



IZA DP No. 5328

## Empirical Methods in the Economics of International Immigration

Fernando A. Lozano  
Michael D. Steinberger

November 2010

# **Empirical Methods in the Economics of International Immigration**

**Fernando A. Lozano**

*Pomona College  
and IZA*

**Michael D. Steinberger**

*Pomona College*

Discussion Paper No. 5328  
November 2010

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Empirical Methods in the Economics of International Immigration**<sup>\*</sup>

In this chapter we provide a brief overview of the main empirical tools used by economists to study international migration. We begin by exploring the three broad research areas that economists examine when researching immigration. We then explore the strengths and shortcomings of the standard methods, and highlight new methods that will likely become more common in future work in the field. We divide the most common tools used in the empirical literature into four broad categories: (1) Ordinary Least Squares and Inference, (2) Difference-in-Difference Estimation, (3) Instrumental Variables Techniques, and (4) Recent Developments and Distributional Estimators. We use recent empirical work to highlight and explain each method, and provide sources for researchers interested in further information on each topic.

JEL Classification: F22

Keywords: immigration, empirical methods

Corresponding author:

Fernando A. Lozano  
Pomona College  
425 N. College Ave  
Claremont, CA 91711  
USA  
E-mail: [fernando.lozano@pomona.edu](mailto:fernando.lozano@pomona.edu)

---

<sup>\*</sup> Forthcoming in *Handbook of Research Methods in Immigration*. We are grateful to Edward Funkhouser and Carlos Vargas for useful comments and Maria Zhu for providing excellent research assistance. All errors are our own.

## **Introduction**

The United Nations reports that by 2010 more than 213 million people, or five percent of the world population, will live and work in a county in which they were not born. Not only is the stock of worldwide international immigrants significant in its own right, the net flow of immigrants has gradually increased over the last 20 years. During the decade of the 1990s, more than 23 million people moved to a different country; during the next decade this figure grew to 35 million people. As international migration increasingly prevails across different regions and countries it is natural to ask how do economists address the causes and consequences of these flows, and what are the strengths and shortcomings of the methodologies employed by economists?

The economics of immigration is a burgeoning field whose interest and research expands over different many facets of immigration. The evolution of economists' interest on immigration is evidenced by the fact that general interest journals are increasingly publishing papers on this topic. Table 1 presents descriptive statistics about the recent research in nine top economic journals covering an immigration topic in the last 20 years. Among the issues economists address: 1) Why do people migrate to a different country, and who chooses to migrate? 2) What explains the labor market success of immigrants in the host country and their economic assimilation? 3) What is the effect of immigrants in the host economy, especially on the host country's most vulnerable populations? 4) What is the effect of emigration on the sending communities? 5) What are the public policy implications of migration? As Table 1 shows, research on immigration is becoming increasingly prevalent: growing from 23 papers on immigration between 1990-1994, to 51 between 2005-2009 in these nine top journals. In addition, while Ordinary Least Squares remains the most common econometric technique, more

quantitatively sophisticated techniques like DiNardo, Fortin and Lemieux (1996) decomposition and other distributional techniques are becoming more popular. Finally, topics of interest for immigration economists are changing, possibly as a result of changing public perceptions of immigration; the most popular topic in 1990-1994 was immigrant assimilation, yet in 2005-2009 the most common topic was immigrants' effect on natives' labor market outcomes.

In this chapter, we explore four econometric methods commonly used by economists in the field of immigration<sup>1</sup>. We particularly highlight the strengths and potential pitfalls of each approach. We begin in the next section by exploring the various ways Ordinary Least Squares (OLS) regression techniques are used by immigration economists and comment on some of the inherent limitations of this approach. Building on that intellectual foundation, we turn our attention to other common econometric methods. In section three we explore the use of difference-in-differences estimation on immigration. In section four, we address the use of instrumental variables techniques on immigration economics research, and the attempts to establish sound causal relationships between the outcomes highlighted above and immigration flows. In section five, we briefly explore two new econometric techniques we believe will become increasingly common in the future of empirical immigration economics, where interest will increasingly focus on distributional analysis of the influence of international immigration.

## **Section 2: Identification Strategies on Immigration Economics**

---

<sup>1</sup> While this paper focuses in the economics of immigration, we do not explore the consequences of immigration on source countries or sender households. For a discussion in this topic see Funkhouser's essay in this volume.

Economists are interested in analyzing labor market or population outcomes across different counterfactual scenarios. For example, an economist may want to investigate whether the wages earned by natives in a local labor markets are affected by a change in the number of immigrants. To properly analyze such a relationship, one needs to compare the outcome in the labor market with an exogenous shock in the number of immigrants, and the same labor market without the shock. Of course only one of these outcomes is observed, while the other remains unobserved. In order to correctly influence policy, and particularly immigration policy, it is important to identify any causal relationship between the outcome and the explanatory variables, and to understand the validity of any control group (explicit or implicit). This is true for other questions which economists are interested in exploring, for example: who chooses to migrate and the quality of immigrants, what explains the labor market success of immigrants in the host economy, or what determines immigrants' decision to invest or not in human capital.

### *2.1 Methodology to Study Immigrant Assimilation into the Host Economy*

The first set of questions in the economics of immigration concern how immigrants fare in the host economy, whether their labor market productivity or earnings differ from that of natives, and whether different cohorts of immigrants assimilate at different rates. For example, Chiswick (1978) and Borjas (1985, 1995) explore whether immigrant earnings assimilate to that of natives. Several empirical issues arise from estimating a model to answer this question. If the empirical model is:

$$\ln w_i = x_i \theta + \gamma_1 I_i + \gamma_2 I_i \times Y_i + \gamma_3 I_i \times Y_i^2 + \varepsilon_i \quad (1)$$

where  $\ln w$  is the natural log of earnings,  $I$  is an indicator variable whether the observation is an immigrant, and  $Y$  represents the observation's number of years in the source country. The vector

$x'$  includes different demographic characteristics. The parameters of interest are  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$ , where  $\gamma_1$  represents the earnings difference between native and immigrants and  $\gamma_2$  and  $\gamma_3$  represent how these earnings differences change as the immigrant's host country experience increases. Chiswick (1978) estimates an equation similar to (1) using the 1970 U.S. Census with annual earnings (for self-employed and employed workers) as the response variable. He finds that  $\gamma_2 > 0$  and  $\gamma_3 < 0$  suggesting that as host country experience increases, immigrants' earnings increase at a decreasing rate. Importantly, these results imply that the earnings of immigrants overtake the earnings of demographically equivalent native workers in ten to fifteen years. This suggests that immigrants' rapid earnings acceleration is due to investments in host country human capital.

