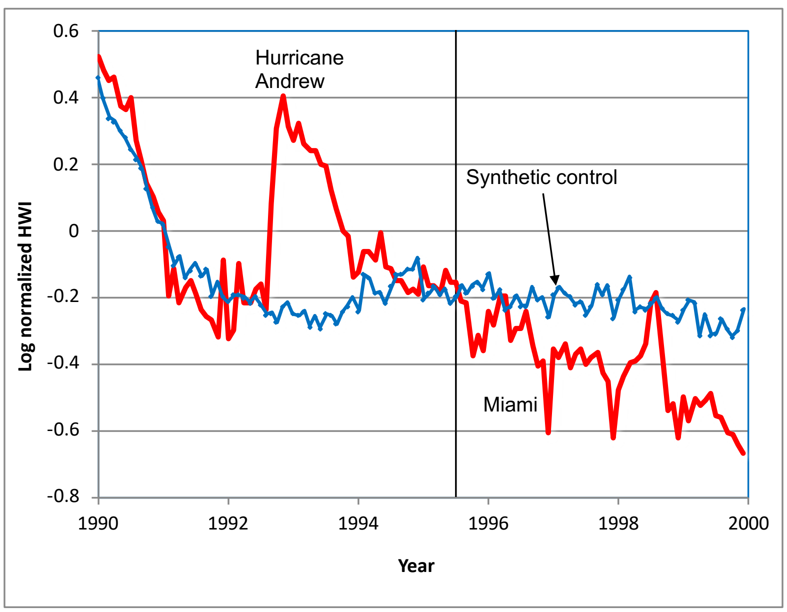
Notes: The HWI for each city/region is rescaled to equal 1 in January 1991-August 1992. The treatment line is drawn as of June 1995.

Figure 10. Job vacancies in Miami relative to the synthetic control, 1990-1999

A. The Help-Wanted Index



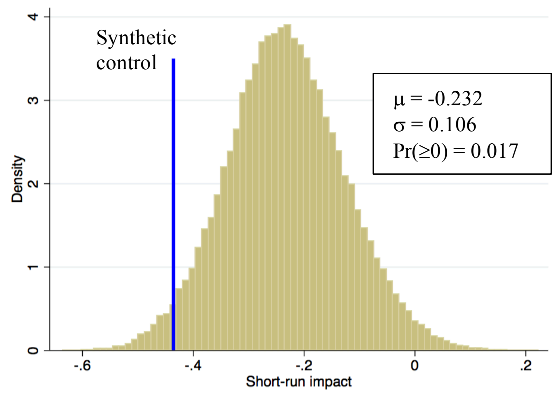
B. The normalized HWI



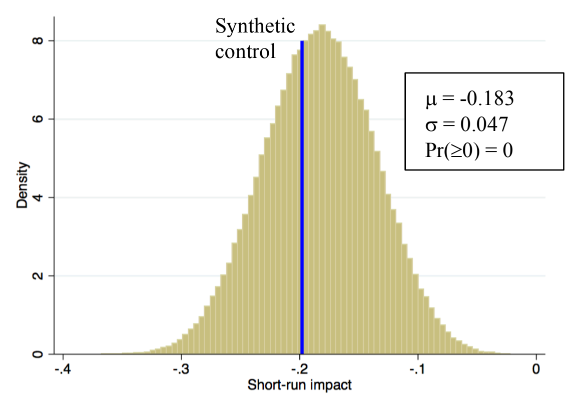
Notes: The normalized HWI is defined as the HWI divided by total non-agricultural employment in the city-year-month cell. The HWI and the normalized HWI for each city are rescaled to equal 1 in January 1991-August 1992. The treatment line is drawn as of June 1995.

Figure 11. Frequency distribution of short-run impacts in Miami (relative to all potential four-city control groups)

