

IZA DP No. 6437

The Short- and Long-Run Determinants of Less-Educated Immigration into U.S. States

Nicole B. Simpson
Chad Sparber

March 2012

The Short- and Long-Run Determinants of Less-Educated Immigration into U.S. States

Nicole B. Simpson

*Colgate University
and IZA*

Chad Sparber

Colgate University

Discussion Paper No. 6437

March 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: 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

The Short- and Long-Run Determinants of Less-Educated Immigration into U.S. States^{*}

This paper uses a gravity model of migration to analyze how income differentials affect the flow of immigrants into U.S. states using annual data from the American Community Survey. We add to existing literature by decomposing income differentials into short- and long-term components and by focusing on newly arrived less-educated immigrants between 2000-2009. Our sample is unique in that 95 percent of our observed immigrant flows equal zero. We accommodate for the zeros by using scaled ordinary least squares, a threshold tobit model from Eaton and Tamura (1994), and the two-part model to analyze the determinants of immigration. Models that include observations with zero flow values find that recent male immigrants respond to differences in (short-term) GDP fluctuations between origin countries and U.S. states, and perhaps to (long-term) trend GDP differences as well. More specifically, short-run GDP fluctuations pull less-educated male immigrants into certain U.S. states, whereas GDP trends push less-educated male immigrants out of their countries of origin. Effects for less-educated women are less robust, as GDP coefficient magnitudes tend to be much smaller than in regressions for men.

JEL Classification: J61, E01

Keywords: immigration, macroeconomics, GDP, gravity

Corresponding author:

Nicole B. Simpson
Colgate University
Department of Economics
13 Oak Drive
Hamilton, NY 13346
USA
E-mail: nsimpson@colgate.edu

^{*} The authors thank Piyush Chandra, Giovanni Perri, Cong S Pham, Todd Sorensen, and seminar participants at Colgate University, Union College, and the 2010 WEAI in Seattle for detailed comments. All errors are our own.

1 Introduction

Income is often cited as an important determinant of immigration, and some measure of income in the origin and/or destination country is included in almost every model explaining international migration. Recently, Clark, Hatton, Williamson (2007), Lewer and Van den Berg (2008), Lewer, Pacheco and Rossouw (2009), Mayda (2010), and Ortega and Peri (2009) all find evidence that per capita GDP (in the origin and/or destination country) is a significant predictor of cross-country immigrant flows. However, none of this work focuses on newly arrived U.S. immigrants with a high-school degree or less education. We add to this literature in three ways: (1) by analyzing recent inflows of less-educated immigrants into U.S. states between 2000-2009; (2) by decomposing GDP into short- and long-run components; and (3) by employing three distinct estimation methodologies: scaled ordinary least squares (SOLS), Eaton and Tamura, and two-part models.

First, we analyze the flow of new immigrants into U.S. states between 2000-2009 using U.S. Census and American Community Survey (ACS) data. Our work complements literature that focuses on the locational choice of new immigrants based on state-specific factors (for example, Bartel, 1989; Zavodny, 1997; Borjas, 1999, Dodson, 2001). These papers often explore the demographic characteristics of immigrants as potential determinants of their selected destination in the U.S. Instead of analyzing individual decisions, we take a macro approach to estimate how U.S. immigrant flows respond to state-specific factors. We focus on the flow of newly arrived less-educated immigrants – that is, immigrants with a high school degree or less who legally or illegally arrived to the U.S. within the past year (from the survey date). We focus on this group since the immigration debate in the U.S. is especially contentious about them. According to Mayda (2006) and Hainmueller and Hiscox (2010), less-educated immigrants likely generate the largest negative political reactions by the U.S. public. Additionally, less-educated immigrants exhibit more volatility in employment than both their native counterparts and other well-educated immigrants (Orrenius and Zavodny, 2009). Our analysis concentrates mainly on men, since past evidence has argued that male migration decisions are more likely to be motivated by economic factors, while females more likely migrate for tied or associational reasons (Taylor, 2006) – a finding supported by our results. Moreover, the male labor market is especially interesting in the wake of the 2007-2009 recession, when male unemployment rates were particularly high (Şahin, Song and Hobijn, 2010).