Although groundbreaking in the literature of immigration economics it does not account for out-migration and that different cohorts of immigrants may be of different quality in the labor market. Omitting these variables results in US experience-earnings profiles that are biased upward. Using the 1970 and 1980 Decennial Census, Borjas (1985) explores the nature of this bias by adjusting equation (1) to include a variable for each immigrant cohort  $C_{it}$ . Borjas' empirical model is:

$$\ln w_{it} = x_{it}'\theta + \gamma_1 I_{it} + \gamma_2 I_{it} \times Y_{it} + \gamma_3 I_{it} \times Y_{it}^2 + \delta_1 I_{it} \times C_{it} + \delta_2 I_{it} \times C_{it}^2 + \varepsilon_i \quad (2)$$

Note that if equation (2) is estimated over a single cross-section for year  $T$  of data the vectors  $\gamma$  and  $\delta$  are not identified because  $T=C+Y$ . Hence equation (2) must be estimated over at least two-pooled cross sections. In addition, the vector of characteristics  $x$  must include year controls as well as other demographic characteristics. The estimates from equation (2) show that assimilation rates are much slower than those in equation (1) (Borjas, 1985) and these results suggest that the years in host country profiles are confounded with the relative quality across

different immigrant cohorts. Note that this approach has two important assumptions: the rate of immigrant assimilation is constant across different cohorts, and cohort labor market performance is independent from out-migration, death or labor market separations from immigrants.

Figure 1 shows the predicted differences of log hourly wages between natives and immigrants. These estimates are constructed with and without controls for cohorts, representing equations (2) and (1), respectively. The data is for all working males in the 1990 and 2000 US Census, and the 2006-2007 American Community Survey. The vector of demographic characteristics  $x_{it}$  includes age, age squared and age cube and includes indicator variables for education characteristics (all these variables are interacted with survey years). The dark line represents estimates from equation (1) and the light lines represent estimates for three cohorts from equation (2). Figure 1 shows that equation 1 predicts a much faster assimilation rate, yet as argued above, these estimates confound the years in the U.S. effect with the cohort effect. Once controls for cohorts are included, the assimilation rate is slower. Further, we see there are differences in the intercept for each cohort: differences are smallest with respect to earlier immigrants and greatest among the most recent ones.

There have been two recent significant developments in the assimilation literature. First, Antecol, Kuhn and Trejo (2006) extend the above methodology to estimate whether immigrants' earnings growth is due to returns to host country experience (wage assimilation) or whether they are due to employment assimilation. In particular they use data from Australia, Canada and the United States to determine what percentage of assimilation is due to changes in wages and what percentage is due to changes in employment. To do this they estimate equation (2) using probability of employment as the dependent variable. Their results show that immigrant assimilation in Australia is mostly due to changes in the probability of employment, while



immigrant assimilation in the US is predominantly due to changes in earnings. Their results for immigrants in Canada are between those of immigrants to the US and immigrants to Australia.

The second significant development in this literature is Lubotsky (2007) who uses administrative data to estimate immigrant assimilation in the US. Lubotsky matches Social Security earnings records to the 1990 and 1991 Survey of Income and Program Participation and to the 1994 March Current Population Survey (CPS) Supplement to explore non-random outmigration of immigrants. His results show that outmigration (whether temporary or permanent) tends to be predominantly from low-earner immigrants, and therefore assimilation profiles using Decennial Census data which cannot control for outmigration tend to be overestimated. The result that temporary outmigration tends to be mostly by low-wage transient immigrants is similar to that of Lozano and Sorensen (2010) who show that assimilation estimates from the March CPS and the CPS Outgoing Rotations Groups (ORG) differ due to very recent immigrants being in the March CPS and not in the ORG.

## *2.2 Methodology to Study Immigrant Selection*

A second issue that arises in the economics of immigration has to do with the fact that the decision of an immigrant to migrate or not is not random, and those immigrants who choose to migrate will have different observable and unobservable characteristics than potential immigrants who choose to stay in the source country. The concavity in the estimates of  $\gamma_2$  and  $\gamma_3$  in equation (1), may suggest that immigrants' assimilation is a result of positive non-random immigrant selection. Consider first the selection of who chooses to migrate. It is clear that the selection of immigrants is non-random, as only the potential immigrants whose expected benefits

from migration are greater than the expected costs from migration choose to migrate.

Appropriately accounting for the selection of immigrants is still an ongoing issue for economists.

There are two ways to think about immigrant selection: one studies the quality of immigrants compared to the population of *all* potential immigrants (including those who do not wish to migrate) in the source country, the second compares immigrant characteristics across different source countries. This issue is widely examined by Borjas (1987), Cobb-Clark (1993), Antecol (2000), Blau and Kahn (2010) or Lopez and Lozano (2010).<sup>2</sup> While the methodology differs across different papers, there is an underlying feature: including source country characteristics to account for immigrant's labor market success. The basic methodology, following Borjas' (1987) includes two stages, the first stage estimates a measure of success of the immigrant across two different cross-sections, for example equation (2), and the second stage consists on recovering the cohort parameters  $\hat{\delta}_{1,c}$  and  $\hat{\delta}_{2,c}$  for immigrants' from country  $c$  and use them as a dependent variable regressed on different country characteristics:

$$\left(\hat{\delta}_{1,c} + \hat{\delta}_{2,c}\right) = w_c^1 \phi + \eta_c \quad (3).$$

where the parameter  $\phi$  represents the correlation between the outcome of immigrants from country  $c$  and the characteristics of the country. Cobb-Clark (1993) analyzes women's immigrant characteristics using the host's country gross domestic product, income inequality, country's return to education, distance between country and the US, among other characteristics<sup>3</sup>. The results in this literature show that immigrants from countries with high income inequality or high returns to education tend to have less desirable observable skills than immigrants from other countries.

---

<sup>2</sup> A third way to think of immigrant selection compares immigrants with non-immigrants from the same country as in Chiquiar and Hanson (2002). The econometric methods used in their analysis are detailed in section 5 below.

<sup>3</sup> Unlike Borjas (1987), Cobb-Clark (1993) uses only one stage where the explanatory variables are both person specific (education, age, family structure) and country characteristics.

One promising avenue for research relies on Chiswick's (1986) research that explores how the immigrant characteristics differ across different source countries. Chiswick compares the earnings of immigrants with the earnings of native born workers who are of the same ethnicity as the immigrant. This is similar to Trejo (2003) who compares the outcomes of Mexican-American immigrants in the U.S. across different generations. Trejo expands equation (2) into a fully interacted model that includes interactions between first (foreign-born), second (children of foreign-born) and third generation immigrants and the other characteristics.