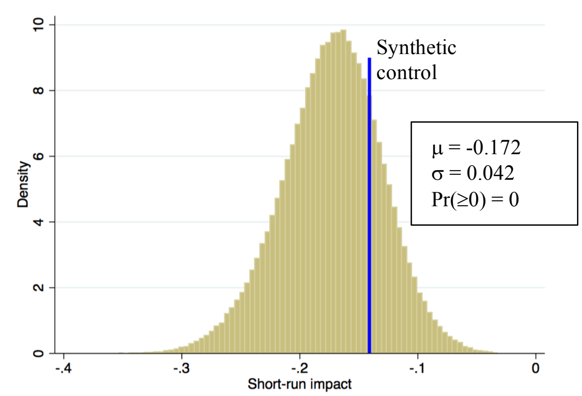
A. The Mariel supply shock



B. The first wave of Cuban refugees



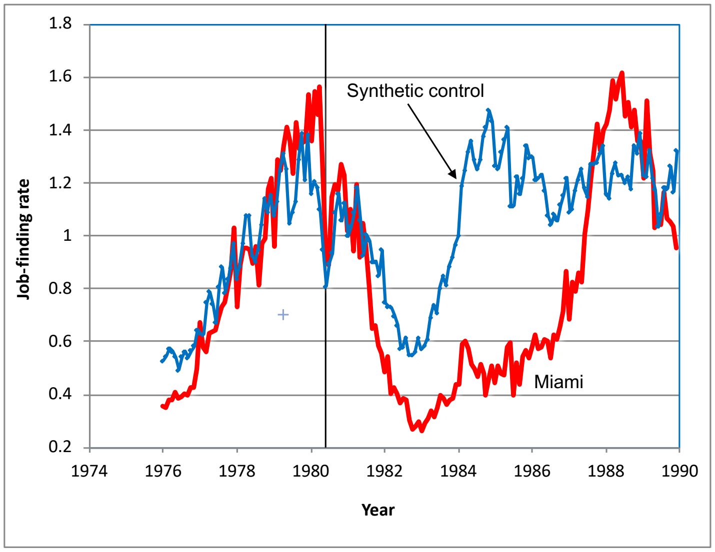
C. The "Mariel Boatlift That Never Happened"



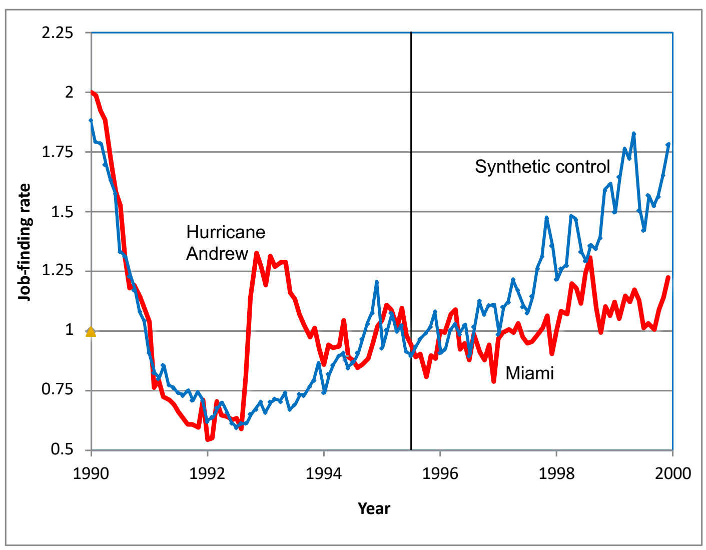
Notes: The sample periods used in the analysis are: 1979-1984 in Panel A, 1958-1963 in Panel B, and 1994-1997 in Panel C. The treatment dates are: July 1994 in Panel A, January 1960 in Panel B, and June 1995 in Panel C. Each panel illustrates the density function for the relevant coefficient from the difference-in-differences (log) HWI regression that compares Miami to all potential 230,300 four-city control groups. All regressions include year, month, and city fixed effects.

Figure 12. The job-finding rate in Miami relative to the synthetic control

A. The Mariel supply shock



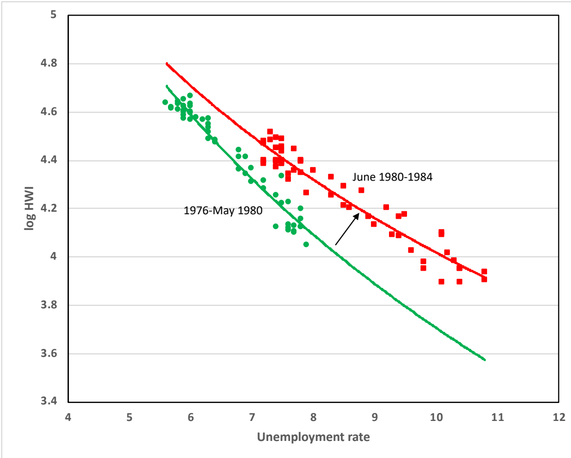
B. The 1994-1995 supply shock



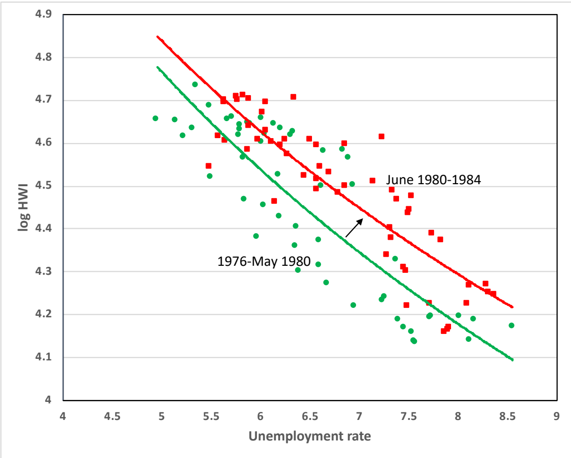
Notes: The job-finding rate is defined as the ratio of the HWI to the unemployment rate in the city-year-month cell. The job-finding rate for each city is rescaled to equal 1 in 1977-1979 (Panel A) or in January 1991-August 1992 (Panel B). The treatment line is drawn as of either June 1980 (Panel A) or June 1995 (Panel B).

