The Impact of Immigration on the Labor Market Outcomes of Native Workers: Evidence using Longitudinal Data from the LEHD

by

Ted Mouw
University of North Carolina-Chapel Hill

CES 16-56 December, 2016

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see www.census.gov/ces or contact J. David Brown, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies 5K034A, 4600 Silver Hill Road, Washington, DC 20233, CES.Papers.List@census.gov. To subscribe to the series, please click here.
Abstract

Empirical estimates of the effect of immigration on native workers that rely on spatial comparisons have generally found small effects, but have been subject to the criticism that out-migration by native workers damps the observed effect by spreading it over a larger area. In contrast, studies that rely on variation in immigration across industries, occupations, or education-based skill-levels often report large negative effects, but rely primarily on repeated cross-sectional data sets which also cannot account for the adjustment of native workers over time. In this paper, we use a newly available data set, the Longitudinal Employer Household Data (LEHD), which provides quarterly earnings records, geographic location, and firm and industry identifiers for 97% of all privately employed workers in 29 states. We use this data to analyze the impact of immigration on earnings changes and the mobility response of native workers. Overall, we find that although immigration has a negative effect on the earnings and employment of native workers, and positive effects on their firm, industry, and cross-state mobility, the overall size of the effects is small.

* Note: This research was supported by a grant to the author from the National Science Foundation (NSF SES-0827347), as well as additional funding from the University of North Carolina which provided continued access to the Census Research Data Center (RDC) and through an NSF grant (ITR-0427889) that supports research at the Triangle RDC. In addition, the author gratefully acknowledges support and assistance from the Census Bureau’s Center for Economic Studies and from William B. Grider, the administrator of the Triangle RDC. Contact information: Ted Mouw, Department of Sociology, University of North Carolina, Chapel Hill, CB#3210, 155 Hamilton Hall, Chapel Hill, NC 27599-3210. tedmouw@email.unc.edu. DISCLAIMER: Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.
INTRODUCTION

As of 2014, 13.3% of the U.S. total population was foreign-born, the highest number since 1920, and a dramatic increase from a low of 5% as recently as 1970 (Colby and Ortman 2015, Singer 2012). Perhaps not surprisingly, this rapid increase has sparked public debate about the impact of immigration and whether it should be reduced or halted (Renwick and Lee 2015). Overall, research on public opinion data suggests that attitudes towards immigration often revolve around concerns about the economic consequences (Fussell 2014; Ceobanu and Escandell 2010; Wallace and Figueroa 2012), and in a 2015 poll, 50% of respondents thought that immigration had a negative effect on the economy, compared to 28% who thought the effect was positive (Pew Research Center 2015).

While there may be a widespread perception among some members of the public that immigration has negative economic consequences, determining whether this is really the case is complicated by the profusion of competing results in academic studies purporting to answer the question (Longhi, Nijkamp, and Poot 2006). In particular, a striking feature of this literature is the discrepancy between the generally very small or null effects in studies based on variation in immigration levels across geographic areas (see Blau and Kahn 2012 for a review) and the often large negative effects in studies that focus on occupational or industrial variation (Kim and Sakamoto 2013; Camarota 1998; Howell and Mueller 2000; Catanzarite 2003), “skill cells” formed by education and experience groups in national labor markets (Borjas 2003), or the effects on African American workers (Adelman, Lippard, Jaret, and Reid 2005; Kposowa 1995; Johannson and Shulman 2003; Borjas, Grogger, and Hanson 2010). In addition, numerous case studies cite evidence of the apparent displacement of native workers by immigrants in low-skill or labor intensive industries (e.g., Waldinger and Lichter 2003; Steusse 2016; Schwartzman 2007), often referring to reports of employer preferences for immigrants as easily exploitable or “willing subordinates” (Alexander 2012; Donato and Bankston 2008) or expressed using code-words about motivation and work effort (Harrison and Lloyd 2013; Zamudio and Lichter 2008; Waldinger 1997; Saucedo 2006; Maldonado 2009).
While some authors have referred to the discrepancy in results between these different approaches as a “paradox” (Rosenfeld and Tienda 1999:64), an alternative possibility is that it indicates the importance of identifying which native workers are actually competing with immigrants as well as understanding the ways that workers adapt to the arrival of immigrants over time. First, because immigrants are often highly clustered in specific industries and occupations and are overrepresented at both ends of the educational distribution, many natives may actually be complements, rather than substitutes, with immigrants in the production process (Linton 2002; Sanders 2012). Indeed, the challenge of pinpointing the degree to which native workers are competing with immigrants even within detailed educational categories has been the subject of heated debate in the recent literature (Ottaviano and Peri 2012; Borjas, Grogger, and Hanson 2011). As a result, it is possible that studies that focus on variation across industries and occupations do a better job of isolating the areas where native workers and immigrants are in direct competition (Kim and Sakamoto 2013; Steinhardt 2009).

At the same time, however, it is likely that native workers and firms adjust to immigration over time in ways that modify its initial impact. Native workers who are competing with immigrants may respond by moving to a different geographic region, which would spread any potential negative effect out over space (Borjas 2006). Alternatively, natives might react by changing industries or occupations in ways that help them avoid direct competition with immigrants (Peri and Sparber 2009; Pais 2013), which might result in net benefits over time (Cattaneo, Fiorio, and Peri 2013). Similarly, industries that employ immigrant workers may respond by increasing production and employment (Olney 2013; Lee 2007) or by adopting labor-intensive technologies (Lewis 2003). Finally, an additional complication for researchers is that immigrants themselves may move to regions with high labor demand (and rising wages), thereby biasing results in studies that rely upon variation in immigrant flows to estimate the effect on native workers (Cadena and Kovak 2013).

In addition to the importance of having an appropriate theoretical model, a considerable part of the challenge of reconciling the findings in the existing literature is the lack of adequate data to identify competing workers and model the dynamic process of adjustment. In studies that use repeated cross-
sections of decennial census data, for example, the 10-year gap in between observations may be too long to model the process of adjustment and any short-term negative consequences that exist may be hidden in the data. Similarly, even with data that appears at shorter intervals, such as the 1-year gap between successive waves of the Current Population Survey (CPS) or the American Community Survey (ACS), the inability to follow individual workers over time makes it difficult to model the ways native workers respond to immigrant inflows.

In this paper, we attempt to improve the estimates of the effect of immigration on native workers by using a unique large scale data set, the Longitudinal Employer Household Dynamics (LEHD) data, which enables us to follow workers over time and has detailed information on the geographic location, industry, and firm that they are employed in. In particular, there are two specific advantages to our data. First, because of the size of the data and the administrative records on industry classification and earnings, we are able to focus our analysis on specific groups of native workers who are likely to be competing with immigrants by calculating changes in immigration levels within earnings quintiles nested within detailed 4-digit industries and local labor markets. Second, our approach enables us to model the degree to which native workers shift their firm, industry, or geographic location away from areas experiencing high rates of immigrant inflows.

The LEHD is created by the Census Bureau by merging individual and firm-level administrative data on the quarterly earnings of workers from state unemployment insurance records (McKinney and Vilhuber 2011). As a result, the LEHD combines the benefits of the detailed geographic information and large sample size of the Decennial Census with the strengths of longitudinal data sets such as the PSID and the NLSY. The version of the LEHD that we use in our paper consists of 96.1 million privately employed workers in 29 states from 1992-2008. In our analysis, we look at the impact of immigration on workers in detailed 4-digit industries within 287 spatially-defined labor market areas. By tracking workers over time, we are able to include measures of the inflow of potentially displaced native workers, which helps us overcome one of the principal critiques of studies that use variation in local labor markets to study the impact of immigration.
As is described in detail below, an additional benefit of the approach that we take in this paper is the use of two instrumental variables to account for, and test, the potential endogenous movement of immigrants into regions and industries that are experiencing high labor demand, which could bias the results. First, we use a version of the ethnic enclave-based instrumental variable (“IV”) that is widespread in the literature (Card 2001), where a lagged measure of the geographic distribution of immigrants from a particular country is used to predict the geographic pattern of current immigrant flows to reduce the correlation with contemporaneous changes in labor demand. Nonetheless, we recognize that a potential problem with this IV is the possibility that persistent geographic patterns of labor demand are correlated with the previous location decisions of immigrants. As a result, we use a second instrumental variable based on industry-level demand shocks to test and calibrate the enclave-based measure of immigrant labor supply. This second IV is based on the change in employment in tradable industries at the national level (Bartik 1991), and it is frequently used in the literature on regional economic dynamics (Moretti 2010; Notowidigdo 2013).

Overall, in our models we find a uniformly consistent—and statistically significant—set of results indicating that immigration has a short term negative impact on the wages and employment of competing native workers, and that native workers adapt by moving geographically and changing industries and firms. However, the magnitude of the results is small even for competing workers in narrowly defined industries, suggesting that in a large, dynamic economy such as the United States, fears of a perceptible negative impact of moderate levels of immigration are overstated.

LITERATURE REVIEW

Theoretical Model

We begin with a brief theoretical discussion, highlighting key points that are relevant for our review of the empirical literature and for our subsequent analysis. Note that there are a number of extended reviews of the theoretical models that we draw on here which go over these points in more detail (Dustman, Glitz, and Fratini 2009; Blau and Kahn 2012; Lewis and Peri 2014; Gaston and Nelson 2007). First, to simplify
things, we imagine that there is an influx of low-skilled immigrants to a specific region (A) and industry (B) of the economy. The question is what impact does this have on native workers? We distinguish between short and long term effects, where “short term” means before native workers and firms have had a chance to adjust. As mentioned above, a key issue is distinguishing between native workers who are complements and those who are substitutes for the immigrant workers. Native workers who are substitutes are competing directly with immigrants, and, in the short term, the increased supply of competing workers should either decrease their wages or cause displacement. In contrast, native workers who are complements to the immigrant flow may experience increased demand for their services, resulting in higher wages and employment.

Theoretically, it also makes sense that we should be concerned about the potential endogeneity of the immigrant flows to a particular area. If immigrants choose to move to cities or regions with strong labor demand, then there may be upward pressure on wages in those areas that could obscure the otherwise negative short term effect of immigrants on competing native workers, resulting in positive bias on the regression coefficients for changes in immigrant density. However, the question of endogeneity could also work the other way if what is going on is a process of replacement where immigrants move into positions in declining industries or occupations vacated by upwardly mobile native workers.

In the medium or long term, workers can respond to the arrival of immigrants. If competing native workers move to a different geographic region in order to avoid competing directly with immigrants, this could diffuse the potential negative impact over a broader area (Borjas 2003). Alternatively, native workers might change their occupation or industry, possibly emphasizing tasks where they have a comparative advantage, such as communication (Peri and Sparber 2009), or by increasing their skill levels (Cattaneo, Fiori, and Peri 2013).