Finally, it is important to recognize that these results do not occur in a policy-less world. Indeed, economists are quite aware that different immigration policies may lead to different immigrant selection. This issue has been addressed by Miller (1999) or Antecol et al (2003) among others. For example, Antecol et al. (2003) compare the selection of immigrants into the United States, Canada and Australia by estimating a version of equation (2) for immigrants in each country using education and earnings as the response variable. Their results suggest that after excluding Latin American immigrants, selection is not that different across these three countries, indicating that policy plays a small role in the selection of immigrants and that selection is predominantly driven by cultural and historical factors.

One final note must be made in the literature that explores the assimilation of immigrants, it is important to recognize that the analysis of outcomes of women differs to the analysis of outcomes of men. Noting the importance of separating outcomes between women and men, Blau and Kahn (2010) study the labor supply assimilation of immigrant women in the United States, and Antecol (2000) explores how labor market performance of immigrant women is determined by source country characteristics. Similarly, Lozano and Lopez (forthcoming) note that border

enforcement policy in the United States generates a different selection of undocumented immigrant women than men.

### *2.3 Estimating the effects of Immigration in Local Labor Markets*

Next we explore what happens to the earnings of natives (and their welfare in general) when immigrants arrive into the local labor market. While some theoretical models argue that an increase in the supply of workers will decrease wages because the demand for labor is downward sloping (Borjas 2003), the empirical evidence is mixed. OLS models in this vein of the literature use cross-sections of data, to evaluate whether local labor markets that receive a higher number of immigrants have depressed wages for native-born workers. There are two problems with this straight-forward approach: first, mobility of native workers implies native workers will move out of areas with lower wages, hence muting the estimated effect. The second potential problem is that immigrants will self-select into labor markets where the demand for their skills is highest. Immigrant self-selection into these labor markets will result in a downward bias of the true effect of immigrants on natives' earnings.

A first generation of papers (see Goldin, 1994 or Lalonde and Topel, 1991) uses two cross sections to look at changes in earnings of natives as a result of changes in the proportion of foreign-born workers. This methodology assumes that the idiosyncratic characteristics which may attract immigrants to each labor market are time invariant. Yet this approach cannot correct for any bias due to outflows in a local market when immigrants arrive. Figure 2 shows the change in the flow of immigrants in American cities between 2000 and 2007, and the change in natives' log hourly earnings during the same period. The association between these two variables is statistically almost zero, which suggests the fact that immigrants may self-select into markets

where their earnings are higher. To address this potential endogeneity problem economists have proposed using instrumental variable methods, which will be discussed in section four below.

### **Section 3: Difference-in-Difference Estimation**

Having explored the three broad themes in the economic literature on international immigration in our overview of OLS methods, we now turn our attention specifically to other empirical tools used by economists that are perhaps lesser known in other disciplines' study of immigration. After OLS regression, one of the more popular econometric tools in the economics literature on immigration is the difference-in-difference (DD) estimator. In fact, the DD estimator itself can be obtained through OLS methods, as will be discussed below.<sup>4</sup>

The DD strategy is not completely unique to the field of economics, for instance in psychology the technique is called “non-equivalent control-group pretest-posttest design” (Campbell 1969). The popularity of the DD technique stems from its intuitive simplicity and the fact it can be employed using data from either panel-data or from repeated cross sections.

Following the logic of a randomized experimental design, the DD technique attempts to determine the effect of an intervention by comparing pre- and post-intervention group averages of a treatment group with a control group. However, unlike a true randomized experiment, the DD estimator must rely on an external shock that affects the treatment group and choose an appropriate control group not affected by the external shock. Because the DD estimator lacks true random assignment of observations into the treatment and control groups, it is referred to as a “quasi-experimental or “natural experiment” design.<sup>5</sup>

---

<sup>4</sup> See the chapter by Funkhouser in this Handbook for further discussion.

<sup>5</sup> See the chapter by McKenzie and Yang in this Handbook for further discussion of experimental studies.

To explore the DD strategy, we use the Mariel Boatlift example from Card (1990). If we are interested in the effect on natives' unemployment of an influx of immigrants, economists are seldom able to adjust immigration policy by region in randomized experiments. Instead, we must rely on real-world policy changes that provide discrete changes in immigration policy. One potential example began on April 20<sup>th</sup>, 1980 when the Cuban government allowed Cuban citizens who wanted to emigrate to leave out of the port of Mariel, Cuba. Over the next six months roughly 125,000 Cubans entered the United States through Florida. Of the immigrants, nearly half settled permanently in Miami, raising the labor force there by seven percent (Card 1990).

To explore the effect of this large influx of immigrants on the unemployment rate of natives, we might first begin by comparing the unemployment rates in Miami before and after the Mariel Boatlift,

$$(U_{M,1981} - U_{M,1979}) \tag{4}$$

Where  $U$  stands for the unemployment rate, the first subscript denotes the city (Miami), and the second subscript denotes the year. Before we ascribe the entire change in unemployment to the influx of immigrants associated with the Mariel Boatlift, we must consider if there were other factors that might have affected the unemployment rate in Miami during this time. Certainly there were, as the US was then in the midst of a recession.

To adjust for other factors that would have led to a change in unemployment in Miami, Card derives the change in unemployment in four comparison cities, Atlanta, Los Angeles, Houston and Tampa-St. Petersburg,

$$(U_{M,1981} - U_{M,1979}) - (U_{C,1981} - U_{C,1979}) \tag{5}$$

By taking the difference between the Miami difference and the difference in the control cities (C), the difference in differences, the DD estimator seeks to determine the effect on unemployment of the influx of immigrants in Miami, exclusive of other changes happening over time. Figure 3 shows the effect that the Mariel Boatlift may have had on Miami's labor market. This figure shows the employment rate of Black males from January 1978 to December 1984, and Miami is represented with the squared markers. The data in the graph suggests that employment of black males in Miami fell compared to employment of black males in the control cities, and these differences only rose after the Mariel boatlift.

This conceptual idea can be expressed in an OLS framework as:

$$\ln w_{it} = \beta_0 + \beta_1 T + \beta_2 M + \beta_3 (M * T) + \varepsilon_{it} \quad (6)$$

Where  $T$  represents the year term(s) for the period after the intervention, in this case 1981,  $M$  again represents Miami, and  $(M * T)$  represents the interaction of the period after the intervention in Miami.