Figure 13. The Beveridge Curve and Mariel

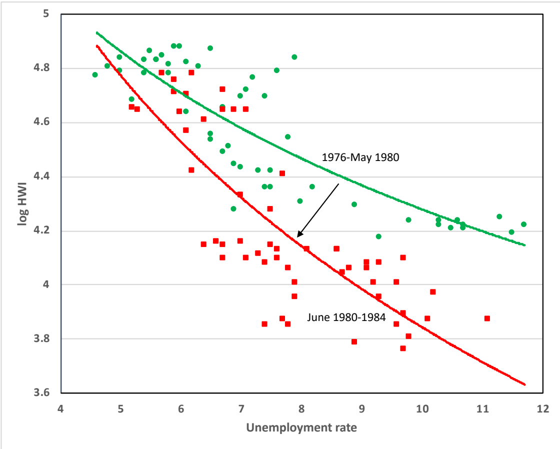
A. The national labor market



B. The synthetic control



C. Miami

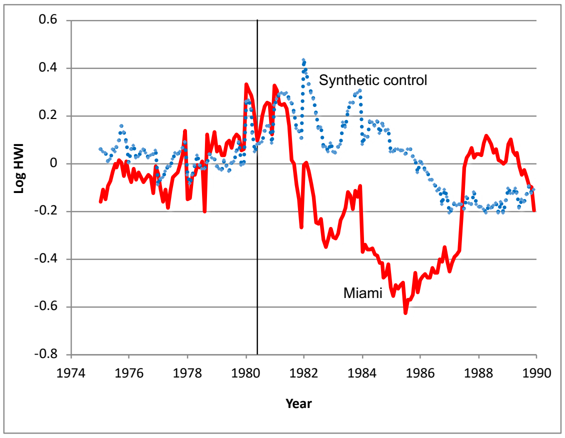


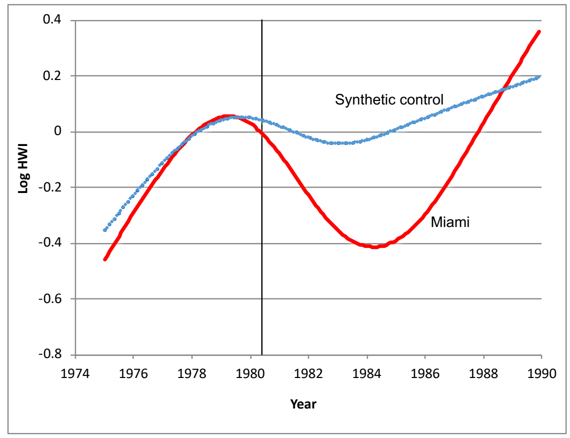
Notes: The figures show the data scatter and the logarithmic trend lines relating the raw HWI and the unemployment rate in a city-year-month cell in the pre-Mariel (January 1976-May 1980) and post-Mariel (June 1980-December 1984) periods. The Beveridge curve for the synthetic control in Panel B is constructed by taking a weighted average of the vacancy and the unemployment rate across metropolitan areas for each year-month cell, weighted by the synthetic control weights obtained in the analysis of the job-finding rate.

Figure 14. Long-term trends in the HWI

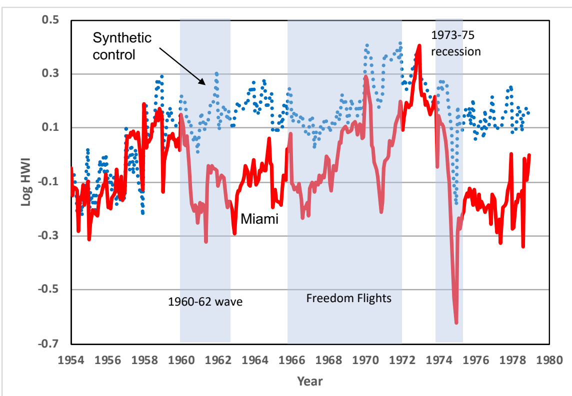
Using year-month fixed effects | **Using HP filter**