Firms and industries may also respond in a parallel fashion. If wages decline for low skilled workers in industry B of region A, then firms might take advantage of reduced labor costs to maximize

---

1 Note that we use “immigrant density” to refer to the proportion of the workforce of a particular sector or region that is immigrant.
profits by increasing production. Furthermore, if there is open trade between regions, then region A could absorb the added workers by shifting its mix of industrial employment to expand the production of tradable goods and export the increased production to other regions (Gaston and Nelson 2007). However, the empirical support for this mechanism of adjustment is surprisingly limited: Lewis (2003), Dustman and Glitz (2015), and Gonzalez and Ortega (2011) find little evidence that immigration induced changes in labor supply have a meaningful effect on the output mix across different regions. Lewis (2003) argues that this is consistent with firms modifying their production techniques to be more labor intensive, thereby absorbing the change in labor supply without increasing the overall level of output. However, the lack of concrete evidence on specific changes in production techniques means that this explanation, while plausible, relies more on indirect support based on the lack of compelling evidence for other competing explanations. Finally, looking beyond identifying the specific effect on the subset of native workers who are competing with immigrants, other authors have noted that immigration might be beneficial economically by stimulating demand for goods and services (Hong and Mclaren 2015, Bodvarsson, Lewer, and Berg 2007; Linton 2002), reducing consumer prices (Cortes 2008), or spurring innovation and productivity growth (Lewis and Peri 2014; Fassio, Montobbio, and Venurini 2015).

Overall, the key point is that there are theoretical reasons why we might expect a short-term negative effect of immigration on affected workers because of increases in labor supply. At the same time, however, there are good reasons to be cautious about interpreting the raw correlations between changes in immigrant density and natives’ labor market outcomes if immigrants’ locational decisions are endogenous to unobserved aspects of labor demand. Moreover, over time, as native workers and firms respond to the short-term impact of immigration, it is possible that the magnitude of the effect will be smaller, and some models conclude that the overall effect for native workers is positive (Ottaviano and Peri 2012). In the next section, we discuss the existing evidence on the various parts of this theoretical narrative, emphasizing the key methodological issues and the primary points of contention.
Empirical evidence

The spatial comparisons approach

The most commonly used approach to estimate the effect of immigration on wages and employment is to use naturally occurring geographic variation in immigrant settlement patterns. The so-called “spatial comparisons approach” analyzes the effect of changes in the percentage of immigrants working in education, occupation, or industry groups at the local geographic level on changes in the wages and employment of native workers in those groups. Overall, a striking result in this literature is that the majority of studies using this framework find little evidence of negative effects of immigration on the wages on native workers, with most of the results centered around zero (see reviews by Longhi, Nijkamp, and Poot 2006, Dustman, Glitz, and Frattini 2008, and Blau and Kahn 2012).

Nonetheless, there are two basic critiques of the spatial comparisons approach that argue that the estimates are substantially underestimated because of the potential non-random movement of immigrants into, and native workers out of, areas experiencing high rates of immigration. The first of these critiques is the worry about the potential endogeneity of immigrant flows to labor demand mentioned above: if immigrants move to areas that are booming economically, then this will bias estimates of the impact on native workers. In response to this critique, most recent studies that use the spatial comparisons approach use some sort of instrumental variable, searching for a factor that is correlated with the flow of immigrants to an area but not correlated with demand shocks. The most commonly used instrumental variable (“IV”) is based on the lagged locational distribution of previous immigrants under the argument that immigrants tend to form “enclaves” (Bartel 1989) with other immigrants of the same national origin group. This clustering effect could be based on the role of social networks in providing information and assistance in the migration process and the subsequent search for housing and employment (e.g., Munshi 2003).

The enclave-based IV is constructed as follows: let $p_{jk,t=1}$ indicate the proportion of national origin group j that lives in region k at time t=1. Let $\Delta N_j$ represent the overall change in the number of
immigrants of group $j$ at some point sufficiently far in the future, say between times 2 and 3. Then the enclave-IV for the flow of immigrants to region $k$ between times 2 and 3 is the national level change in the number of immigrants from each group multiplied by the lagged share in each region:

$$\Delta N_{k}^{IV} = \sum_{j} P_{jk,t=1} \times \Delta N_{j}$$

If labor demand shocks between times 2 and 3 are not correlated with the past locational decisions of immigrants at time 1, then this approach should result in a prediction of immigrant flows unaffected by contemporaneous changes in labor demand. Various forms of this IV have been widely adopted in the literature (Card 2001; Cortes 2008). Nonetheless, as noted by Longhi, Nijkamp, and Poot (2006) a potential weakness of this instrument is that if labor demand shocks persist over a substantial period of time then the lagged locational patterns might be correlated with current labor demand. In the absence of some direct measure of labor demand shocks there may be no way to intuit whether the assumptions underlying the use of this IV are true or not.

**Adjustment via geographic mobility**

The second often-mentioned critique of the spatial comparisons approach is that native workers may respond to the influx of immigrants by moving to a geographic area less affected by immigration (Borjas 2003). In this way, their geographic displacement is a form of spatial arbitrage which reduces the wage differences between areas and spreads the negative effect of immigration across the entire economy. If the period of time between observations in the data is large enough then the movement of native workers away from immigrants may reduce the wage differences across areas and hide any potential negative effects of immigration. Nonetheless, aside from Borjas (2006) and Cortes (2008), most of the current literature suggests little evidence of substantial state or regional-level displacement effects for
native workers due to immigration (see, for example, Wozniak and Murray 2012; Card 2001; Peri and Sparber 2011).2

The National Skill-Cell Approach

A concern about the potential geographic mobility of native workers has led to the development of what might be called the “national skill-cell” approach which analyzes the labor market effects of immigration at a national level. This approach assumes a constant elasticity of substitution (CES) production function for the overall output of the economy and a national labor market for workers within cells based on combinations of broad education and experience categories (Borjas 2003, Ottaviano and Peri 2012). An important feature of the use of the CES production function is the ability to simulate the overall effect of immigration which includes not only the direct effect of competing workers, but also the potential positive effects resulting from labor market complementarity among different types of workers. Borjas (2003) uses aggregate data from the 1960-2000 Censuses and regresses average native wages on immigrant density using fixed effects for all two-way combinations of time, education, and experience. He finds that a 10 percent increase in labor supply due to immigration reduces the weekly wages of competing native workers by 4 percent.3 However, in a reanalysis of these findings, Bohn and Sanders (2007) suggest that the results in Borjas (2003) are sensitive to slight changes in specification, as the main finding is not statistically significant if high school dropouts (or younger and older workers) are excluded from the analysis.

Ottaviano and Peri (2012) use the same basic approach as Borjas (2003), but they allow for imperfect substitution between natives and immigrants within the education-experience cells, alternative groupings based on education, and a readjustment in capital stock of the economy in response to the increased supply of labor. They conclude that the overall effect of immigration on the average native

---

2 See Peri and Sparber (2011) for a methodological critique of the two estimates (Borjas 2006 and Cortes 2008) that find a geographic displacement effect.

3 With the fixed effects, the results rely on variation within education-experience cells over time for identification.
worker is between -0.6 to +1.0 percent, depending on the exact specification of the model that they use.  

One of the points of contention is the degree to which capital flows are assumed to respond to immigration. In the CES production function, an increase in the supply of labor raises the return to capital, which should increase its supply. Because Borjas’ (2003) model doesn’t allow capital to readjust to the increased supply of labor, it should be interpreted as a short term effect of immigration. On the other hand, however, Ottaviano and Peri’s (2012) results assume that the capital/labor ratio will return to its pre-immigration level, which (although it may be a reasonable long-term assumption) mechanically drives the overall wage impact of immigration to zero.

Overall, two comments are relevant for summarizing the evidence from the national skill-cell approach. First, as noted in Borjas (2003:1362) and de la Rica, Glitz, and Ortega (2014), this approach assumes that immigration into the country is exogenous to changes in labor demand within skill cells, but this seems questionable given that economic models of the decision to migrate are based on the expectation of wage differences between destination and origin countries. Second, the CES production functions that are used by all of the authors adopting this approach make substantial simplifications in the number of cross-group parameters to estimate in order to get tractable results with a limited number of aggregate data points. In a review of the evidence, Blau and Kahn (2012) argue that the apparent sensitivity of the results to different specifications may be a fundamental shortcoming of this approach.

Adjustment via occupational and industrial mobility

The basic premise of the national skill-cell approach is that effect of immigration can be estimated within the skill cells because (unlike geographic mobility) native workers cannot change their education-experience categories (except as they age or go back to school). If, however, jobs (as defined by occupation-industry combinations) rather education-experience cells are true the inputs in the economy-wide production function, then the industrial and occupational mobility of native workers generates the

---

4 A comparison of Models A and B in Table 6 of Ottaviano and Peri (2012) suggests that a large part of the discrepancy between the two papers turns out to depend on whether or not the detailed educational categories are first nested within a high/low division of education based on a college degree.
same sort of potential problems for the national skill-cell approach that provoked criticism of the spatial comparisons approach. The question is whether native workers switch sectors away from immigrant inflows and what implications would this have on the process of adjustment.

Peri and Sparber (2009), for instance, argue that there is evidence that immigration pushes native workers into occupations that are more communication intensive, where they have a comparative advantage compared to immigrants. Gu and Sparber (2015) calculate that this shift to more communication intensive occupations reduces the impact of immigration on competing workers by 20%. Pais (2013) finds that the immigration level in the metropolitan area at the time of native workers’ entry into labor market entry has a positive impact on their subsequent wage growth. Similarly, Cattaneo, Fiorio and Peri 2013 use longitudinal data from 11 European countries and find that immigration density within 1-digit occupations increases the rate of occupational mobility, with a higher rate of upward rather than downward movement, and has no effect on the probability of movement into non-employment. They argue that it is possible that the influx of immigrants either causes or coincides with the expansion of opportunities for natives, which allows them to move up and vacates these positions for immigrants.

Estimates relying on industry and occupational variation

An alternative approach to the education-experience cells discussed above is to focus on variation in immigrant density across industry and occupational sectors as a way of more precisely identifying the groups of workers likely to be competing with immigrants. As emphasized above in the introduction, a striking finding here is that many authors report substantial negative effects of immigrant density at the industry or occupation level (Howell and Mueller 2000; Catanzarite 2003; Camarota 1998; Kim and Sakamoto 2013; Orrenius and Zavodny 2007; Card 2001; Steinhardt 2009). Kim and Sakamoto (2013), for instance, find that increasing the proportion immigrant in an occupation by 10 percentage points, reduces wages for native workers by 0.5 percent. Catanzarite (2003) shows that working in “brown collar” occupations with large numbers of recent Hispanic immigrants is negatively correlated with the earnings of native workers (with larger effects for black workers), and Howell and Mueller (2000) report
similar findings using cells composed of occupation and industry combinations and suggest that this reflects a crowding effect based on segregation.