If we want to control for changes in the underlying distribution of characteristics between control and treatment cities over time, we can add a vector of observable characteristics,  $x_{it}$ , to the regression<sup>6</sup>

$$\ln w_{it} = \beta_0 + \beta_1 T + \beta_2 M + \beta_3 (M * T) + x'_{it} \beta_4 + \varepsilon_{it} \quad (7)$$

The critical assumption of the DD estimator is that conditional on the controlled characteristics, the unemployment rates for Miami and the control cities would have followed parallel paths over time. Therefore studies that use DD estimators must explore the trends in the outcome variable of interest between the treatment and control groups before the treatment date to provide evidence that the chosen control group is appropriate. This is done by exploring the

---

<sup>6</sup> Meyer 1995 notes that if the treatment has an asymmetric affect on differential groups in the population, the OLS design may need to account for this by including an interaction term which interacts each included characteristic with the treatment effect.

time series data before the treatment date to insure that conditional on the other characteristics, the outcome variable of interest changes equally in the treatment and control groups over time.

The choice of an appropriate control group is of critical importance to immigration research utilizing the DD strategy. To insure a proper control group, a researcher using the DD estimation strategy must be careful that the “exogenous” treatment shock was not precipitated or caused by unique changes in the treatment group that were not mirrored in the control group (Ashenfelter and Card 1985). For instance, if a majority of Mariel immigrants choose to relocate to Miami because of uncharacteristically low unemployment rates in 1979 relative to the other control cities, the DD estimator would be biased.

A second major consideration in DD estimators is correct estimation of the standard errors for the interaction term. Bertrand, Duflo and Mullainathan (2004) show consistent estimation of the standard errors for the OLS regression (7) must correct for serial-correlation of the outcome variables over time for the treatment and control groups. They show that failure to properly account for serial-correlation can result in 45 percent of pseudo interventions being statistically significant at the 5 percent level. Incorrectly estimated standard errors will lead to misinterpretation of the statistical significance of results and incorrect inference of the immigration topic in question.

Angrist and Krueger (1999) provide an example of the problem of incorrectly matched control groups and improperly measured standard errors in a follow-up study to Card’s Mariel Boatlift example. They track the changes in the unemployment rate in Miami and the four same control cities from 1993 to 1995. In 1994 a large group of potential Cuban immigrants were redirected from their intended destination of Miami to the US military base at Guantanamo Bay, Cuba. Since the immigrants did not arrive in Miami, they should not have affected the labor



market conditions there. However, Angrist and Krueger show that if researchers would have naively utilized a DD strategy on the 1993-1995 unemployment rates in the same five cities, they would have incorrectly found a statistically significant increase in the unemployment rate for blacks in Miami as a result of “The Mariel Boatlift that Did not Happen.” This example serves as a warning to researchers of the importance of choosing the correct control cities and appropriately adjusting standard errors to adequately account for serial-correlation in the observations.

The Difference-in-Difference model is a powerful tool for immigration economist researchers. The conventional DD estimator presented above can be expanded to include multiple time periods and control groups or combinations of control groups (synthetic control groups) (see Abadie, Diamond and Hainmueller 2007). The strategy can also be generalized to include semi-parametric approaches (see Abadie 2005 and Athey and Imbens 2006). While DD strategies are not true randomized experiments, and are hence subject to the weaknesses inherent in natural experiments (Meyer 1995), they still represent a significant contribution for empirical immigration research.

#### **Section 4: Instrumental Variables Techniques**

A second technique used by economists to simulate the causal interpretations possible with randomized experiments is instrumental variables (IV) regression. In several areas of international immigration research, it is unlikely that standard regressions will give correct estimates of the causal effect of an explanatory variable on the response variable because the assumptions of OLS regressions cannot be met. Using IV regression techniques, economists attempt to overcome the bias in OLS estimates by isolating a source of exogenous variation in

order to estimate an explanatory variable's true effect on the dependent variable of interest. The IV technique utilizes a third variable called the instrument which is correlated with variation in the explanatory variable, but has no direct mechanism (except for through the explanatory variable) to affect the response variable. While IV regressions are quite powerful and allow economists to address a much wider array of questions regarding immigration, difficulties in finding suitable instruments that can satisfy the significant restrictions of IV often limit application of this econometric tool.

Turning back to the question of the effect that immigrants have on the wages of natives, we use another article by Card (2001) to explain IV regression methods. We are interested in the effect that an increase in the share of immigrants in an area has on the wages of natives in that area. Economic theory predicts that an increase in the number of new immigrants in a local labor market will tend to depress wages if immigrants and natives compete for the same jobs, but can raise the wages of natives if immigrants take jobs that help make natives more productive.<sup>7</sup> So which is true, does international immigration lower or raise the wages of natives?

To explore this question, we might first attempt an OLS regression that looks at changes in average wages in areas of the country that have seen a large increase in new immigrants. We can model this as:

$$\ln \bar{w}_{jc} = x_{jc}'\theta + \alpha_{OLS} s_{jc} + \varepsilon_{jc} \quad (8)$$

Where  $\bar{w}_{jc}$  is the average wage in occupation  $j$  and city  $c$ ,  $s_{jc}$  is log population share of new immigrants in the city working in the occupation,  $x_{jc}$  is a vector of explanatory characteristics and  $\varepsilon_{jc}$  is an error term with the usual properties. The estimated coefficient  $\hat{\alpha}_1$

---

<sup>7</sup> One way immigrants may raise the wages of natives suggested in Peri and Sparber (2009), is that immigrants may take jobs that require a high degree of manual skills, like construction workers, which will increase the need for jobs with more management and complex skills, like construction supervisors.

would be an estimate of the effect of an increase in the log share of immigrants on averages wages in occupations in the cities.

While OLS analysis would be straight forward in this case, it is unlikely to give unbiased estimates of the true parameter  $\alpha$ . New immigrants likely do not randomly move to cities, nor do they randomly choose occupations once they arrive there. Instead, recent shocks that cause high wages in a particular city or occupation will likely induce new immigrants to take jobs in that area. This omitted variable, recent demand shocks, leads to bias in the standard regression estimates because it violates a necessary assumption of OLS, namely that  $E(\varepsilon | s) = 0$ ; all things equal, we would expect areas with positive wage shocks to pull a larger fraction of new immigrants.

Card uses an IV strategy to address the problem. In order to approximate the random experiment of adding more new immigrants to a particular area, he suggests as an instrument a variable that likely influences the share of new immigrants in an area, but has no direct effect on current local demand shocks to wages. Noting that new immigrants often tend to migrate to ethnic enclaves established by previous immigrants, Card suggests using the fraction of earlier immigrants from the same source country to an occupation in a given city as an instrument for new immigrants.