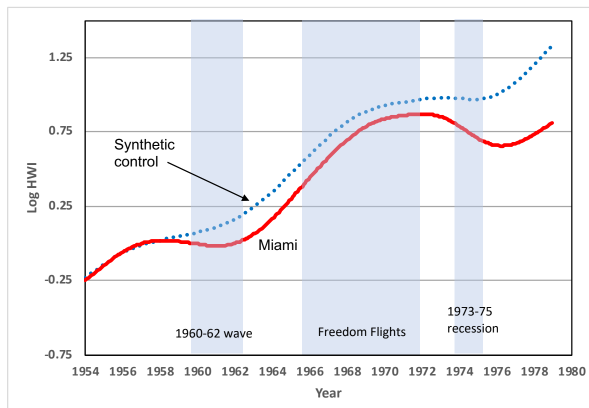
A. The Mariel supply shock



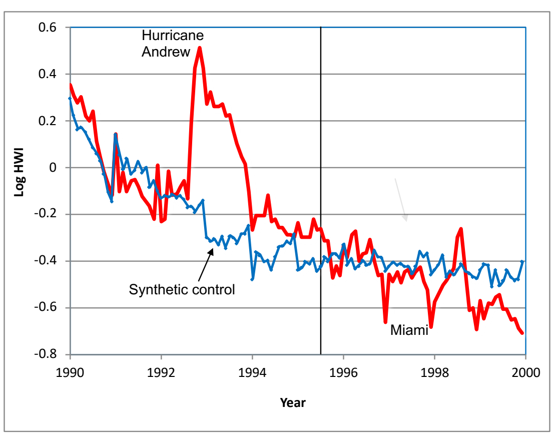


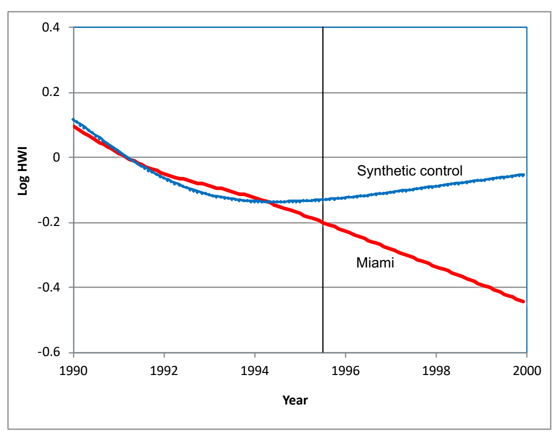
B. The first wave of Cuban refugees





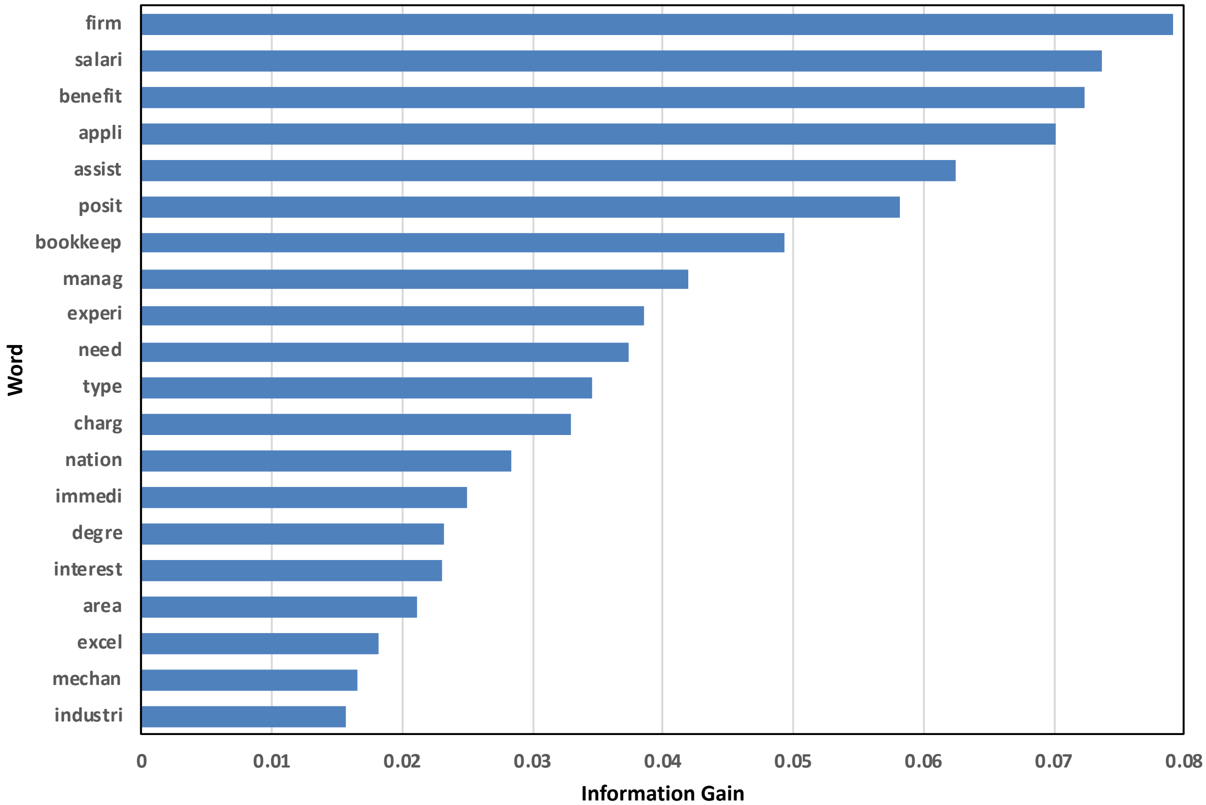
C. The 1994-1995 supply shock





Notes: The HWI in the left column is the residual from a regression that stacks all observations across metropolitan areas and time periods and the regressor is a vector of interacted year-month fixed effects. The HWI in the right column is the long-term trend implied by the HP filter with a smoothing parameter of 129,600. The HWI for each city is rescaled to equal 1 in 1956-1958 (Panel A), 1977-1979 (Panel B), or January 1991-August 1992 (Panel C).

Figure 15. Words that Best Distinguish High-Skill from Low-Skill Ads



Note: Importance is measured as information gain across trees.