Second, variation across countries and U.S. states allows us to consider if differences in short-run GDP (i.e., fluctuations) and long-run GDP (i.e., trends) have distinct effects on gross immigrant flows. Surprisingly, there is little work that analyzes the response of im-

migrant *flows* to macroeconomic cycles (exceptions include Hanson and Spilimbergo (1999), Davis and Haltiwanger (1992), Borger (2008), and Mandelman and Zlate (2010)). In addition, we further disentangle GDP differentials to separately identify push and pull factors, adding to recent work by Warin and Svaton (2008), Zaiceva and Zimmermann (2008), Pedersen, Pytlikova and Smith (2008), and Mayda (2010). This allows us to assess whether less-educated immigrants leave countries that are experiencing short-run downturns (i.e., recessions) or are attracted by states experiencing short-run booms. Similarly, we ask whether U.S. immigrants are pulled into U.S. states with higher income, or are instead being pushed out by persistent poverty in their origin country.

Third, we estimate gravity models of immigration in the spirit of Karemera, Oguledo and Davis (2000), Lewer and Van den Berg (2008), Ortega and Peri (2009), Mayda (2010) and Beine et al. (2011). However, we employ a number of techniques, including the two-part and Eaton and Tamura (1994) threshold tobit models – methods that, to our knowledge, have not yet been used to analyze the determinants of immigration.¹ The use of these models is necessitated by unique features of our data. Specifically, we observe annual bilateral gross flows of less-educated workers into each U.S. state from 112 different source countries, but approximately 95 percent of our sample has an immigration flow value of zero. This presents estimation challenges since the standard gravity model adopts log-flows as the dependent variable. This problem of log-zero values is quite common in other economics fields. We therefore appeal to those literatures for alternative estimation techniques. We first estimate our gravity model using a standard OLS regression in which we add one to each observed flow (i.e., SOLS). Next, we follow the trade literature and apply a threshold tobit model in the spirit of Eaton and Tamura (1994) to account for the zero flows. Lastly, we employ the two-part model – commonly used in health economics – that estimates a probit regression, followed by an OLS specification that drops all observations with zero immigrant flows.

Our results indicate that fluctuations in GDP positively affect male less-educated immigration, but only when the entire sample of immigrant flows are considered. If the observations with zero immigrant flows are dropped, we find no relationship between GDP differentials and less-educated male immigration. Effects from long-run GDP differentials follow a similar pattern, though baseline Eaton and Tamura estimation also fails to find a significant relationship. In subsequent push and pull analysis, however, models that include zero flow values robustly find that long-run GDP trends push less-educated men out their origin countries, and that recent booms in U.S. states attract less-educated men from abroad. Conversely, there are no pull effects from long-run state GDP trends, nor do short-run origin-

¹A growing literature has provided ample support for the Eaton-Tamura technique, including Head and Ries (1998), Rauch and Trindade (2002), and Martin and Pham (2008).

country cyclical fluctuations spur men to emigrate.

We also consider alternative sub-samples of U.S. immigrants to see if certain groups respond similarly to GDP differentials. We briefly discuss how Mexican immigrants, who represent the vast majority of new U.S. immigrants, affect the analysis. Results are robust to the exclusion of Mexico, however Mexican immigrants themselves are not responsive to long- or short-run GDP differentials. Next, we find that the flows of less-educated female immigrants are much less responsive to short-run GDP fluctuations than their male counterparts. Finally, we assess effects for “well-educated” male immigrants (those with at least some college experience). These workers exhibit a response to short-run GDP differences similar to that of less-educated men, but our results for long-run GDP differentials are generally less robust.