To employ the IV regression to eliminate the omitted variable bias, Card first creates an estimate for the share of new immigrants expected in each local area based only on the experience of previous immigrants:

$$s_{jc} = \eta_1 z_{jc} + \varepsilon_{jc} \tag{9}$$

In this first stage, the distribution of previous immigrants across occupations and cities,  $z_{jc}$ , is used to predict how current immigrants would distribute across the same occupations and cities. While the decisions of previous immigrants were likely based upon previous demand shocks to wages, it is unlikely that those decisions are correlated with current demand shocks. Therefore a prediction of how current immigrants will distribute across areas based only on the decisions of past immigrants will be free of the bias from the omitted variable (current demand shocks).

In the second stage of the IV process, Card takes the estimated values for the log share of immigrants, and uses those values instead of the actual values to estimate:

$$\ln \bar{w}_{jc} = x_{jc}^I \theta + \alpha_{IV} \hat{s}_{jc} + \varepsilon_{jc} \quad (10)$$

Where  $\hat{s}_{jc}$  is the log share of new immigrants in occupation j and city c predicted using OLS estimates from equation 9. The difference between the actual log share of new immigrant,  $s_{jc}$ , and the predicted share,  $\hat{s}_{jc}$ , is that the actual share reflects decisions new immigrants made in response to recent local demand shocks, whereas the predicted share excludes any influence from this effect. Hence OLS regressions of equation 10 using predicted values will not have the same omitted variable bias as the standard OLS regression on actual values (equation 8). Using this IV approach, Card finds that immigration by new immigrants decreased wages and employment rates of natives in traditionally high immigrant cities by 1-3 percent.

Utilizing IV regression techniques places high requirements on the proposed instrument. If the instrument is not highly correlated with the explanatory variable, IV estimates will be unreliable. To avoid this “Weak Instruments” problem usually requires a high t-statistics for proposed instrument(s) in the first stage regression which predicts the values for the endogenous explanatory variable (i.e. equation 8) (Staiger and Stock 1997). A second consideration in IV

regressions is called the Local Average Treatment Effect (LATE). For an instrument to be useful in estimating the true effect of the explanatory variable on the response variable, the subpopulation affected by the instrument cannot be dramatically dissimilar from the entire population; or else the IV estimate will not be true of the entire population, but only provide an estimate for the sub, or “local”, population. For instance, if new immigrants who settle in areas based on ethnic networks are radically different than all other immigrants, the Card estimate will not provide the true effect for an increase in the share of all new immigrants, but will instead only provide a LATE estimate for the effect of an increase in the subpopulation of immigrants affected by the instrument (Imbens and Angrist 1994). Finally IV regressions require relatively large samples for the estimates to be unbiased and hence are not appropriate for smaller data sets.

Instrumental Variables regressions allow estimation of the relationship between an explanatory variable and a response variable even when unobserved factors prohibit using standard OLS regression techniques. Given the power of IV techniques, there has been a number of econometric innovations which will likely become increasingly common in economic analysis of international immigration research. Limited Information Maximum Likelihood (LIML) and Control Function techniques are alternative methods for dealing with endogenous explanatory variables which are closely related to IV methods. LIML analysis tends to perform better in situations with multiple instruments, and Control Function techniques tend to be more precise, while less robust, than IV methods (Wooldridge 2002). While finding appropriate instruments remains difficult, the power of IV and related methods mean they will continue to be a popular tool in the economics of international immigration research.

## **Section 5: Recent Developments in Research Methods**

Given the increasing levels of international migration in the past two decades, we predict that economic research of immigration will only continue to grow. While future analysis will continue to use ordinary least squares, difference-in-difference and instrumental variable methods, the literature is beginning to incorporate new econometric methods into the field. These new methods expand the research focus onto distributional analysis of the influence of international immigration.

The methods presented in the first sections of this chapter have focused on analyses of the mean. Researchers have assessed questions such as the average rate of wage assimilation of immigrants, or the average effect of immigrants on local labor markets. But are rates of wage assimilation different for low wage and high wage immigrants? Do immigrants have the same influence across the distribution of wages in the local labor market as they do on the average wage? To answer these research questions, economists have begun to use new econometric tools which shift the focus away from mean differences and instead focus on differences across the entire distribution. Most prominent among these new methods are quantile regressions (Koenker and Hallock 2001) and DiNardo, Fortin and Lemieux (1996) reweighting analyses. These new methods represent a substantive addition to the economics of immigration literature, and we believe they will become increasingly common in the field as researchers seek to explore differences in the relationship between immigrants and labor market outcomes at various points along the distribution of earnings.

### 5.1: Quantile Regression

Since Chiswick's seminal work in 1978, economic research on immigration has focused on analysis at the mean of the distribution. Quantile regression is very similar to OLS regression methods in concept, but instead of focusing on associations at the mean, quantile regression presents associations at a specified quantile of the conditional dependent variable. Quantile regressions are essential when the relationship between the explanatory and response variables is not constant across the distribution.

To see the importance of quantile analysis we present the male native-immigrant wage gap from the 2000 US Census at each decile of wages in Figure 4. The figure is constructed by taking the difference between the given decile of log wage in the immigrant distribution minus the same decile of log wage in the native distribution. Whereas the average immigrant wage gap is -0.14, the wage gap actually goes from a low of -0.23 at the median to 0.04 in the 9<sup>th</sup> decile of wages. With such a large difference in the immigrant wage gap along the distribution, it is natural to ask if a given characteristic affects immigrants equally at all points in the wage distribution. For instance, we might wish to know if the change in the median wage for an extra year of education is the same as the change in the average wage for another year of schooling.

Referring back to equation 1, which relates the wage of immigrants on a variety of variables, we remember that OLS regression coefficients denote the change in the response variable (wage) for a one unit change in the explanatory variable (for instance, years of education).

$$\ln w_i = x_i \theta + \gamma_1 I_i + \gamma_2 I_i \times Y_i + \gamma_3 I_i \times Y_i^2 + \varepsilon_i \quad (1)$$

Quantile regression is intuitively very similar to the standard OLS regression, but instead of estimating the conditional mean of the response variable, quantile regressions estimate the conditional specified quantile of the response variable.

In this section we follow Chiswick, Le and Miller (2008), who define  $Q_\theta(w|x)$  as the  $\theta^{th}$  quantile of the conditional log wage distribution given the vector of characteristics  $x$  ( $0 < \theta < 1$ ). Quantile regression seeks to estimate a vector of coefficients  $\beta(\theta)$  such that

$$Q_\theta(w|x) = x'\beta(\theta) \quad (11)$$

Quantile regression estimates  $\beta(\theta)$  by minimizing the weighted sum of the absolute value of the errors between the estimated conditional quantile function and the true log wage value. See Koenker and Hallock (2001) for more on estimation of quantile regression functions.