Table 1. Supply shocks of Cuban immigrants in Miami labor market, 1960-2000

Episode:	High school dropouts	High school graduates	Some college	College graduates	All workers
A. Mariel, 1980					
No. of workers in Miami, 1980	176.3	187.5	171.5	124.1	659.4
No. of Cuban immigrants	32.5	10.1	8.8	4.2	55.7
Percent increase in supply	18.4	5.4	5.1	3.4	8.4
Education in occupation employing:					
Average native	11.8	12.5	13.2	14.9	13.3
Average Cuban immigrant	11.5	11.9	12.4	13.5	11.9
B. First wave, 1960-62					
No. of workers in Miami, 1960	172.1	111.0	43.1	35.5	361.8
No. of Cuban immigrants	21.7	15.5	7.7	15.0	59.9
Percent increase in supply	12.6	14.0	17.9	42.2	16.6
Education in occupation employing:					
Average native	10.2	11.6	12.2	14.6	11.7
Average Cuban immigrant	10.0	10.7	11.8	13.2	11.2
C. Guantanamo, 1994-1995					
No. of workers in Miami, 1990	246.9	222.9	260.8	193.7	932.3
No. of Cuban immigrant	13.5	11.5	4.5	6.8	36.2
Percent increase in supply	5.5	5.2	1.7	3.5	3.9
Education in occupation employing:					
Average native	12.1	12.8	13.4	15.2	13.6
Average Cuban immigrant	11.8	12.2	12.5	13.5	12.3

Notes: The pre-existing number of workers in Panel A is obtained from the 1980 census; the number of Cuban immigrants (at least 18 years old as of 1980) is obtained from the 1990 census; and a small adjustment is made because the 1990 census only identifies immigrants who arrived in 1980 or 1981. The pre-existing number of workers in Miami reported in Panel B is calculated from the 1960 census; the number of Cuban immigrants (at least 18 years old as of 1962) comes from the 1970 census; and a small adjustment is made because the 1970 census only identifies immigrants who arrived between 1960 and 1964. The pre-existing number of workers in Miami reported in Panel C is obtained from the 1990 census; the number of Cuban immigrants (at least 18 years old as of 1995) is obtained from the 2000 census.

.

Table 2. Difference-in-differences impact of 1980 supply shock

Variable:	Card control	Low-skill control	Synthetic control	All cities
A. Log HWI				
June 1980-1982	-0.204	0.006	-0.237	0.029
	(0.045)	(0.048)	(0.037)	(0.044)
1983-1984	-0.482	-0.186	-0.436	-0.296
	(0.030)	(0.027)	(0.023)	(0.028)
1985-1986	-0.446	-0.225	-0.448	-0.493
	(0.038)	(0.024)	(0.030)	(0.019)
1987-1989	0.109	0.129	0.138	-0.041
	(0.034)	(0.034)	(0.030)	(0.030)
B. Log normalized HWI				
June 1980-1982	-0.179	-0.059	-0.258	-0.039
	(0.039)	(0.046)	(0.035)	(0.042)
1983-1984	-0.449	-0.298	-0.410	-0.364
	(0.027)	(0.025)	(0.021)	(0.027)
1985-1986	-0.403	-0.358	-0.395	-0.544
	(0.032)	(0.024)	(0.027)	(0.018)
1987-1989	0.130	-0.017	0.138	-0.099
	(0.031)	(0.035)	(0.028)	(0.030)

Notes: Robust standard errors are reported in parentheses. The data consist of monthly observations for each city between 1975 and 1989. All regressions include vectors of city, year, and month fixed effects. The table reports the interaction coefficients between a dummy variable indicating if the metropolitan area is Miami and the timing of the post-Mariel period (the baseline period goes from January 1975 through May 1980). The regressions that use the Card control has 720 observations; the regressions using the low-skill control has 900 observations; the regressions using "all cities" have 9,180 observations; and the regressions that use the synthetic placebo have 360 observations.

Table 3. Difference-in-differences impact of 1960-62 supply shock

	Log HWI			Log normalized HWI		
Variable:	(1)	(2)	(3)	(1)	(2)	(3)
1960-1961	-0.210	-0.210	-0.161	-0.239	-0.239	-0.203
	(0.022)	(0.022)	(0.020)	(0.023)	(0.023)	(0.022)
1962-1963	-0.269	-0.269	-0.245	-0.278	-0.278	-0.260
	(0.015)	(0.015)	(0.016)	(0.017)	(0.017)	(0.019)
1964-1965	-0.218	-0.218	-0.232	-0.134	-0.134	-0.144
	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
1966-1968	---	-0.174	-0.242	---	-0.155	-0.205
		(0.015)	(0.018)		(0.014)	(0.018)
1969-1971	---	-0.181	-0.196	---	-0.181	-0.192
		(0.023)	(0.017)		(0.021)	(0.018)
1972-1974	---	-0.098	-0.086	---	-0.186	-0.176
		(0.023)	(0.021)		(0.033)	(0.031)
1975-1978	---	-0.287	-0.178	---	-0.502	-0.422
		(0.014)	(0.023)		(0.016)	(0.027)
Unemployment rate	---	---	-0.049	---	---	-0.036
			(0.007)			(0.008)

Notes: Robust standard errors are reported in parentheses. The data consist of monthly observations for Miami and the synthetic control between 1954 and 1965 (in column 1) or 1954 and 1978 (in columns 2 and 3). All regressions include vectors of city and year fixed effects. The table reports the interaction coefficients between a dummy variable indicating if the metropolitan area is Miami and the timing of the post-Mariel period (the excluded period goes from January 1954 through December 1959). The "unemployment rate" variable is an interaction between the Miami dummy variable and the national unemployment rate in the year-month cell. The regressions in columns 2 and 3 have 600 observations.