The paper is organized as follows. First, we motivate our empirical specification with a simple model and provide a thorough explanation of the estimation techniques. We then summarize some of the important trends regarding recent U.S. immigration, and describe our data in detail. Lastly, we present the results and discuss how they add to existing literature.

2 Empirical Strategy

2.1 Theoretical Motivation

The canonical theoretical model of migration consists of an income maximization problem where a potential immigrant from origin country o chooses the destination d based on the relative returns to migrating after factoring out migration costs. Assume there is a discrete number of origin countries $o = \{1, 2, \dots, O\}$ and a discrete number of destinations $d = \{1, 2, \dots, D\}$.

Following the work of Ortega and Peri (2009), Grogger and Hanson (2011) and Beine et al. (2011), we assume a linear utility function. The utility of an agent from country o who remains in country o is therefore

$$u_{o,o} = Y_o + A_o + \varepsilon_o$$

where Y_o represents income in the origin country, A_o represents country-specific factors (such as amenities, etc.), and ε_o is the extreme-value distribution error term. Immigration researchers use either aggregate measures of income (i.e., GDP) or micro-level measures of income (i.e., wages) to model Y_o . We choose the former since we are assessing how less-educated immigrant flows respond to macroeconomic differences across a large set of destinations; that is, we are not trying to measure the response of immigrants to variations in

the return to skill, for example.²

The utility of an agent from country o who decides to migrate to destination d is

$$u_{d,o} = Y_d + A_d - C_{d,o} + \varepsilon_d$$

where migration costs are denoted by $C_{d,o}$. Migration costs may include costs that are specific to the destination (i.e., immigration restrictions), bilateral costs between the destination and origin country (i.e., language differences), or costs that are individual-specific (i.e., family members left back home).

The agent chooses the destination k that maximizes his/her utility:

$$\max_{k=\{1,\dots,D\}} \{u_{k,o}\} \quad (1)$$

Using this simplified model, the probability that an individual born in country o will move to destination d is then:

$$pr \left(u_{d,o} = \max_k \{u_{k,o}\} \right) = \frac{M_{d,o}}{M_o} \quad (2)$$

where $M_{d,o}$ is the number of immigrants from origin country o in destination d , and M_o is the native population of the origin country o . When the random term follows an iid extreme-value distribution, we can apply the results in McFadden (1984) to deliver:

$$\frac{M_{d,o}}{M_o} = \frac{\exp(u_{d,o})}{\exp(u_{o,o})} \quad (3)$$

or, equivalently:

$$\frac{M_{d,o}}{M_o} = \frac{\exp(Y_d + A_d - C_{d,o})}{\exp(Y_o + A_o)}. \quad (4)$$

Taking natural logarithms of both sides yields:

$$\ln \left(\frac{M_{d,o}}{M_o} \right) = Y_d - Y_o + A_d - A_o - C_{d,o} \quad (5)$$

or equivalently:

$$\ln(M_{d,o}) = Y_d - Y_o - C_{d,o} + A_d - A_o + \ln(M_o) \quad (6)$$

Thus, immigrant flows depend on the aggregate income differential between the destination and origin ($Y_d - Y_o$), moving costs that depend on the destination and origin of the

²For a recent discussion of this issue, we refer the reader to Rosenzweig (2007).

immigrant ($C_{d,o}$), origin- and destination-specific factors (A_d, A_o), and the population of the origin country (M_o). Recall that the focus of this paper is on immigration into U.S. states, hence the set of possible destinations is the 48 contiguous U.S. states. We abstract from internal migration within the origin country and immigration to other destinations outside of the U.S.