For example, the coefficient from a 0.5 quantile regression can be used to estimate the change in the median of the conditional wage distribution for immigrants for a one unit change in the education variable. Conceptually, the OLS coefficient is estimated by fitting a best fit line through the conditional means of log wage at each year of education. Similarly, a 0.5 ( $\theta$ ) quantile regression coefficient is estimated by fitting a best fit line through the conditional medians (or other quantiles) of log wage at each year of education. If the conditional medians increase at the same slope as the conditional means, the estimated coefficients of the quantile regression and the OLS regression will be the same. If the conditional medians rise at a different slope, the coefficients will not be the same, and quantile analysis has informed us of a potentially interesting deviance in the distributional analysis of our variable of interest.

In addition to addressing the problem of outliers, the real power of quantile regression for immigration researchers is that it allows analysis of multiple points along the wage distribution. Researchers can evaluate the relationship between the dependent variable and independent



variables across the distribution of the dependent variable. An example from Chiswick, Le and Miller (2008) illustrates the usefulness of quantile analysis for researchers. They compare the quantile regression coefficient estimates for each decile with the OLS coefficients in assessing how immigrants compare to natives across the distribution of earning. They use the 2000 US Census to compare the log hourly wages of adult men. They find that while immigrants from non-English-speaking countries earn less than natives on average, using quantile analysis and controlling for differences in observable characteristics, immigrants from non-English speaking countries in the bottom decile actually have a wage advantage relative to their native counterparts.<sup>8</sup>

Quantile regression analysis can be used to look at the same research questions as OLS analysis, and informs a more nuanced understanding of the relationship between the variables of interest. The OLS results are certainly a useful benchmark for understanding the relationship between the dependent and independent variables of interest. By further comparing the OLS conditional mean function results with the 0.1, 0.5(median) and 0.9 conditional quantile function results, researchers can evaluate if the relationship is constant over the distribution of the response variable. Further, unlike OLS regression focused on the mean, quantile regressions are also robust to outlier log wage values.

## *5.2: DiNardo, Fortin, Lemieux (1996) Reweighting Analysis*

A second recent econometric method that allows researchers to shift their focus beyond the mean and on other moments of the distribution is the DiNardo, Fortin, Lemieux (1996)

---

<sup>8</sup> To be precise, Chiswick, Le and Miller (2008) use a technique called quantile decomposition to obtain their results. Quantile decomposition, as introduced in Machado and Mata (2005), uses hundreds of vectors of quantile regression coefficient estimates to obtain a synthetic distribution of wages if the immigrant population was paid according to the same conditional wage functions as natives.

(henceforth DFL) reweighting technique. Like quantile regression analysis, the DFL reweighting method allows researchers to explore the relationship of the explanatory and response variables at multiple points along the distribution. Using DFL reweighting methods, economists researching immigration can answer such questions as ‘How has the widening of the distribution of wages in the US and changes in source country of birth of recent immigrants affected the immigrant wage gap at median wages?’ (Butcher and DiNardo 2002), or ‘If Mexican immigrants in the US were paid according to the going rates they would receive in Mexico, where would they be in the distribution of Mexican wages?’ (Chiquiar and Hanson 2005). Analysis using DFL methods allow researchers to explore in greater detail the selection of immigrants across the distribution of skills in the source country, compare immigrant assimilation at wages above and below the median, and gain greater specificity on the effect of immigration on local labor markets.

The power of the DFL technique comes not only from its usefulness in distributional analysis, but also because it is semi-parametric and imposes fewer restrictive assumptions than the other methods discussed in this chapter. OLS, Difference-in-Difference, Instrumental Variables and Quantile Regression analyses all impose a polynomial functional form on the equation of interest. That is, they must estimate a linear relationship between an explanatory variable (including characteristics raised to a power) and the response variable. The DFL reweighting technique is advantageous over these methods because it does not need to estimate a functional form for the equation of interest. This allows the researchers to avoid biases if the explanatory and response variables of interest do not have a linear relationship. The DFL technique functions by reweighting one sample so that it has the same distribution of observable characteristics as a second sample. This is a powerful tool if a researcher wants to analyze how

much of differences between groups can be explained by differences in other observable characteristics.

The DFL method allows Butcher and DiNardo (2002) to explore the increase in the immigrant/native-born wage gap since 1970 without estimating a wage function. In 1970, 22 percent of immigrant men to the US were Hispanic. By 1990, this number had risen to 45 percent. So, compared to 1970, male Hispanic immigrants were relatively more common in 1990, and immigrant men from other areas were relatively less common. To explore how much of the increase in the immigrant-native wage gap from 1970 to 1990 might be attributable to differences in the racial and ethnic composition of immigrants, the DFL technique would decrease the relative proportion of Hispanic immigrants in 1990, and increase the relative proportion of all other immigrants in that year. Specifically, the DFL technique would multiply the sample weight of each 1990 Hispanic immigrant male by 0.49 ( $\approx 22/45$ ) and would multiply the sample weight of each non-Hispanic immigrant male by 1.42 ( $\approx 78/55$ ). The resulting “reweighted” sample of 1990 immigrant men would thus contain 22 percent Hispanic men and 78 percent men of other ethnicities. The wage distribution of this reweighted 1990 sample would be different than the actual 1990 sample. The difference between the reweighted 1990 sample and the actual 1990 sample is completely due to changes in the distribution of immigrant ethnicity over the period. If the only reason for the change in the immigrant/native-born wage gap was because of the change in the ethnicity of immigrants between the two periods, then the reweighted 1990 sample would be identical to the 1970 wage distribution. While Butcher and DiNardo (2002) find a large role for race and ethnicity changes, these variables do not explain all changes in the distribution of wages over the period.

This example is useful to highlight several advantages of the semi-parametric aspect of the DFL reweighting approach. While previous research has focused on mean comparisons, the reweighted 1990 sample can be used to explore the effect of the change in ethnic composition of immigrants on the entire distribution of wages, including the median and each percentile of the distribution. Further, the analysis does not force the researcher to estimate the exact functional form of the relationship between ethnicity and wages in either period. Because the procedure only changes the sample weight of each observation, there is no need to estimate a wage function; each observation keeps their actual wage. The change in the wage structure comes from the change in the relative frequency of observations with particular characteristics, but it is not necessary to estimate the exact relationship between characteristics and the dependent variable of interest. This can be particularly useful when the researcher does not want to impose restrictions on the functional form of the relationship between the independent and dependent variables of interest. For a more detailed explanation of the DFL decomposition see DiNardo, Fortin, and Lemieux (1996).