Table 4. Difference-in-differences impact of 1994-95 supply shock

Dependent variable	Synthetic control	All cities	All cities	All cities
A. Log HWI				
January 1990-August 1992	-0.044	0.256	0.217	0.254
	(0.040)	(0.032)	(0.034)	(0.032)
September 1992-December 1993	0.366	0.477	0.452	0.471
	(0.039)	(0.039)	(0.038)	(0.038)
June 1995-December 1997	-0.141	-0.175	-0.148	-0.161
	(0.022)	(0.021)	(0.022)	(0.021)
January 1998-December 1999	-0.215	-0.314	-0.325	-0.309
	(0.025)	(0.026)	(0.028)	(0.026)
B. Log normalized HWI				
January 1990-August 1992	0.054	0.286	0.257	0.285
	(0.039)	(0.030)	(0.032)	(0.031)
September 1992-December 1993	0.396	0.491	0.470	0.486
	(0.041)	(0.041)	(0.040)	(0.040)
June 1995-December 1997	-0.162	-0.190	-0.170	-0.179
	(0.022)	(0.023)	(0.023)	(0.022)
January 1998-December 1999	-0.242	-0.339	-0.368	-0.344
	(0.025)	(0.026)	(0.029)	(0.027)
Interacts year fixed effects and percent of adults in city:				
Who own a personal computer	No	No	Yes	No
Who use the Internet	No	No	No	Yes

Notes: Robust standard errors are reported in parentheses. The vacancy rate proxy is defined as the HWI divided by total employment in the city-year-month cell. The data consist of monthly observations for each city between 1990 and 1999. All regressions include vectors of city and year fixed effects. The table reports the interaction coefficients between a dummy variable indicating if the metropolitan area is Miami and the timing of the post-Mariel period (the excluded period goes from January 1994 through May 1995). The regressions that use the synthetic control have 240 observations, and the regressions using "all cities" have 5,760 observations.

Table 5. Labor market conditions and the HWI

		Log HWI interacted with:		
	Log HWI	High school dropout	High school graduate	Some college
March CPS (1972-1999)				
1. Log weekly wage	-0.009	0.125	0.069	0.038
	(0.013)	(0.021)	(0.013)	(0.013)
2. Employment propensity	0.002	0.043	0.034	0.023
	(0.004)	(0.006)	(0.004)	(0.004)
CPS-ORG (1979-1999)				
1. Log hourly wage	-0.030	0.107	0.063	0.043
	(0.012)	(0.021)	(0.012)	(0.012)
2. Employment propensity	0.022	0.036	0.032	0.025
	(0.004)	(0.006)	(0.004)	(0.005)

Notes: Robust standard errors in parentheses. The number of observations in the wage regressions are 4,604 in the March CPS and 3,626 in the CPS-ORG; the number in the employment regressions are 4,604 in the March CPS and 3,626 in the CPS-ORG. Both the wage and employment variables are age- and gender-adjusted. The employment variable in the March CPS gives the probability that the person worked at some point during the calendar year prior to the survey, while the employment variable in the CPS-ORG gives the probability that the person worked during the CPS reference week. All regressions are weighted by the number of observations used to calculate the dependent variable.

Table 6. Performance statistics for the classifier and prediction results

	Test data	1978	1982
A. Performance statistics			
Accuracy (%)	72.0		
Sensitivity rate (%)	80.4		
Specificity rate (%)	54.0		
B. Prediction exercise			
Total number of ads		1054	400
Number of high-skill ads		277	115
Number of low-skill ads		777	285
Percent of low-skill ads		73.7	71.3