Nonetheless, the possibility of adjustment by native workers through a process of occupational and industrial mobility described above could help us explain the difference between the findings between the spatial and sector approaches that was emphasized above in the introduction. The studies cited above in the sector approach either use repeated cross sections of quantitative data or are qualitative studies of specific occupations and industries, and as pointed out by Kim and Sakamoto (2013:97) and Waldinger and Licther (2003) this type of data doesn’t allow researchers to follow up on what happened to the native workers who used to work in the sectors that immigrants have moved into. This is where the results on occupational mobility as a response to immigration in Cattaneo, Fiorio, and Peri (2013) are provocative: if there is upward mobility away from sectors that immigrants are moving into then it calls into question the idea of fixed “skill cells” (education-experience or occupation/industry) to measure the consequences of immigration, and suggests that focusing on negative cross-sectional correlations between immigrant density and native wages within these cells may miss the point.

Quasi-experimental evidence

An alternative approach to estimate the labor market effect of immigration is to look large scale migration events that might approach a “quasi-experiment” of the impact of immigration. For example, in light of the concern about the endogeneity of immigrant flows, Card (1990) compares the labor market outcomes of native workers before and after the arrival of Cuban refugees from the 1980 Mariel boatlift, which increased Miami’s labor force by 7 percent. Comparing wage and employment trends in Miami versus a set of other comparable cities, Card (1990) finds little evidence that the trend in wages for native workers in Miami differed from the wage trend for native workers in the comparison cities. Nonetheless, Angrist and Krueger (1999:1328) argue that these results are sensitive to idiosyncrasies in labor market trends in Miami and in the choice of the comparison cities. As an example, they compare the impact of the 1980 boatlift to an averted boatlift from Cuba in 1994, when the U.S. Navy diverted the refugees to
Guantanamo Bay instead of allowing them to proceed to Florida. They find that using the same approach for this “non-event” resulted in an increase in unemployment for blacks in Miami between 1993 and 1995 compared to the same four comparison cities, which they argue might have been misinterpreted as a causal effect of immigration had this second Mariel boatlift actually happened.

A number of other papers have used immigration shocks caused by the sudden movement of refugees or natural disasters to attempt to estimate the effect of immigration (Friedberg 2001; Kugler and Yuksel 2008; Glitz 2012; de Silva 2009). Overall, while this approach is good at identifying push factors that result in a flow of migrants to the destination country, they still end up having to rely on an instrumental variable strategy to deal with potential endogeneity about which geographic areas within the country (e.g., Kugler and Yuksel 2008), or occupations (e.g., Friedberg 2001) the immigrants move into.

Recent analyses using longitudinal administrative data
The recent availability of large administrative data sets allows for a new type of analysis combining the spatial and sector comparisons approaches, by following individual workers over time to account for their mobility while still having sufficient numbers of cases to calculate the immigrant share of the workforce in local labor markets and sectors (Ortega and Verdugo 2013; Foged and Peri 2016; Lengermann, McKinney, and Pedace 2004). Foged and Peri (2016), for example, use register data on the universe of works from Denmark to estimate the effect of immigration in 98 municipalities and four broad industrial sectors from 1994-2008 following individual workers over time, using the lagged location of immigrants as an instrumental variable for current immigrant flows. They find that immigration increases the earnings and occupational complexity of native workers, which they argue constitutes evidence that native workers are complements, rather than substitutes, with immigrants in the labor market. However, because they measure immigrant density solely at the geographic level, it is possible that they are missing potential negative effects on competing native workers within the specific industries where immigrants are concentrated.
Lengermann, McKinney, and Pedace (2004) use data for seven states in the LEHD from 1995-1998 to analyze the impact of immigration on the earnings of native workers. They find that a 10 percentage point increase in the concentration of immigrants at the metropolitan level increases the earnings of male workers by 3.7 percent. However, as they note (on p.5) the positive association between immigration and native earnings may be due to unobserved labor demand factors correlated with both higher wages and larger immigrant inflows, and they do not attempt to account for the potential nonrandom movement of immigrants between geographic areas.

Overall, the recent availability of high quality administrative data allows for an improvement in the estimation of the effect of immigration by following individual workers over time. However, this research is still open to the criticism that the estimates do not target the impact on the subset of native workers most likely to be competing with immigrants or that the endogeneity of immigrant flows to local economic conditions may upwardly bias the resulting estimates.

DATA

In this paper, we use longitudinal data on 99.6 million workers for 1992-2008 from the Longitudinal Employer Household Data (LEHD) which is an administrative data set constructed by the Census Bureau that provides quarterly earnings data on 97% of all privately employed workers in participating states. The main source of administrative data for the LEHD comes from quarterly unemployment insurance (UI) reports by firms to state unemployment insurance commissions. Workers who are not covered by UI insurance are not in the data: these include workers in the military, federal government, agriculture (farms with less than 10 employees), and the self-employed (Stevens 2007). The firm-level data include the geographic location of the firm and detailed industry codes. For the analysis in this paper, we use the 4-digit NAICS (North American Industrial Classification System) codes, resulting in 311 different industries. Workers are identified over time by a randomly assigned 9-digit number—the “Protected Information Key” (PIK)—which represents a scrambled combination of their name and social security number (SSN). Matching workers’ UI earnings history to the firm-level data creates a linked employer-
employee data set. A full description of the steps involved in processing and creating the data is available in McKinney and Vilhuber (2011). As of the 2008 version of the data that we have access to (covering the time period from 1992 to 2008), there are 29 available states. The first year that each state's data is available varies from 1992 to 2003, and a majority of the states are available by 1998.5

Identifying immigrants in the data.

Information on workers' basic demographic characteristics in the LEHD comes from the Social Security Administration’s (SSA) “Numerical Identification” (Numident) file, which contains information on gender, race, age and place of birth. An important issue for this paper is that because the nativity variable comes from the SSA it is likely that the LEHD has incomplete coverage of undocumented immigrants (Andersson, Garcia-Perez, Haltiwanger, and Sanders 2014). As a result, it is important to consider the degree to which undocumented workers are missing in the LEHD, and what additional data can be used supplement it.

First, we note that there is a variable in the LEHD (the “invalid SSN flag”) that indicates whether a particular SSN has not been activated by the Social Security Administration. As argued by Lengerman et. al. (2004) and Brown, Hotchkiss, and Quispe-Agnoli (2012), these are likely to be undocumented workers, and this variable can be used as a proxy measure of overall flows of undocumented workers in the data. Alternatively, it is also possible that workers who cannot be matched to the Numident data with a PIK because of an incorrect name and SSN combination are deleted from the data during the processing of the UI records (McKinney and Vilhuber 2011). Below, we discuss remedies to potential undercounts of immigrants in the LEHD based on these two possibilities.

Lengerman et. al. (2004) provide a test of the accuracy of the LEHD immigration measures, using LEHD data on 7 states and comparing the count of immigrants in 2000 to the same states from the 2000 Census. They find that when they combine immigrants (with valid SSNs) with workers with the invalid

---

SSN flag, the number of immigrants in the LEHD data is proportionately 0.09 lower than the Census, but the overall difference in immigrant density is 0.005 (0.190 in the Census 2000 data compared to 0.185 in the LEHD).\(^6\) To test the accuracy of the immigration data in our 29-state sample of the LEHD data, we took a similar approach. We used data from the 5% sample of the 2000 Census to calculate the immigrant density of employed workers in each of the geographic areas in our sample and compared it to the immigrant density in our LEHD data for 2000, combining immigrants with valid SSNs with workers with invalid SSN flags. To make the Census data comparable to the LEHD, we only include workers in industries covered by unemployment insurance. Overall, the mean level of immigrant density in the LEHD data is slightly lower than the mean level from the Census data (but consistent with the results of Lengerman et. al. 2004), and the two measures are correlated at the 0.988 level across the 181 labor market areas in our data as of 2000. We argue that the high level of correlation between the LEHD and Census estimates means that the administrative data from the LEHD does a good job of capturing the spatial variation in immigrant density that is measured with household based data such as the Census microdata.

Nonetheless, although we have successfully benchmarked the LEHD immigration data to the 2000 Census, we want to be cautious about the accuracy of the LEHD coverage of immigrants. There are three potential approaches for adding supplemental data to augment the LEHD measures of immigrant density. These are (A) using the invalid SSN flag discussed above as a variable in the statistical models, (B) using state and industry level differences between the LEHD and the American Community Survey (ACS) as a proxy for the level of undocumented immigration, and (C) using systematic differences in employment counts between the LEHD and an alternative source of administrative data derived from the UI records (the Quarterly Census of Employment and Wages, “QCEW”) to measure the number of cases in the LEHD that have been deleted because they could not be matched to the SSA data. Although we discuss all three methods here, only method A is implemented in the current version of the paper.

\(^6\) Author’s calculations based on their results in Figure 1.
Method A for supplementing the LEHD immigration data involves using the invalid SSN variable as a proxy for undocumented immigrants. As mentioned above, the invalid SSN flag indicates social security numbers that have not been activated. Although this variable does not identify all undocumented immigrants in the data, it is likely to be correlated with the flow of undocumented workers across geographic areas and industries. As a result, including the change in the density of invalid SSNs along with our other migration variables as a proxy for flows of undocumented workers mitigates any potential problems caused by a divergence between flows of documented and undocumented immigrants.

An alternative method for supplementing the LEHD data involves using the 2000 Census microdata combined with the 2005-2008 ACS to test for changes in the accuracy of the LEHD over time. Although the ACS begins in 2000, sub-state geographic identifiers are only available starting in 2005, so we will interpolate changes between the 2000 Census and the 2005 ACS. If the density of migrants in a particular labor market area (LMA, see below) and industry is significantly larger (at the 0.05 level of significance) than the corresponding measure in the LEHD, then we will interpret this to indicate an undercount of undocumented workers. We will then add the estimated undercount to the analysis at the LMA level and the industry by LMA level.

Finally, Method C will test for whether undocumented immigrants have been deleted from the LEHD data because they didn’t match to the SSA Numident file based on their PIK. The Quarterly Census of Employment and Wages (QCEW) is a public source of aggregate employment counts derived from the same UI reporting process as the LEHD (but not processed by linking it to the SSA data), and we can match it to our LEHD sample at the county and industry level for each year of the data.\(^7\) If there is a systematic divergence between the LEHD and QCEW employment counts at the industry level within counties, then we will include this difference in counts as an alternative measure of potentially missing undocumented workers in our regression models.