## **Section 6: Summary**

In this chapter we have highlighted the main empirical tools used by Economists when studying international immigration. In particular we highlight four common methodologies that have made important headway in the literature. The fact that economists rely on large data sets and on identifying causal relationships gives us a unique perspective to contribute in the current immigration debate. Because of the relevance of population flows in the foreseeable future we argue that the field will continue to grow in the years to follow. Non-economists should take advantage of recent econometric developments to contribute to the current immigration debate.

## Section 7: Bibliography

- Abadie, Alberto. (2005). "Semiparametric Difference-in-Difference Estimators". *Review of Economic Studies*. Vol. 72. (#1). 1-19.
- Abadie, Alberto, Alexis Diamond, Jens Hainmueller. (2007). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". NBER Working Papers 12831, National Bureau of Economic Research, Inc.
- Alfey, Susan, Guido Imbens. (2006). "Identification and Inference in Nonlinear Difference-in-Differences Models". *Econometrica*. Vol. 74. (Mar, #2). 431-497.
- Angrist, Joshua D., Alan B. Krueger. (1999). "Empirical Strategies in Labor Economics". *Handbook of Labor Economics*. Vol. 3. Ed. 1. 1277-1366.
- Antecol, Heather. (2000). "An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates." *Labour Economics*. Vol. 4. (July, #4). (409-426)
- Antecol, Heather, Deborah A. Cobb-Clark, Stephen J. Trejo. (2003). "Immigration Policy and the Skills of Immigrants to Australia, Canada, and the United States". *The Journal of Human Resources*. Vol. 38. (Winter, #1). 192-218.
- Antecol, Heather, Peter Kuhn, Stephen J. Trejo. (2006). "Assimilation via Prices or Quantities? Sources of Immigrant Earnings Growth in Australia, Canada, and the United States". *The Journal of Human Resources*. Vol. 41. (Fall, #4). 821-840.
- Ashenfelter, Orley, David Card. (1986). "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs". *The Review of Economics and Statistics*. Vol. 67. (November, #4). 648-660.
- Bertrand, Marianne, Esther Duflo, Sendhil Mullainathan. (2004). "How Much Should we Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*. Vol. 119. (Feb, #1). 249-275.
- Borjas, George J. (1985) "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics*. Vol. 3. (Oct, #4). 463-489.
- Borjas, George J. (1987) "Self-Selection and the Earnings of Immigrants," *American Economic Review*. Vol. 77. (Sep, #4). 531-553.
- Borjas, George J. (1995). "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?" *Journal of Labor Economics* Vol. 13. (April, #2). 201-245.

- Borjas, George J. (2003). "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market". *The Quarterly Journal of Economics*. Vol. 118. (Nov, #4). 1335-1347.
- Butcher, Kristin F., John DiNardo. (2002). "The Immigrant and Native-born Wage Distributions: Evidence from United States Censuses". *Industrial and Labor Relations Review*. Vol. 56. (Oct, #1). 97-121.
- Capbell, D. T. (1969). "Reforms as Experiments." *American Psychologist*. Vol. 24. 409-429.
- Card, David. (2001). "Immigrant Inflows, Native Outflows and the Local Labor Market Impacts of Higher Immigration". *Journal of Labor Economics*. Vol. 19. (Jan, #1). 22-64.
- Card, David. (1990). "The Impact of the Mariel Boatlift on the Miami Labor Market". *Industrial and Labor Relations Review* Vol. 43. (Jan, #2). 245-257.
- Chiquiar, Daniel, Gordon H. Hanson. (2002). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States". *Journal of Political Economy*. Vol. 113. (April, #2). 239-281.
- Chiswick, Barry R. (1978) The Effect of Americanization on the Earnings of Foreign-born Men *The Journal of Political Economy*, Vol. 86, No. 5 (Oct., 1978), pp. 897-921
- Chiswick, Barry R., Anh T. Le, Paul W. Miller. (2008). "How Immigrants Fare across the Earnings Distribution in Australia and the United States. *Industrial and Labor Relations Review*. Vol. 61. (April, #3). 353-373.
- Cobb-Clark, Deborah A. (1993), "Immigrant Selectivity and Wages: The Evidence for Women," *American Economic Review*, Vol. 83. (Sep, #4): 986-993.
- DiNardo, John, Nicole M. Fortin, Thomas Lemieux. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*. Vol. 64. (September, #5). 1001-1044.
- Goldin, Claudia (1994), "The Political Economy of Immigration Restriction in the United States, 1890-1921." In Goldin, C., and G. Libecap, eds., *The Regulated Economy: A Historical Approach to Political Economy*. Chicago: University of Chicago Press. 223-57.
- Imbens, Guido W., Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects". *Econometrica*. Vol. 62. (Mar, #2). 467-475.
- Koenker, Roger, Kevin F. Hallock. (2001). "Quantile Regression". *Journal of Economic Perspectives*. Vol. 15. (Fall, #4). 143-156.

- LaLonde, Robert, and Robert Topel (1991) "Labor Market Adjustments to Increased Immigration". In Abowd, J., and R. Freeman, eds., *Immigration, Trade and the Labor Market*. Chicago: University of Chicago Press, 1991. 167-200.
- Lopez, Mary, Fernando Antonio Lozano. (2009). "The Labor Supply of Immigrants in the United States: The Role of Changing Source Country Characteristics". *American Economic Review*. Vol. 99. (May, #2). 35-40.
- Lozano, Fernando and Todd Sorensen (2010) "Mexican Immigrants, the Labor Market Assimilation and the Current Population Survey: The Sensitivity of Results across Seeming Equivalent Surveys" *International Migration (forthcoming)*.
- Lubotsky, Darren. (2007). "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings". *The Journal of Political Economy*. Vol. 115. (Oct, #5). 820-867.
- Meyer, Bruce D. (1995). "Lessons from the U.S. Unemployment Insurance Experiments". *Journal of Economic Literature*. Vol. 33. (Mar, #1). 91-131.
- Miller, Paul W. (1999). "Immigration Policy and Immigrant Quality: The Australian Points System". *American Economic Review*. Vol. 89. (May, #2). 192-197.
- Munshi, Kaivan. (2003). "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market". *The Quarterly Journal of Economics*. Vol. 118. (May, #2). 549-599.
- Peri, Giovanni and Chad Sparber (2009). "Task Specialization, Immigration, and Wages". *American Economic Journal: Applied Economics*. Vol. 1. (3). 135-169.
- Staiger, Douglas, James H. Stock. (1997). "Instrumental Variables Regression with Weak Instruments". *Econometrica*. Vol. 65. (May, #3). 557-586.
- Trejo, Stephen J. (2003). "Intergenerational Progress of Mexican-Origin Workers in the U.S. Labor Market". *The Journal of Human Resources*. Vol. 38. (Summer, #3). 467-489.