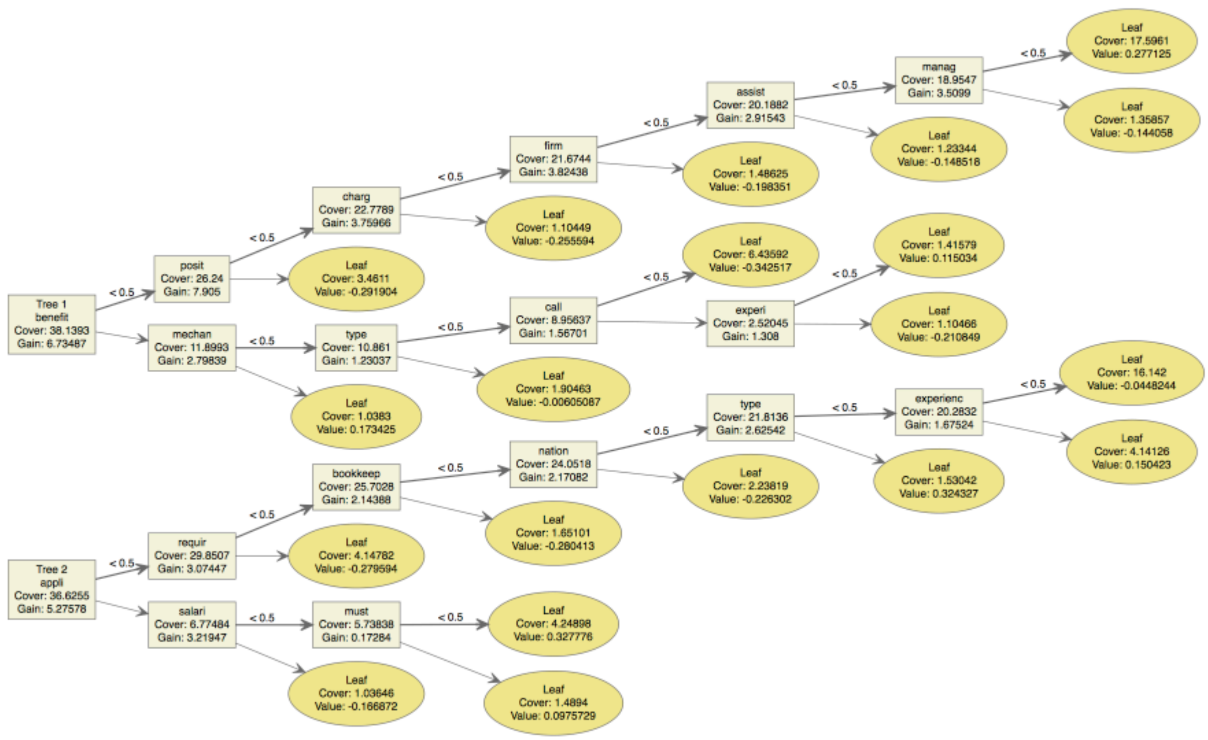
Notes: The accuracy rate is the percent of ads assigned to the correct skill group by the classifier; the sensitivity rate is the true positive rate, or the percent of low-skill ads classified correctly; the specificity rate is the true negative rate, or the percent of high-skill ads classified correctly.

Table 7. Supply shocks and the HWI, 1960-2000
(Dependent variable = Decadal change in city's log HWI)

		Education specific shocks			
Definition of supply shock:	All immigrants	High school dropouts	High school graduates	Some college	College
A. $M(\tau, \tau\text{-}1)/N(\tau\text{-}1)$					
1. All immigrants	-1.027	---	---	---	---
	(0.354)				
2. Education-specific shocks introduced separately	---	-0.371	-1.335	-1.697	-1.282
		(0.151)	(0.375)	(0.553)	(0.500)
3. Education-specific shocks introduced at same time	---	-0.173	-1.444	-2.110	2.619
		(0.292)	(2.005)	(3.246)	(1.985)
4. All immigrants (IV)	-1.809	---	---	---	---
	(0.606)				
B. $M(\tau, \tau\text{-}1)/N(\tau)$					
1. All immigrants	-1.291	---	---	---	---
	(0.402)				
2. Education-specific shocks introduced separately		-0.262	-1.348	-2.355	-2.166
		(0.111)	(0.335)	(0.642)	(0.608)
3. Education-specific shocks introduced at same time	---	0.022	-1.452	-1.435	1.439
		(0.171)	(2.345)	(3.980)	(2.992)
4. All immigrants (IV)	-1.707	---	---	---	---
	(0.553)				
C. $M(\tau)/N(\tau) - M(\tau\text{-}1)/N(\tau\text{-}1)$					
1. All immigrants	-1.104	---	---	---	---
	(0.471)				
2. Education-specific shocks introduced separately	---	-0.153	-1.149	-1.831	-2.045
		(0.092)	(0.307)	(0.674)	(0.734)
3. Education-specific shocks introduced at same time	---	0.393	-5.934	2.859	1.473
		(0.170)	(2.087)	(2.557)	(2.126)
4. All immigrants (IV)	-1.835	---	---	---	---
	(0.703)				