---
\(^7\) See Spear (2011) for a comparison between the QCEW and the LEHD.
Additional Individual and industry characteristics

The LEHD includes data on workers’ gender, race, and date of birth. A disadvantage, however, is the lack of data on occupation and education, which are not reported in the unemployment insurance reports or in the SSA data. While we can use the detailed industry codes to find workers who are in the same industry as immigrants, we also want to identify workers who are most likely to be competing with those immigrants based on similar skill levels. In the main analysis for this paper, we use earnings quintiles within the industry as a proxy variable for workers’ skill levels. Earnings quintiles are defined separately for each quarter using inflation adjusted earnings for all full-quarter workers in the data. The assumption is that, within 4-digit industry categories, workers in the same wage quintile are likely to be similar in skill levels—and more likely to be competing with each other.

Full-quarter workers.

Because only the quarterly earnings of workers are reported by firms, not the number of hours or weeks worked, we need to differentiate between partial and full quarters of employment. If the worker only was at the firm for part of a quarter—i.e. in situations where the worker begins or ends a job—she would receive only a portion of her full earnings during that quarter. In order to prevent partial quarters of work from affecting the results, we identify “full quarter workers” as those workers at quarter T who were also employed at the firm at quarters T-1 and T+1: these are “interior” earnings quarters excluding the beginning and end of a worker’s tenure at the firm (Abowd and Vilhuber 2005).

Geographic data

The basic geographic identifier in the LEHD data is the county. We aggregate individual counties together into larger groups as a better measure of spatially defined local labor markets. First, we combine urban counties together at the Metropolitan Statistical Area (MSA) level. Rural counties are combined into Super Public Use Micro Areas (“Super PUMAs”), which are contiguous geographic areas consisting of at least 450,000 people. We refer to the combined MSA-Super PUMA geographic codes as “labor
market areas” (LMA). Overall, the 29 states in the 2008 version of the LEHD comprise 287 LMAs defined by MSAs and rural Super PUMAs.

To identify the LMA of workers, we use the variable in the LEHD for the county of work. For multi-unit firms, which represent about 40% of the firms in the data, the geographic location of each of the units is known, but the UI data only report the state level firm identification. For workers in these firms, the Census creates 10 imputations of the worker’s unit, using the location of the firm’s units, the county of residence, and detailed data on individual commuting behavior (McKinney and Vilhuber 2011). In order to accurately combine data on the geographic location of workers in single and multi-unit firms, we include each of the imputations of the county of work for multi-unit workers, and then aggregate these up to the to the LMA level, weighting each multi-unit worker based on the fraction of the imputations that are in each LMA. Because we are using labor market areas defined by aggregating multiple counties together, the geographic imputation of workers at multi-unit firms has much less impact on our data, as most imputations fall in the same LMA.

Measuring Inflows of Displaced Native Workers

As discussed above, if native workers move out of areas or sectors that are experiencing high immigrant inflows, then the negative effect of immigration may diffuse outward to other areas even without the direct movement of immigrants to those areas. For each earnings quintile within each industry, we identify native workers who worked in a different industry in the previous year. We divide these new entrants into two groups based on whether the immigrant density in their former industry-quintile combination was higher (“inflow1”) or lower (“inflow0”) than the immigrant density in their current one. Both of these variables are expressed as a proportion of the total employment in the industry and quintile within the particular LMA. If it is true that that the flight of native workers away from sectors with high concentrations of immigrants is exerting downward pressure on wages in the destination industry, then we would expect to that the coefficient on inflow1 would be smaller than the corresponding coefficient for
In addition, as discussed below, we also include an alternate measure of this native worker inflow rate derived from our instrumental variable approach for the flow of immigrant workers.

**METHODS: MODELING THE DYNAMIC RESPONSE TO IMMIGRATION**

**Dependent Variables**

In addition to changes in earnings and employment status, the dependent variables for this analysis represent measures of possible adjustment mechanisms discussed above in the literature review section. We focus on identifying the short term effect that might be missed by repeated cross-sectional data sets with longer intervals between observations by measuring the dependent variables as changes over a 1 year period. For earnings, we use the change in workers’ log earnings from time T to time T+4, where time is measured in quarters. To measure workers who drop out of the labor force, we create a dichotomous variable which is 1 if the worker is employed at time T, but is not observed to be working at time T+4. A worker who does not appear in the LEHD at time T+4 may have dropped out of the labor force, but he or she may also have migrated to a non-LEHD state or moved to a job that is not covered by the unemployment insurance reporting mechanism.

Geographic mobility is defined as a worker who switches states between time T and T+4. Note that because not all states are included in our 29-state sample, and because the number of states in the data increases over time, this measure of state mobility only picks up workers who are working in one of the covered states. However, as discussed below, because we use combined geographic and time fixed effects in our models, we will be comparing differences in the rate of geographic mobility across industries within each labor market area (and time period), so changes over time in the number of states that are included will not affect our results.

Industry and firm mobility indicate whether the worker switched industries or firms between time T and T+4. In order to differentiate firm mobility from geographic mobility, this variable is conditional upon working in the same state at time T+4. Similarly, industry mobility is conditional upon the worker
switching firms. Finally, we divide workers who switch industries into two groups based on whether their earnings increased or decreased between time T and T+4. Workers who shift industries and experience earnings declines are arguably a better measure of the degree of involuntary displacement due to immigration. In contrast, industry mobility accompanied by earnings gains indicates the type of voluntary mobility associated with skill upgrading—i.e. replacement rather than displacement.

Aggregation of Cases and Key Independent Variables

A practical problem in estimating models with the LEHD is its size. Because our combined data set consists of over 99.6 million uniquely identified native workers followed over time, the overall individual level data set is over 300 gigabytes in size and cannot be analyzed efficiently, even with the computing resources available at the Census Research Data Centers. In order to facilitate the estimation of models with a data set this large, we aggregate native workers together into larger groups by combining workers of the same gender, race, and wage quintile working in the same 4-digit NAICS industry within the same labor market area. This allows us to include demographic characteristics (race and gender) in the analysis while also reducing the sample size to 11.4 million cases. All of the regressions are weighted by the number of workers in each of the aggregate groups for each dependent variable, so there is no loss in accuracy by aggregating the individual cases together.

The key independent variables for the net flow of immigrant and native workers are measured as a proportional change over the past year. In particular, we define the change in the number of immigrant workers at time T as the proportional change in the number of immigrant workers in industry j, earnings quintile q and labor market area i over the previous four quarters:

---

8 The regression weights for the aggregate groups (see below) for these conditionally constructed dependent variables take the changes in the sample size for each variable into account.
(3) $\Delta \text{immigrant}_{ijqt} = \frac{N\text{immig}_{ijq,t} - N\text{immig}_{ijq,t-4}}{N\text{immig}_{ijq,t-4}}$, with a similar measure for the change in the number of native workers over the previous year. As discussed below in relation to Table 2, we also estimate models with a number of alternative measures of changes in the number of immigrant and native workers.

Regression Models

In our analyses we estimate models with combined fixed effects for time and geographic regions. The reason for this approach is that it allows us to exploit the specific advantages of the LEHD data by following workers over time with detailed measures of exposure to competition with immigrants and industry-specific measures of labor demand shocks.

Equation 4 describes the basic setup for our statistical models. In Equation 4, the dependent variable $\Delta Y_{jgrqit}$ represents one of the labor market outcomes for native workers discussed above where the set of subscripts $(j, q, i, r, \text{and } g)$ represent a specific industry, earnings quintile, area, race, and gender respectively between time $t$ and time $t+4$ quarters:

$$
(4) \Delta Y_{jgrqit} = \beta_0 + \beta_1 \Delta \text{immigrant}_{jqi} + \beta_2 \Delta \text{native}_{jqi} + \beta X_{jgrqit} + u_{ji}^{\text{industry}} + u_{ji}^{\text{area}} + \varepsilon_{jgrqit}
$$

$\Delta \text{immigrant}_{jqi}$ and $\Delta \text{native}_{jqi}$ are the key independent variables representing measures of the change in the number of immigrant and native workers, $X_{jgrqit}$ denotes a set of control variables, and $\varepsilon_{jgrqit}$ is a case-specific error term. Finally, $u_{ji}^{\text{industry}}$ and $u_{ji}^{\text{area}}$ indicate the potentially confounding effects of unmeasured industry and geographic labor demand shocks which are hypothesized to be correlated with wage changes and the inflow of immigrant workers.

For a small number of cases where the proportional change in the number of native or immigrant workers was greater than 1, we cap the maximum change at 1 to prevent extreme values—such as due to clerical errors in the administrative data—from dominating the results.
As mentioned above, we use time by area fixed effects which eliminates the unobserved area-specific effects (the $u_{jtu}$ term). We then use a combination of two instrumental variables to deal with the possibility of industry specific unobserved factors (the $u_{itu}$ term), which is what we turn to next.

**Instrumental variables for immigrant flows and labor demand shocks**

A. Immigrant flows

As discussed above, a central criticism of the spatial comparisons approach is that if immigrants move to areas that are booming economically, then estimates of the effect of immigration on natives’ wages will be biased. As described in the literature review, the use of lagged locational patterns of immigrants as an instrumental variable for current flows is widespread in the recent literature. If we believe that these lagged locations of immigrants are not correlated with the spatial pattern of current demand shocks, then it may be a reasonable choice as an instrumental variable.

In this paper, we modify the construction of this instrumental variable as is commonly used in the immigration literature (depicted above in Equation 1) to predict the inflow of immigrants to earnings quintiles in detailed industries within geographic areas, as illustrated by Equation 5:

\[
\Delta N_{qikt}^{IV} = \sum_j p_{qijk,t=1} \times \Delta N_{qik}^{IV}
\]

where $\Delta N_{qik}^{IV}$ is the predicted change in the number of immigrants in earnings quintile q of industry i of region k, $p_{qijk,t=1}$ is the 3-year lagged proportion of immigrants of national origin group j in region k working work in quintile q of industry i, and $\Delta N_{jk}$ is the overall change in the number of immigrants of origin group j in region k working in earnings quintile q between quarter T and T-4 (see Steinhardt 2009 for a similar approach). Based on Equation 5, we calculate the enclave-IV for immigrant flows by substituting $\Delta N_{qik}^{IV}$ for the numerator of Equation 3 to get:

\[
IV_{enclave}^{immig} = \Delta N_{qik}^{IV} / N_{qij,1-4}^{immig}
\]
The theoretical justification for doing this relies upon the role that social networks and employment niches play in generating a correlation between previous and current industry employment patterns of specific immigrant groups (e.g., Waldinger and Lichter 2003).10

In addition, we also include an alternative version of this IV that incorporates the flow of potentially displaced native workers. As discussed above on page 20, the variables inflow1 and inflow0 measure the inflow of native workers from industry-quintiles with higher and lower immigrant density than the current location. Conceptually, if native workers are displaced due to immigration in industry J and move to industry K (thereby increasing the labor supply and driving down wages), then the effect of immigration in industry J has essentially been projected onto industry K. To capture this effect, we first calculate the average level of the instrumental variable for the inflow of immigrants in the origin industry-quintiles of incoming native workers, $IV_{native-inflow}^{qik}$. Then we calculate the native-inflow adjusted IV as:

$$ IV_{adjusted-enclave}^{qikt} = (1 - inflow1 - inflow0)(IV_{enclave}^{qik}) + (inflow1 + inflow0)(IV_{native-inflow}^{qik}) $$

where inflow1 and inflow0 are the inflows of native workers into the industry-quintile described above, and $IV_{enclave}^{qik}$ is the enclave IV from Equation 6.