2.2 Empirical Specification

Equation 6 motivates the basic empirical specification in Equation 7 – a gravity model of immigration, similar to Karemera, Oguledo and Davis (2000), Mayda (2010), Lewer and Van den Berg (2008), Ortega and Peri (2009), and Beine et. al. (2011).

$$\begin{aligned} \ln(M_{t+1,d,o}) &= \alpha + \beta \cdot (Y_{t,d} - Y_{t,o}) \\ &\quad + \delta \ln(Dist_{d,o}) + FE_d + FE_o \\ &\quad + \gamma \ln(Stock_{t,d,o}) + \eta \ln(Pop_{t,d}) + \mu \ln(Pop_{t,o}) + FE_t + \varepsilon_{t,d,o} \end{aligned} \quad (7)$$

The dependent variable $M_{t+1,d,o}$ measures the flow of immigrants from origin country o to destination d at time $t + 1$. The income differential is measured using time t per capita GDP differentials, $Y_{t,d} - Y_{t,o}$. Notice that we lag the independent variables (by one year) to mitigate endogeneity issues. We follow the literature in identifying control variables that proxy for migration costs. We include the natural log of the distance between the origin country’s capital city and the state’s capital city ($Dist_{d,o}$). Time-variant factors include the natural log of the immigrant stock from country o residing in state d who have been in the U.S. for more than one year ($Stock_{t,d,o}$), the natural log of the state’s population ($Pop_{t,d}$), and the natural log of the origin country population ($Pop_{t,o}$). Year fixed effects (FE_t) account for time trends as well as U.S. immigration policy decided at the national level (see Clark, Hatton, and Williamson (2007) for further discussion of the importance of policy). Destination and origin fixed effects (FE_d and FE_o) account for region-specific factors that do not change over time and imply that all coefficients of interest will be identified by variation within regions over time. The error term is represented by $\varepsilon_{t,d,o}$, and α , β , δ , γ , η and μ are the coefficients to be estimated.

We modify this framework by further decomposing GDP into two components. First, we consider a long-run country-specific GDP trend, $\hat{Y}_{t,c} = \hat{a}_c + \hat{b}_c \cdot T$ for $c = \{o, d\}$ with time trend T . The coefficients \hat{a}_c and \hat{b}_c are obtained by estimating the following country-specific regressions, where $e_{c,t}$ is an error term:

$$Y_{t,c} = a_c + b_c \cdot T + e_{c,t} \quad (8)$$

We compute short-run fluctuations in GDP from its long-term trend, such that $\Delta Y_{t,c} = Y_{t,c} - \hat{Y}_{t,c}$. Thus, we can rewrite the model from Equation 7 as

$$\begin{aligned} \ln(M_{t+1,d,o}) &= \alpha + \beta_1 \cdot (\hat{Y}_{t,d} - \hat{Y}_{t,o}) + \beta_2 \cdot (\Delta Y_{t,d} - \Delta Y_{t,o}) \\ &\quad + \delta \ln(Dist_{d,o}) + FE_d + FE_o \\ &\quad + \gamma \ln(Stock_{t,d,o}) + \eta \ln(Pop_{t,d}) + \mu \ln(Pop_{t,o}) + FE_t + \varepsilon_{t,d,o} \end{aligned} \tag{9}$$

Our dataset includes immigrant flows from 112 origin countries into the 48 contiguous U.S. states. Note that the time fixed effects account for national immigration policy. Our state fixed effects absorb time-invariant state policies, but not changes in state policies over time. We feel that these effects will not dramatically impact aggregate immigrant flows.

2.3 Estimation Techniques

Anderson (2011) reports that gravity models were initially introduced to study immigration flows by Ravenstein (1889). However, they have been used extensively by trade economists to analyze bilateral export and import flows. Immigration economists have recently adopted them to help identify the determinants of migration. The characteristics and limitations of the gravity model are therefore shared by these two fields, so knowledge about the trade literature is informative for our estimation technique.