Table 1. Research in Economics of Immigration

	(1)	(2)	(3)	(4)
	1990-1994	1995-1999	2000-2004	2005-2009
Number of Papers	23	31	40	51
by Empirical Technique				
Ordinary Least Squares	11	14	11	20
MLE, Probit, Logit, Tobit	3	3	9	9
Difference-in-Difference	3	1	2	0
Instrumental Variables	2	4	12	8
Other	4	9	6	14
By Topic				
Assimilation	10	14	17	14
Immigrant Selection	6	6	7	8
Natives Outcomes	5	8	9	17
Other	2	3	7	12

Note: Papers on immigration in the following journals: American Economic Review, Quarterly Journal of Economics, Journal of Political Economy, Review of Economic Studies, Economic Journal, Review of Economic and Statistics, Journal of Labor Economics, Journal of Human Resources, Industrial and Labor Relations Review.



Figure 1. Native Immigrant Differences in Predicted Earnings Differences between Estimates with and without Cohort Controls

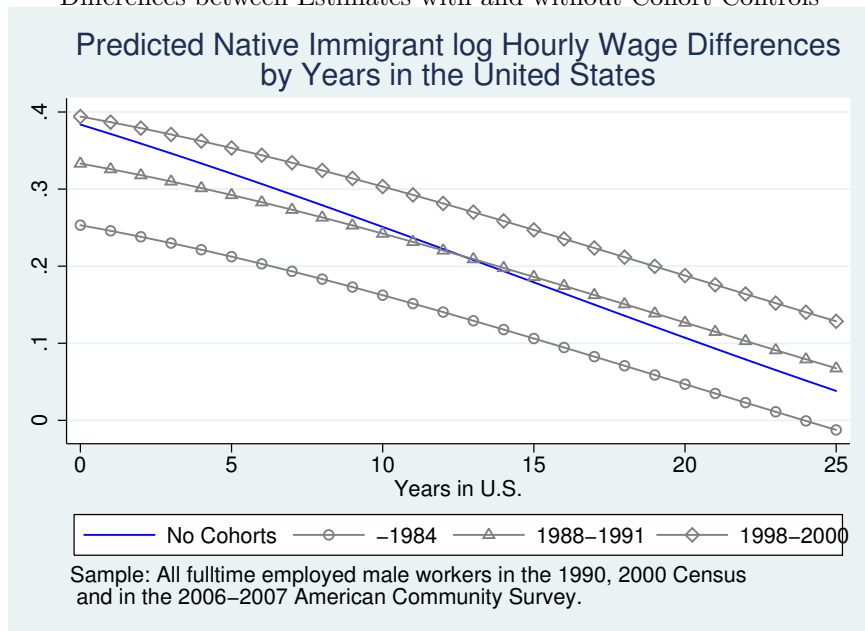


Figure 2. Immigrants and the Earnings of Natives

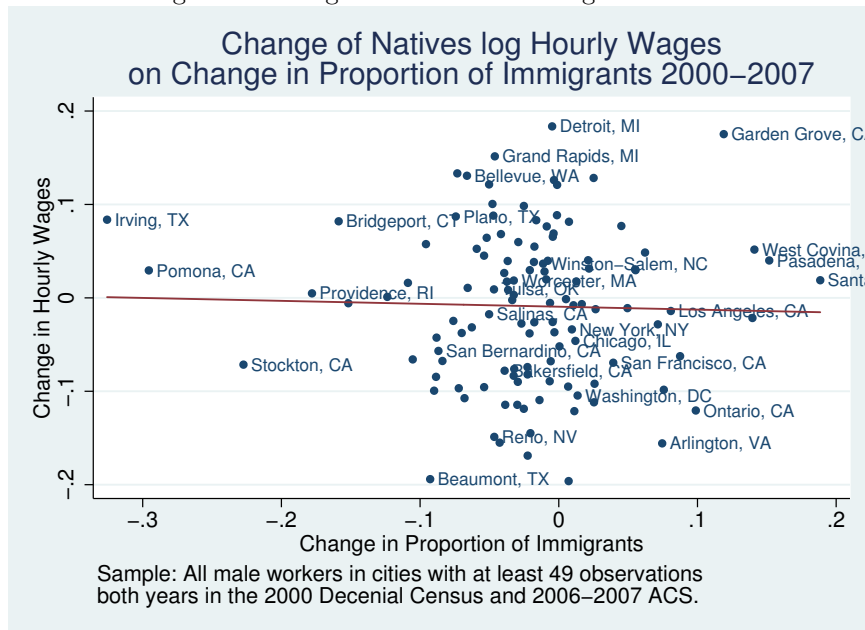


Figure 3. The Mariel Boatlift

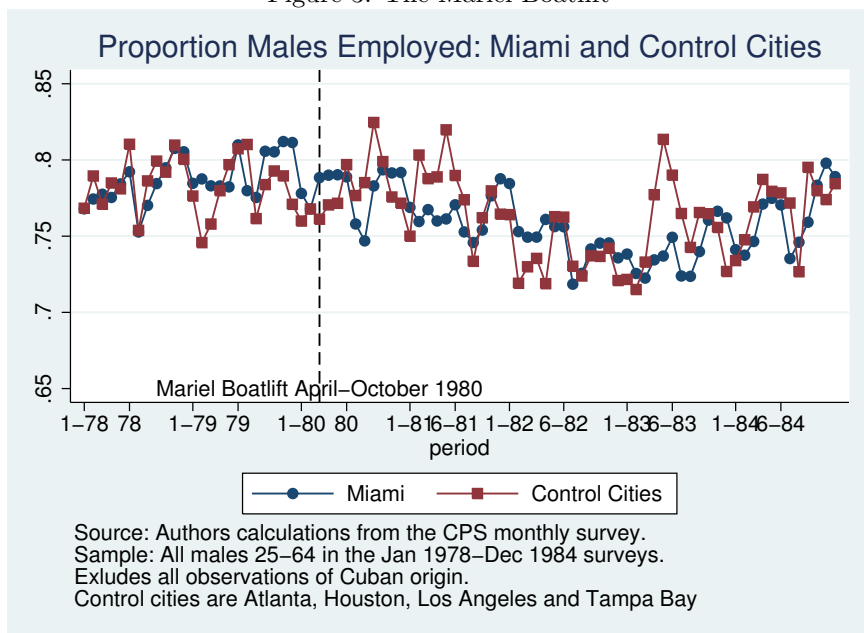


Figure 4. Immigrant-Natives Log Hourly Wage Differences