B. Labor Demand Shocks

As discussed above in the literature review, a number of authors have expressed concern with the enclave-based IV for immigrant inflows, arguing that the lagged locations of immigrant groups could be correlated with current patterns of labor demand. In this paper, we test the validity of the enclave IV in order to strengthen the credibility of results using this approach vis-à-vis competing approaches. To do this we use a second instrumental variable. In this case it is a measure of labor demand shocks in tradable industries, where a positive labor demand shock is defined as an increase in the demand for labor (i.e.,

10 In our analysis, we use a lag of 3 years in the final analysis, but we also estimated lags of 4 and 5 years, which didn’t affect the results (but reduced the sample size).
due to a change in consumer preferences) that results in an increase in employment in a particular industry. We use an IV for demand shocks because they are otherwise difficult to disentangle from changes in labor supply; the challenge is that changes in employment are the product of both labor supply and demand changes, and an increase in supply, such as through immigration, can result in an increase in employment without any change in demand.

Bartik (1991, 2013) argues that for tradable industries—for industries that produce goods that can be traded across regions—a suitable instrument for changes in labor demand that is independent of changes in local labor supply is the proportional change in the number of workers in that industry at a national level, excluding the local labor market. The logic behind the Bartik measure is that the production of tradable goods—products that can be transported across space and consumed in areas other than the ones they were produced in—responds to national-level changes in the demand for those goods. While the change in local employment in a particular tradable industry might be the result of either changes in demand or changes in local labor supply, the national level change in employment (excluding the local labor market) is a proxy for the change in demand affecting local producers. The calculation of this measure is depicted in Equation 8:

\[
(8) \quad b_{it} = \sum_{j} \frac{e_{ij,t-1} - e_{ij,t-1}}{\tilde{e}_{j,t-1}}
\]

where \(b_{it}\) is the IV estimate of the labor demand shock in geographic area \(i\) at time \(t\), \(e_{ij,t-1}\) is employment in industry \(j\) and area \(i\) at time \(t-1\), and \(\tilde{e}_{j,t}\) indicates national level employment in industry \(j\) excluding area \(i\) at time \(t\) (see Saks 2005:40).

In this paper, because our analysis focuses on the impact of immigration at the industry level within labor market areas (LMAs), we use an industry specific version of this instrument:

11 The following are classified as tradable industries: Agriculture, mining, manufacturing, wholesale trade, transportation, information, finance, professional and technical services, and arts-entertainment.
which is just the term within the parentheses in Equation 8. In our analysis, we will use this industry-specific measure of labor demand shocks to test whether our enclave-IV for immigrant flows is indeed uncorrelated with measures of contemporaneous labor demand shocks.

In some of the analyses below we will focus exclusively on within-region differences in tradable industries to take advantage of our ability to test the relationship between the enclave-IV for immigrant supply shocks and the Bartik IV for labor demand shocks. In theory, firms in tradable industries can absorb immigrants easier because any increase in production can be shipped outside the local area. Nonetheless, in the short term we would still expect to see downward pressure on wages, which provides the basic signal for firms to increase employment and expand production. In the medium or long term we might predict a shift in product mix towards tradable industries that are hiring immigrants (Gaston and Nelson 2007; Lee 2007). However, as discussed above, little evidence of this shift is apparent in the existing literature (Lewis 2003; Dustman and Glitz 2015).

RESULTS

Descriptive Statistics

In our analysis below, we use three different samples. Sample 1 is tradable industries for states that have been in the LEHD data for at least three years (due to the lag time needed for the enclave IV). Sample 2 expands Sample 1 to include all industries, and Sample 3 is all industries and all time points. Table 1 shows summary statistics for each of these three samples, weighted by the number of workers in each of the aggregate cells. For example, the mean value of the variable “firm mobility” for Sample 2 (in row 4) indicates that among workers who stayed in the same state over the one year period from time $T$ to $T+4$ quarters, 17% changed firms.

In Table 1, we see that the weighted average change in the number of immigrant workers in a particular industry-quintile cell was 0.022 (or 2.4%) in row 7 of Sample 2, while the average change in
the number of native workers was 0.072. We use these measures of the proportional change in the number of workers—rather than the change in the percentage immigrant in the industry-quintile—in order to estimate separate effects of increases in labor supply for native and immigrant workers. In addition, we also estimate models using the proportional change in the number of immigrant and native workers over the previous four years, and the overall change at the industry level.

The variable “change in undocumented workers” (row 8) is the proportional change in the number of undocumented workers over the past year, using the flag in the data for an invalid social security number (SSN) discussed above. The variables inflow1 and inflow0 indicate the influx of native workers from industry-quintiles with higher and lower immigrant densities than their current industry-quintile respectively, as discussed above. Next, in addition to the enclave-based IV, \( IV_{enclave} \) (in row 11), calculated using Equation 6, Table 2 also shows summary statistics for the modified version of this IV (\( IV_{adjusted-enclave} \) in row 16) which incorporates the effect of potentially displaced incoming native workers as discussed above on page 25.

The final three rows of Table 3 show counts of the (unweighted) average number of workers within each of the cases that result from aggregating the individual cases by gender, race, LMA, industry, and wage quintile. Referring to Sample 2, on average there are 38.68 individual workers in each of these cells, and 538.89 if we combine the gender and race groups together within the industry, wage quintile and geographic area, and an average of 2,552.85 at the industry level within geographic areas.

Table 2 shows the results of baseline models to test and calibrate the enclave-IV used in the analysis. All of the models in Table 2 use Sample 1 (tradable industries), with the exception of Model 6, which uses Sample 3. In Models 1 and 2, we test the Bartik IV for labor demand shocks. In Model 1, the

\[12\] Note that these measures of change are larger than the average overall growth of because they are weighted by the current number of cases in each aggregate cell.

\[13\] If we weight by the number of workers (to represent the number of individual cases in the typical workers’ aggregate cell) then the averages across the three levels of aggregation described above are 1425, 5441, and 19301 respectively.
dependent variable is $\Delta\text{immigrant}_{jqi}$, the proportional change in the number of immigrant workers, and in Model 2 the dependent variable is $\Delta\text{native}_{jqi}$. The results for both of these models show that industry specific labor demand shocks result in employment increases for both immigrant and native workers in those industries. Similarly, in Models 3 and 4 we find that the enclave IV for immigrant inflows is positively correlated with $\Delta\text{immigrant}_{jqi}$ and the change in the number of undocumented immigrants.

In Models 5 and 6 we test the correlation between immigrant inflows and the proxy variable for flows of undocumented immigrants based on the flag for invalid social security numbers. Model 5 uses tradable industries (Sample 1) and Model 6 uses all industries and available years (Sample 3). In both cases, we see a strong positive association between the two. This is important because it means that although it is likely that the LEHD undercounts the number of undocumented immigrants in the data, the evidence indicates that these flows—as proxied by the flagged SSN variable—are positively correlated with the flow of immigrants measured by $\Delta\text{immigrant}_{jqi}$.

Finally, in Model 7 we see that there is no relationship between the Bartik IV for labor demand shocks and the enclave IV for immigrant flows. While the Bartik IV has a positive effect on immigrant and native employment in Models 1 and 2, and the enclave-IV has a positive effect on immigrant labor supply in Models 3 and 4, they have no correlation with each other: The coefficient for the bivariate regression of $IV_{enclave_{qikt}}$ on the Bartik is -0.000127 (p-value = 0.142).\(^{14}\) This gives us confidence that our instrumental variable for immigrant inflows based on lagged employment patterns is not being affected by current patterns of labor demand.

[Table 3 about here]

Next, Table 3 presents models for the change in log earnings. Note that all of the models include dummy variables for gender, race, and earnings quintiles (which are not presented in the tables to save

\(^{14}\) Note, that, if anything, there is a very small negative correlation, which would suggest that the enclave-IV is slightly associated with negative demand shocks, which would tend to make the effect of immigration look more negative.
space) as well as fixed effects for LMA by time period. The top three rows of Table 3 show the model type, the model number, and the sample used. Because we repeat many of same types of models for different samples and dependent variables, the model type refers to the specific combination of explanatory variables in the model. Mode Type A, for example, estimates a baseline model with no instrumental variables for both Sample 1 (in Model 1) and Sample 2 (in Model 4). In Model 1, there is a positive relationship between \( \Delta \text{immigrant}_{jqi} \) and wage changes for native workers. This is what we might expect based on the potential endogeneity of immigrants’ location choices—in the sense that they may choose to seek work in places and industries that are booming—and it is consistent with the results in Model 1 of Table 3, which shows a positive relationship between \( \Delta \text{immigrant}_{jqi} \) and industry-specific labor demand shocks.

In Model 2 of Table 4, we use the IV variable for the change in the number of immigrants, \( IV_{\text{enclave}}^{qikt} \), instead of the observed change, \( \Delta \text{immigrant}_{jqi} \). In this model, the effect on wages is negative, but the magnitude of the effect is small (-0.0162). Note also in Models 1 and 2 that the coefficient on the inflow of native workers from industry-quintiles with higher levels of immigrant density than their current one (flow1) is positive (i.e., 0.203 in Model 2) and larger than the effect of workers arriving from industry-quintiles with lower levels of immigrant density (flow0). As discussed above, the use of these measures of the inflow of native workers allows us to incorporate the mobility of potentially displaced native workers in the models.