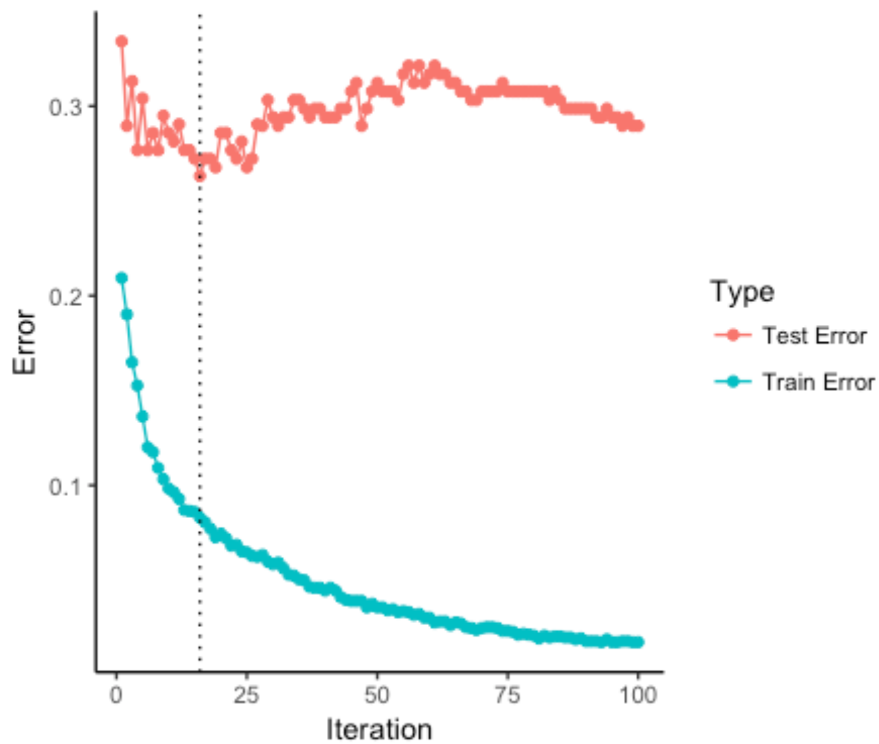
Notes: Robust standard errors are reported in parentheses. The variable $M(\tau)$ and $N(\tau)$ give the number of immigrants and natives (in the relevant city-year-education cell) enumerated in census year τ, and $M(\tau\ \text{-}1, \tau)$ gives the number of immigrants who arrived between the two census years. The regressions in Panels A and B have 198 observations and the regressions in Panel C have 148 observations. The instrument is the predicted size of the immigrant flow settling in a particular city based on the geographic settlement of earlier waves of immigrants belonging to the same national origin group (as constructed by Jaeger, Ruist, and Stuhler, 2018). All regressions are weighted by the size of the city's adult population at the time of the census.

Appendix Figure A1. Two trees grown as part of the final trained model



Notes: The importance of words, as measured by information gain, for distinguishing between low- and high-skill classified ads are displayed from left (most important) to right (less important).

Appendix Figure A2. Cross-validated average training and test error for each training iteration



Notes: The final model chosen was the one with the lowest average test error as indicated by the dotted line.

Appendix Table A1. Newspapers sampled by the Conference Board

City	Paper Used for HWI Since at Least 1970
Albany	*The Times Union*
Allentown	*Allentown Morning Call*
Atlanta	*Atlanta Constitution* (became *Atlanta Journal Constitution* in 1982)
Baltimore	*Baltimore Sun*
Birmingham	*Birmingham News*
Boston	*Boston Globe*
Charlotte	*Charlotte Observer*
Chicago	*Chicago Tribune*
Cincinnati	*Cincinnati Enquirer*
Cleveland	*Cleveland Plain Dealer*
Columbus	*Columbus Dispatch*
Dallas	*Dallas Times Herald* until 1991, then *Dallas Morning News*
Dayton	*Dayton Daily News*
Denver	*Denver Rocky Mountain News*
Detroit	*The Detroit News*
Gary	*Gary Post-Tribune*
Hartford	*Hartford Courant*
Houston	*Houston Chronicle*
Indianapolis	*Indianapolis Star*
Jacksonville	*Florida Times-Union*
Kansas City	*Kansas City Star*
Knoxville	*Knoxville News-Sentinel*
Los Angeles	*Los Angeles Times*
Louisville	*Louisville Courier-Journal*
Memphis	*Memphis Commercial Appeal*
Miami	*Miami Herald*
Milwaukee	*Milwaukee Sentinel*
Minneapolis	*Minneapolis Star Tribune*
Nashville	*Nashville Tennessean*
New Orleans	*The Times-Picayune*
New York	*New York Times*
Newark	*Newark Evening News*
Oklahoma City	*The Daily Oklahoman**
Omaha	*Omaha World-Herald*
Philadelphia	*Philadelphia Inquirer*
Phoenix	*Phoenix Arizona Republic*
Pittsburgh	*Pittsburgh Post-Gazette*
Providence	*Providence Journal*
Richmond	*Richmond Times-Dispatch*
Rochester	*Rochester Times-Union*
Sacramento	*Sacramento Bee*
Salt Lake City	*Salt Lake Tribune*
San Antonio	*San Antonio Express-News*
San Bernardino	*San Bernardino Sun*
San Diego	*San Diego Union*
San Francisco	*San Francisco Examiner*
Seattle	*Seattle Post-Intelligencer*
St. Louis	*St. Louis Post-Dispatch*
Syracuse	*Syracuse Herald Journal*
Toledo	*Toledo Blade*
Tulsa	*Tulsa World*
Washington D.C.	*Washington Post*

*We have been unable to confirm that the surveyed paper in Oklahoma City was the *Daily Oklahoman*.

Appendix Table A2. Confusion matrix for predictions on the test data

	Reference	
Prediction	High-Skill	Low-Skill
High-Skill	13	10
Low-Skill	11	41

Notes: The confusion matrix provides information about where the human coders and the machine learning algorithm agreed and disagreed about the classification of low- and high-skill ads in the test data of 75 ads. The classifier and the coders agreed on 54 ads, giving an accuracy rate of 54/75 = 0.72. The classifier incorrectly assigned 10 low-skill ads, giving a sensitivity rate of 80.4% (or 41/(41 + 10)). The classifier incorrectly assigned 11 high-skills ads, giving a specificity rate of 54 percent (or 13/(13+11)).