Nonetheless, while Model 2 suggests that inflows of potentially displaced native workers from high-immigration industries do not have a negative effect on earnings, it is possible that that the native worker inflow variables are themselves reflecting the non-random movement of immigrants that affected the variable \( \Delta \text{immigrant}_{jqi} \) in Model 1. For example, if immigrants move towards industries experiencing positive demand shocks, then native workers who move from those booming industries may only do so if they receive a sufficiently high wage in the destination industry to justify the move. As a result, in Model 3 we use the enclave-IV from Equation 7 that has been adjusted to incorporate the
in the origin industries of incoming native workers. In Model 3 the coefficient on

$IV_{qikt}^{\text{enclave}}$ is larger than from Model 2 (-0.0292 compared to -0.0162), but it is still relatively small in magnitude.

Finally, Models 4-6 repeat model types A, B, and C for all industries (Sample 2), and Model 7 repeats model type C for Sample 3 (all industries and time periods). The results in Models 4-7 closely resemble the results for the corresponding models 1-3 for tradable industries discussed above.

Table 4 presents additional models for the change in log earnings. Due to space constraints, only the key coefficients of interest are presented. Model 8 presents a more conventional version of Model 1 without the inflow0 and inflow1 variables to make it consistent with the parameterization of models in the existing literature. In this case the coefficient on $\Delta\text{immigrant}_{qi}$ is negative but close to zero (-0.00254).

Model 9 is the same as Model 8, but it uses the values of $\Delta\text{immigrant}_{qi}$ and $\Delta\text{native}_{qi}$ calculated over the past four years. This longer time period doesn’t affect the results. Similarly, Model 10 (which is Model Type E) is the same as Model 3 using $IV_{qikt}^{\text{adjusted-enclave}}$ as the key independent variable, except that it is calculated over a four year period as well. Again, this doesn’t affect the results qualitatively. Finally, Model 11 (which is Model Type D) is the same as Model 2, except that $IV_{qikt}^{\text{enclave}}$ is calculated at the industry level, rather than the industry-quintile level. Interestingly, this is the largest result in terms of magnitude, suggesting that a 10% increase in the number of immigrants leads to a 0.5 percentage point decrease in the wages of native workers. Next, Models 12-15 of Table 4 repeat Models 8-11 for Sample 2 (all industries) with almost identical results.

Table 5 presents results for the other dependent variables. All of the models in Table 5 use linear probability models with robust standard errors. Linear probability models are used for convenience so that we can make direct comparison of the magnitude of the effects without having to worry about the
rescaling of the coefficients across models that occurs in logit and probit models (Cramer 2007). Table 5 uses Model Types A, C, D, and E corresponding to similar models already discussed for earnings changes in Tables 4 and 5.

First, Column 1 of Table 5 shows the results for cross-state mobility. Keep in mind that, as reported in Row 3 of Table 2, the average level of cross-state mobility is 0.016, so the magnitude of the effect for geographic mobility must be interpreted with this average level in mind. For Model Type C, the coefficient on immigrant inflows is 0.0141 for tradable industries and 0.0263 for all industries. In general, for model types C, D, and E the effect for all industries is about twice the size as the effect for tradable industries. Two comments about this result: First, it suggests that there is a real effect, and that geographic mobility may be one of the mechanisms of adjustment in response to an increase in the supply of immigrants to a particular industry. For the all-industries result, it means that a 10% increase in immigrants leads to an increase of 0.00263 in the probability of migrating across states, which is a 16.4% increase in the migration rate (100*0.00263/0.016). Second, the magnitude of the effect is still small, implying that adjustment to immigration via geographic mobility works slowly over time.

Next, Column 2 of Table 5 presents models where the dependent variable is firm mobility. Similar to the results for geographic mobility, the effect is considerably larger for Sample 2 compared to Sample 1 for all the models using the enclave-IV variable (C, D, and E). Given that the average level of firm mobility is 0.171 (Row 4 of Table 2), the magnitude of the coefficients suggests that the impact of immigration on firm mobility is similar to that of geographic mobility as discussed above. For example, for the all-industries result for Model Type C, a 10% increase in the number of immigrants would increase the probability of firm mobility by 0.0157 which is a 9.2% increase (100*0.0157/0.171). Nonetheless, the impact of immigration on firm mobility is consistent with either a displacement effect, where natives are pushed out, or a replacement effect, where natives move up vacating the positions for immigrants.

In order to differentiate between displacement and replacement, we turn to the models of industry mobility in columns 3 and 4 of Table 5. First, as mentioned above, we define industry mobility
conditional upon firm mobility in order to differentiate the two effects. In row 2 of Table 1, the average rate of industrial mobility among workers who change firms is 71.7%. In the analysis in Table 5, this is divided into two components: The variable “Ind-Up” is mobility with an earnings gain, and the variable “Ind-Down” is mobility with an earnings loss. An important finding here is that for all of the model types, the effect of immigration has a greater effect on Ind-Up than on Ind-Down, and in almost every case the difference in the effects is about two-fold. This indicates evidence in support of the argument made by Cattaneo, Fiorio, and Peri (2013) that one of the mechanisms of adjustment to immigration might actually be the upward mobility of native workers.

Finally, column 5 of Table 5 reports results for the dependent variable for non-employment. As noted above, this is coded as 1 for anyone who is in the LEHD data at time T but not at time T+4 quarters. As a result, it is also picking up migration to states that are not in the sample as well as movement to economic sectors not covered by unemployment insurance (as noted above, the LEHD covers approximately 97% of private employment). The mean level of a move to non-employment is 0.087 for Sample 2 in Table 1. For the results in Table 5, we see a familiar pattern. The effect of Δimmig is negative in Model Type A, but models that use the enclave-IV (Model Types C, D, and E) all show positive effects of immigrant inflows on the non-employment of native workers. Taking the result for Sample 2, which is 0.0449, we would conclude that a 10% increase in the number of immigrants increases the rate of movement to non-employment by 5.2% (100*.1*0.0449/0.087).

Table 6 provides an overall assessment of the magnitude of these predicted effects. For an upper bound of the estimated coefficients, we use the results from Model Type D with Sample 2 (all industries), because these coefficients are the largest. For the lower bound, we use Model Type C, also with Sample 2.\textsuperscript{15} Panel A of Table 6 shows the coefficients for each of the outcome variables. For industry mobility (which is conditional on changing firms), we multiply the coefficient for firm mobility with the sum of

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Industry Mobility} & \textbf{Firm Mobility} & \textbf{Sum} \\
\hline
Model Type A & 0.1 & 0.5 \\
Model Type B & 0.2 & 0.8 \\
Model Type C & 0.3 & 0.9 \\
Model Type D & 0.4 & 1.0 \\
Model Type E & 0.5 & 1.1 \\
\hline
\end{tabular}
\caption{Coefficient Multiplication Table}
\end{table}

\textsuperscript{15} We note that in general the magnitude of the effects is smaller for tradable industries (Sample 1) which makes sense given that these industries could presumably respond more rapidly to immigration through inter-city trade, as is discussed above.
the coefficients for upward and downward industry mobility.\textsuperscript{16} The total out mobility effect, in Column E of Panel A, is the sum of the estimated mobility coefficients in Columns B-D. Column F presents a combined earnings coefficient, which is calculated as

\begin{equation}
\beta_{\text{combined}} = \frac{\beta^A}{1 - \beta^E},
\end{equation}

where $\beta^A$ is the coefficient on earnings growth in Column A, and $\beta^E$ is the combined out-mobility effect in Column E. Equation 10 is adapted from Borjas (2006:227), who provides a theoretical model to describe how the out-migration of natives ($\beta^E$) leads to an underestimation of the effect on earnings. This relationship should be considered suggestive rather than definitive, but it does provide a way to combine the earnings and mobility responses together. For the upper bound coefficients in Panel A, the total out-mobility effect (0.1875) suggests that the impact on earnings is underestimated by about 23% based on Equation 10.

In Panel B we estimate the predicted impact, multiplying the coefficients from Panel B with the annual increase in the number of immigrants from Column H. The estimated combined earnings effect is given in Column K, using $\beta_{\text{combined}}$ from Panel A. For Scenarios B, C, and D of Panel B, we first calculated the annual proportional increase in immigrants implied by the increase in immigrant density at the industry, earnings quintile, and LMA level of native workers from 2000 to 2008, using data from the 2000 Census and the 2007-2009 American Community Survey. For instance, for the average native high school dropout, the average industry-quintile exposure to immigrants went up from 0.1028 to 0.1321 over this time period. This implies an average annual rate of increase in the size of the immigrant population of 0.0355.\textsuperscript{17} For all three groups that are considered in Panel B—all native workers, native high school dropouts, and African Americans—the estimated combined short-term effect on earnings growth are real but modest in size. For instance, for native high school dropouts, the upper bound estimate in Column K

\textsuperscript{16} For example, the coefficients for firm mobility, Ind-Up, and Ind-Down are 0.205, 0.359, and 0.163 respectively for Sample 2 and Model Type D in Table 5. We estimate the overall industry effect as $0.205 \times (0.359 + 0.163) = 0.1071$.

\textsuperscript{17} The implied annual rate is calculated using the exponential growth formula for the rate based on the proportional increase in the relative size of the immigrant population over the time period.
is a 0.23 percentage point reduction in the rate of annual earnings growth. Nonetheless, one should keep in mind that because effects on earnings growth accumulate over time, over the 8 year period from 2000-2008 this upper-bound effect on the growth rate suggests that wage levels would be about 1.8% lower for high school dropouts, and about 1.2% lower for black workers.

Overall, Panel B of Table 6 finds small negative effects of immigration on earnings growth that over time would have a modest cumulative impact on affected groups of native workers. On the other hand, however, it is likely that the estimates in Panel B overstate the long term effect because of the potential divergence between the short and longer term impact of immigration. Potential medium- or long-term positive effects, either through complementarities in production or through increases in demand caused by immigrants as consumers, are likely to be widely diffused through the labor market and thus difficult to detect except through the assumption of a production function approach such as in Ottaviano and Peri (2012) or with data on regional variation in consumer prices such as in Cortes (2008). Nonetheless, one piece of evidence consistent with longer term results that dampen short term negative effects is the finding for industry mobility in Table 5 which indicates that immigration is associated with a larger effect on upward, rather than downward, industrial mobility for native workers. This suggests that some of what is going on is actually replacement, rather than displacement, at least for industry switchers. In this sense, it is possible that as the market adjusts to the increase in demand caused by immigration, it opens up new positions higher up in the earnings distribution and natives respond via upward mobility.

DISCUSSION AND CONCLUSION

In this paper, we have used data from the LEHD to estimate the effect of immigration on the earnings growth, employment, and the mobility of native workers. The size of the data, and its level of geographic and industry detail, allows us to calculate detailed measures of changes in immigration levels over time in order to precisely identify native workers who are competing with immigrants. We also incorporate measures of the inflow of native workers to account for the degree of industry integration in these
markets, and we use measures of demand shocks to test the validity of the instrumental variable that we use to measure changes in the number of immigrant workers over time.

In our results, we find consistent evidence of negative effects of immigration on the wage growth and employment rates of competing native workers as well as evidence that native workers respond through spatial, firm, and industrial mobility. In addition, as discussed above, we find evidence that part of what is going is a stronger process of replacement rather than displacement, where native workers who switch industries due to immigration are more likely to experience positive rather than negative earnings changes.

Overall, we would argue, as suggested by Kim and Sakamoto (2013), that concerns about the economic impact of immigration should be taken seriously because it is likely that there are effects on native workers who are competing directly with immigrants by virtue of their skill level and employment sector. To the degree that this occurs, immigration will have consequences for changes in inequality (Blau and Kahn 2012) similar to international trade, where some workers face acute negative effects such as job loss, while the positive effects are broadly distributed through increases in labor demand and consumption. At the same time, however, we emphasize that the effect of immigration estimated here is fairly small, even when we attempt to isolate the effect on narrowly demarcated sets of competing native workers. Indeed, one of the benefits of the large size of our data set is that it enables us to obtain precise estimates of small effect sizes. These estimates indicate that while the direction of the effects is as a textbook model of the labor market might predict in terms of lower earnings growth, reduced employment, and outmigration of affected native workers, the magnitude of the effects is not large. If the level of immigration was dramatically higher and workers and firms were unable to adjust, then we would expect the effects to be correspondingly higher. Nonetheless, because the U.S. has a large dynamic economy that is in constant flux, we would argue that the important finding of this paper is that a steady but modest level of immigration similar to the current one (about 0.4% of the total population per year) is likely to be absorbed without much of an overall perceptible effect on native workers.
References


<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Sample 1: Tradable industries, ≥ 3 years in sample.</th>
<th>Sample 2: All industries, ≥ 3 years in sample.</th>
<th>Sample 3: All industries, all years.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>1. Change in earnings</td>
<td>0.067 (0.155)</td>
<td>0.066 (0.154)</td>
<td>0.063 (0.159)</td>
</tr>
<tr>
<td>2. Industry mobility</td>
<td>0.717 (0.235)</td>
<td>0.714 (0.234)</td>
<td>0.749 (0.223)</td>
</tr>
<tr>
<td>3. Moved states</td>
<td>0.016 (0.027)</td>
<td>0.016 (0.027)</td>
<td>0.019 (0.031)</td>
</tr>
<tr>
<td>4. Firm mobility</td>
<td>0.171 (0.138)</td>
<td>0.172 (0.137)</td>
<td>0.179 (0.143)</td>
</tr>
<tr>
<td>5. Non-employment</td>
<td>0.089 (0.073)</td>
<td>0.087 (0.072)</td>
<td>0.087 (0.072)</td>
</tr>
<tr>
<td>Explanatory variables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Bartik IV for labor demand</td>
<td>1.082 (0.229)</td>
<td>1.056 (0.146)</td>
<td>1.054 (0.164)</td>
</tr>
<tr>
<td>7. 1-yr. change in immigrant workers</td>
<td>0.024 (0.110)</td>
<td>0.022 (0.103)</td>
<td>0.025 (0.115)</td>
</tr>
<tr>
<td>8. 1-yr. change in undocumented workers</td>
<td>0.0004 (0.0083)</td>
<td>0.0005 (0.0091)</td>
<td>0.0005 (0.0097)</td>
</tr>
<tr>
<td>9. 4-yr. change in immigrant workers</td>
<td>0.054 (0.146)</td>
<td>0.053 (0.145)</td>
<td>0.058 (0.159)</td>
</tr>
<tr>
<td>10. 4-yr. change in undocumented workers</td>
<td>0.001 (0.012)</td>
<td>0.001 (0.013)</td>
<td>0.001 (0.013)</td>
</tr>
<tr>
<td>11. IV for 1-yr. change in immigrant workers</td>
<td>0.008 (0.041)</td>
<td>0.007 (0.023)</td>
<td>0.008 (0.025)</td>
</tr>
<tr>
<td>12. 1-yr. change in native workers</td>
<td>0.078 (0.255)</td>
<td>0.072 (0.245)</td>
<td>0.075 (0.264)</td>
</tr>
<tr>
<td>13. 4-yr. change in native workers</td>
<td>0.162 (0.332)</td>
<td>0.158 (0.329)</td>
<td>0.152 (0.350)</td>
</tr>
<tr>
<td>14. Native worker inflow, from lower immig. industries</td>
<td>0.039 (0.036)</td>
<td>0.039 (0.035)</td>
<td>0.042 (0.038)</td>
</tr>
<tr>
<td>15. Native worker inflow, from higher immig industries</td>
<td>0.036 (0.035)</td>
<td>0.037 (0.035)</td>
<td>0.034 (0.036)</td>
</tr>
<tr>
<td>16. adjusted IV including inflow of native workers</td>
<td>0.008 (0.040)</td>
<td>0.007 (0.021)</td>
<td>0.008 (0.024)</td>
</tr>
<tr>
<td>17. Average N (within aggregate cell), unweighted</td>
<td>30.54 (185.29)</td>
<td>38.68 (231.59)</td>
<td>39.67 (240.81)</td>
</tr>
<tr>
<td>18. Average N (within industry-quintile), unweighted</td>
<td>444.20 (1692.35)</td>
<td>538.89 (1809.42)</td>
<td>563.31 (1913.31)</td>
</tr>
<tr>
<td>19. Average N (industry level), unweighted</td>
<td>2068.17 (6094.08)</td>
<td>2552.85 (7389.96)</td>
<td>2675.73 (7807.82)</td>
</tr>
</tbody>
</table>
Table 2: Baseline models to test and calibrate the IVs

<table>
<thead>
<tr>
<th>Sample Used</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in immigrant workers</td>
<td>1-Tradable</td>
<td>1-Tradable</td>
<td>1-Tradable</td>
<td>1-Tradable</td>
<td>1-Tradable</td>
<td>3-All</td>
<td>1-Tradable</td>
</tr>
<tr>
<td>Change in native workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV change in immigrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VARIABLES

- Change in immigrant workers
- Change in native workers
- Change in immigrant workers
- Change in undoc. immigrants
- Change in immigrant workers
- Change in immigrant workers
- IV change in immigrants

<table>
<thead>
<tr>
<th>industry level Bartik IV for labor demand shocks</th>
<th>0.0392***</th>
<th>0.104***</th>
<th>-0.000127</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.000418)</td>
<td>(0.000955)</td>
<td>(8.66e-05)</td>
</tr>
<tr>
<td>IV change in immigrants (IV_enclave_qikt)</td>
<td>0.923***</td>
<td>0.0146***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00226)</td>
<td>(0.000194)</td>
<td></td>
</tr>
<tr>
<td>Change in undoc. immigrants</td>
<td></td>
<td></td>
<td>1.548***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.805***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.00562)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.00372)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.00305***</td>
<td>0.00467***</td>
<td>3.85e-05****</td>
</tr>
<tr>
<td></td>
<td>(5.40e-05)</td>
<td>(0.000123)</td>
<td>(5.31e-05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.55e-06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5.35e-05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.99e-05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.12e-05)</td>
</tr>
<tr>
<td>Observation (rounded to the nearest 10,000)</td>
<td>4,390,000</td>
<td>4,388,000</td>
<td>4,390,000</td>
</tr>
<tr>
<td></td>
<td>4,390,000</td>
<td>4,390,000</td>
<td>11,390,000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.002</td>
<td>0.003</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>0.017</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses
### Table 3: Models for the Change in Log Earnings (shaded cells indicate the key independent variable in each model)

<table>
<thead>
<tr>
<th>Model Type</th>
<th>A (1)</th>
<th>B (2)</th>
<th>C (3)</th>
<th>A (4)</th>
<th>B (5)</th>
<th>C (6)</th>
<th>C (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>1- Tradable</td>
<td>1- Tradable</td>
<td>1- Tradable</td>
<td>2- All</td>
<td>2- All</td>
<td>2- All</td>
<td>3-all time periods</td>
</tr>
<tr>
<td><strong>VARIABLES</strong>&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δimmigrant&lt;sub&gt;jqi&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td>0.000586</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.000628)</td>
<td></td>
<td></td>
<td></td>
<td>(0.000466)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δundoc. immigrants</td>
<td>-0.00283</td>
<td></td>
<td></td>
<td>0.102***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.00615)</td>
<td></td>
<td></td>
<td></td>
<td>(0.00438)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δnative&lt;sub&gt;jqi&lt;/sub&gt;</td>
<td>0.00532***</td>
<td>0.00600***</td>
<td>0.00988***</td>
<td>0.00525***</td>
<td>0.00582***</td>
<td>0.00799***</td>
<td>0.00883***</td>
</tr>
<tr>
<td>(0.000275)</td>
<td>(0.000228)</td>
<td>(0.000226)</td>
<td>(0.000197)</td>
<td>(0.000161)</td>
<td>(0.000160)</td>
<td>(0.000138)</td>
<td></td>
</tr>
<tr>
<td>Native worker inflow, from lower immig industries (flow0)</td>
<td>0.0643***</td>
<td>0.0650***</td>
<td></td>
<td>0.0352***</td>
<td>0.0361***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.00166)</td>
<td>(0.00166)</td>
<td></td>
<td>(0.00118)</td>
<td>(0.00118)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native worker inflow, from higher immig industries (flow1)</td>
<td>0.204***</td>
<td>0.203***</td>
<td></td>
<td>0.142***</td>
<td>0.140***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.00169)</td>
<td>(0.00168)</td>
<td></td>
<td>(0.00118)</td>
<td>(0.00118)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV-Δimmigrant&lt;sub&gt;jqi&lt;/sub&gt; (IV&lt;sub&gt;enclave&lt;/sub&gt;)</td>
<td>-0.0162***</td>
<td></td>
<td></td>
<td>-0.0126***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.00250)</td>
<td></td>
<td></td>
<td>(0.00182)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV-Δimmigrant&lt;sub&gt;jqi&lt;/sub&gt; including inflows (IV&lt;sub&gt;adjusted-enclave&lt;/sub&gt;)</td>
<td>-0.0292***</td>
<td></td>
<td></td>
<td>-0.0215***</td>
<td>-0.0149***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.00271)</td>
<td></td>
<td></td>
<td>(0.00195)</td>
<td>(0.00154)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy variables for gender, race, and earnings quintile?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0454***</td>
<td>-0.0454***</td>
<td>-0.0483***</td>
<td>-0.0511***</td>
<td>-0.0511***</td>
<td>-0.0532***</td>
<td>-0.0537***</td>
</tr>
<tr>
<td>(0.000116)</td>
<td>(0.000116)</td>
<td>(0.000113)</td>
<td>(8.56e-05)</td>
<td>(8.56e-05)</td>
<td>(8.26e-05)</td>
<td>(7.35e-05)</td>
<td></td>
</tr>
<tr>
<td>Observations (rounded to the nearest 10,000)</td>
<td>4,390,000</td>
<td>4,390,000</td>
<td>4,390,000</td>
<td>8,730,000</td>
<td>8,730,000</td>
<td>8,730,000</td>
<td>11,390,000</td>
</tr>
<tr>
<td>Number of unique individuals</td>
<td>45,200,000</td>
<td>45,200,000</td>
<td>45,200,000</td>
<td>91,000,000</td>
<td>91,000,000</td>
<td>91,000,000</td>
<td>99,600,000</td>
</tr>
<tr>
<td>Number of unique firms</td>
<td>2,310,000</td>
<td>2,310,000</td>
<td>2,310,000</td>
<td>5,740,000</td>
<td>5,740,000</td>
<td>5,740,000</td>
<td>6,630,000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.400</td>
<td>0.400</td>
<td>0.398</td>
<td>0.452</td>
<td>0.452</td>
<td>0.451</td>
<td>0.446</td>
</tr>
</tbody>
</table>

Notes: 1All models include LMA by time fixed effects, as described in the text. Standard errors are in parentheses. *** p<0.001
<table>
<thead>
<tr>
<th>Model (and notes)</th>
<th>(sample number)</th>
<th>Key independent variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Mod. 1 (Model Type A) without native inflow1 &amp; inflow0</td>
<td>(1) Tradable</td>
<td>Δ immig</td>
<td>-0.00254*** (0.000628)</td>
</tr>
<tr>
<td>9. The same as model 8, except with 4-year changes in native and immigrant employment</td>
<td>(1) Tradable</td>
<td>4-year Δ immig.</td>
<td>0.00614*** (0.000457)</td>
</tr>
<tr>
<td>10. [Model Type E] The same as Model 3 except with 4-year changes in native employment and the IV adjusted−enclave.</td>
<td>(1) Tradable</td>
<td>4-year IV Δ immig, including native inflow ( (IV_{\text{adjusted−enclave}}) )</td>
<td>-0.0170*** (0.00326)</td>
</tr>
<tr>
<td>11. [Model Type D] The same as Model 2, but labor flows calculated at the industry level.</td>
<td>(1) Tradable</td>
<td>Industry level IV Δ immig ( (IV_{\text{enclave}}) )</td>
<td>-0.0517*** (0.00372)</td>
</tr>
<tr>
<td>12. Mod. 4 (Type A) without native inflow1 &amp; inflow0</td>
<td>(2) All industries</td>
<td>Δ immig</td>
<td>-0.00450*** (0.000336)</td>
</tr>
<tr>
<td>13. The same as model 12, except with 4-year changes in native and immigrant employment</td>
<td>(2) All industries</td>
<td>4-year Δ immig.</td>
<td>-0.00450*** (0.000336)</td>
</tr>
<tr>
<td>14. [Model Type E] the same as Model 6 except for 4-year changes.</td>
<td>(2) All industries</td>
<td>4-year IV Δ immig, including native inflow ( (IV_{\text{adjusted−enclave}}) )</td>
<td>-0.0188*** (0.00236)</td>
</tr>
<tr>
<td>15. [Model Type D] The same as Model 5, but labor flows calculated at the industry level.</td>
<td>(2) All industries</td>
<td>Industry level IV Δ immig ( (IV_{\text{enclave}}) )</td>
<td>-0.0511*** (0.00246)</td>
</tr>
</tbody>
</table>

Notes: All models include LMA by time fixed effects, as described in the text. Standard errors are in parentheses. *** p<0.001
### Table 5: Linear Probability Models for Non-Employment and Industry, Firm, and Geographic Mobility

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Δ immigr.</strong></td>
<td>(1) Tradable</td>
<td>-0.00920***</td>
<td>-0.0359***</td>
<td>0.159***</td>
<td>0.0544***</td>
<td>-0.0121***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000144)</td>
<td>(0.000571)</td>
<td>(0.000705)</td>
<td>(0.000535)</td>
<td>(0.000281)</td>
</tr>
<tr>
<td></td>
<td>(2) All</td>
<td>-0.00131***</td>
<td>-0.0396***</td>
<td>0.182***</td>
<td>0.0622***</td>
<td>-0.0152***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.97e-05)</td>
<td>(0.000436)</td>
<td>(0.000532)</td>
<td>(0.000383)</td>
<td>(0.000221)</td>
</tr>
<tr>
<td><strong>Δ undoc. immigr.</strong></td>
<td>(1) Tradable</td>
<td>0.0127***</td>
<td>0.0655***</td>
<td>-0.0388***</td>
<td>-0.000805</td>
<td>0.0465***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00141)</td>
<td>(0.00559)</td>
<td>(0.00713)</td>
<td>(0.00523)</td>
<td>(0.00275)</td>
</tr>
<tr>
<td></td>
<td>(2) All</td>
<td>0.0137***</td>
<td>0.127***</td>
<td>-0.0820***</td>
<td>-0.0305***</td>
<td>0.0691***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000938)</td>
<td>(0.00410)</td>
<td>(0.00506)</td>
<td>(0.00360)</td>
<td>(0.00207)</td>
</tr>
<tr>
<td><strong>C. IV Δ immigr.</strong></td>
<td>(1) Tradable</td>
<td>0.0141***</td>
<td>0.0238***</td>
<td>0.147***</td>
<td>0.0877***</td>
<td>0.0185***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000625)</td>
<td>(0.00251)</td>
<td>(0.00313)</td>
<td>(0.00231)</td>
<td>(0.00122)</td>
</tr>
<tr>
<td></td>
<td>(2) All</td>
<td>0.0263***</td>
<td>0.157***</td>
<td>0.281***</td>
<td>0.128***</td>
<td>0.0449***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000421)</td>
<td>(0.00194)</td>
<td>(0.00228)</td>
<td>(0.00161)</td>
<td>(0.000933)</td>
</tr>
<tr>
<td><strong>D. industry-level IV Δ immigr.</strong></td>
<td>(1) Tradable</td>
<td>0.0168***</td>
<td>0.0546***</td>
<td>0.223***</td>
<td>0.129***</td>
<td>0.0260***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000853)</td>
<td>(0.00338)</td>
<td>(0.00428)</td>
<td>(0.00317)</td>
<td>(0.00166)</td>
</tr>
<tr>
<td></td>
<td>(2) All</td>
<td>0.0316***</td>
<td>0.205***</td>
<td>0.359***</td>
<td>0.163***</td>
<td>0.0489***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000525)</td>
<td>(0.00230)</td>
<td>(0.00281)</td>
<td>(0.00202)</td>
<td>(0.00116)</td>
</tr>
<tr>
<td><strong>E. 4-year IV Δ immigr.</strong></td>
<td>(1) Tradable</td>
<td>0.0154***</td>
<td>0.0107***</td>
<td>0.278***</td>
<td>0.143***</td>
<td>0.0152***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000752)</td>
<td>(0.000302)</td>
<td>(0.000383)</td>
<td>(0.000278)</td>
<td>(0.000147)</td>
</tr>
<tr>
<td></td>
<td>(2) All</td>
<td>0.0322***</td>
<td>0.180***</td>
<td>0.465***</td>
<td>0.213***</td>
<td>0.0484***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000509)</td>
<td>(0.000234)</td>
<td>(0.000279)</td>
<td>(0.000195)</td>
<td>(0.000113)</td>
</tr>
</tbody>
</table>

Notes: All models include LMA by time fixed effects, as described in the text. Standard errors are in parentheses. *** p<0.001
Table 6: Estimates of the Overall Effect

Panel A: Summary of estimated coefficients

<table>
<thead>
<tr>
<th>Column</th>
<th>A. earnings growth (direct effect)</th>
<th>B. Geographic migration</th>
<th>C. Industry mobility</th>
<th>D. Non-employment</th>
<th>E. Total out mobility (B+C+D)</th>
<th>F. Combined earnings coefficient $\beta_{combined}^{HIGH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated upper bound coefficients, $\hat{\beta}^{HIGH}$, (Model Type D, Sample 2)</td>
<td>-0.0511</td>
<td>0.0316</td>
<td>0.1071</td>
<td>0.0489</td>
<td>0.1875</td>
</tr>
<tr>
<td></td>
<td>Estimated lower bound coefficients, $\hat{\beta}^{LOW}$, (Model Type C, Sample 2)</td>
<td>-0.0215</td>
<td>0.0263</td>
<td>0.0642</td>
<td>0.0449</td>
<td>0.1354</td>
</tr>
</tbody>
</table>

Panel B: Estimated Effect, by Scenario

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Effect size used</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect size used</td>
<td>Average Annual $\Delta$ immig.</td>
<td>Earnings growth (direct effect)</td>
<td>Total out mobility</td>
<td>Combined earnings Effect</td>
<td></td>
</tr>
<tr>
<td>A. 10% increase</td>
<td>Upper</td>
<td>0.10</td>
<td>-0.00511</td>
<td>0.01875</td>
<td>-0.0063</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>0.10</td>
<td>-0.00215</td>
<td>0.01345</td>
<td>-0.0025</td>
<td></td>
</tr>
<tr>
<td>B. Native workers (2000-2008)</td>
<td>Upper</td>
<td>0.0273</td>
<td>-0.0014</td>
<td>0.0051</td>
<td>-0.0017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>0.0273</td>
<td>-0.0006</td>
<td>0.0037</td>
<td>-0.0007</td>
<td></td>
</tr>
<tr>
<td>C. High school dropouts (2000-2008)</td>
<td>Upper</td>
<td>0.0355</td>
<td>-0.0018</td>
<td>0.00665</td>
<td>-0.0023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>0.0355</td>
<td>-0.0008</td>
<td>0.00481</td>
<td>-0.0009</td>
<td></td>
</tr>
<tr>
<td>D. African American workers, High school or less (2000-2008)</td>
<td>Upper</td>
<td>0.0250</td>
<td>-0.0013</td>
<td>0.00469</td>
<td>-0.0015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>0.0250</td>
<td>-0.0005</td>
<td>0.00339</td>
<td>-0.0006</td>
<td></td>
</tr>
</tbody>
</table>

Overall increase in immigrant density, 2000-2008 at the industry, earnings quintile, and Super PUMA level for full time workers with at least 20 weeks worked and 20 hours per week. All workers: 0.1255 to 0.1634. Native workers: 0.0863 to 0.1052. High school dropouts: 0.1028 to 0.1321. African American workers: 0.1047 to 0.1250. (Source: 2000 Census PUMS and American Community Survey 2007-2009)