Gravity models of international trade regress log bilateral trade flows (either exports or imports) on the economic mass of each trading partner, the geographic distance between them, and other covariates. Our procedure simply replaces trade flows with gross migrant flows. Estimation problems arise, however, when country pairs experience zero flows since log values are undefined. This is a nontrivial issue both in trade and in our analysis. For example, half of the observations used in recent important work by Santos Silva and Tenreyro (2006) and Helpman, Melitz, and Rubinstein (2008) equaled zero. Summarizing trade data on the 10-digit harmonized system of goods classification (HS10), Baldwin and Harrigan (2007) report that “The U.S. imports nearly 17,000 different HS10 categories from 228 countries, for a total of over 3.8 million potential trade flows [but] over 90 percent of these potential trade flows are zeros” (p.23). In our dataset of immigrant flows from origin countries to U.S. states, we encounter values of zero in roughly 95% of the observations. Thus, the number of zero values in our dataset is quite similar to that confronted in trade.

The simplest method of overcoming this limitation – scaled ordinary least squares (SOLS) – adds a scalar (usually one) to each flow value before taking natural logs. Analysts may augment this approach by performing tobit estimation (e.g., Dodson, 2001) and censoring

log values less than zero. Others estimate a truncated model (i.e., drop observations of zero flows). The two-part model – explained in section 2.3.2 and often employed in health economics – first estimates a probit model to identify the determinants of whether positive values exist, and then performs OLS estimation of the truncated model. Our analysis will explore both SOLS and two-part estimation.

The similar nature of our data limitations to that of trade studies, however, motivates us to also consider more advanced techniques from recent trade literature. Martin and Pham (2008) thoroughly evaluate the efficacy of alternative estimation strategies when many zero values are present. Among the simpler solutions, the authors find that truncated OLS models outperform censored regressions, and that “just solving the ‘zero problem’ and adding the zero valued observations to the sample is quite an unhelpful strategy” (p. 20). Among the advanced procedures, they conclude that the smallest biases arise when using Eaton and Tamura tobit estimators (after controlling for heteroskedasticity). The Heckman two-step estimator performs well only if the true underlying data is governed by a Heckman selection-model data generating process. Otherwise, the Heckman model commonly fails to converge or produces massive biases.³ A well-known paper by Santos Silva and Tenreyro (2006) advocates the Pseudo Poisson Maximum Likelihood (PPML) procedure. Martin and Pham (2008) agree that PPML performs well “for analysis of nonlinear relationships in models where zero values of the dependent variable are infrequent” (p. 2), but they go on to emphasize that it provides severely biased estimates and is inferior to the Eaton and Tamura procedure when many observations equal zero.

Economists studying the determinants of migration are obviously aware of the problem of zero immigration flows. Those preferring the gravity approach usually adopt truncated, SOLS, or censored methodologies.⁴ Some eschew the gravity model and instead measure flows or emigration rates in levels (not logs).⁵ A few, however, are beginning to take the issue of zero immigration flows more seriously. For example, the Falck et. al. (2010) analysis of linguistic determinants of German regional migration is robust across truncated and PPML methodologies. PPML seems appropriate in their setting since only about four percent of their flows equal zero. Alternatively, Beine et. al. (2011) estimate the role of diasporas (i.e., stock of current immigrants) in predicting the current flow of immigrants using bilateral

³Moreover, the Heckman model requires one variable used in the first (selection) stage of the model to be omitted from the second (quantity) stage. In the context of immigration, this would require a variable that is related to the probability of positive immigration flows but unrelated to the size of immigrant flows among observations with positive values. It is difficult to imagine such a variable.

⁴See Lewer, Pacheco & Rossouw (2009), Lewer and Van den Berg (2008), Ortega and Peri (2009), Falck et. al. (2010), or Beine et. al. (2011) for recent examples.

⁵See Adsera and Pytlikova (2010), Pedersen, Pytlikova & Smith (2008), Zavadny (1997), Dodson (2001) or Mayda (2010).

OECD data. They too have a large number of observations with zero values, and favor the Heckman selection and Instrumental Variable (IV) methods to help with endogeneity issues.

Given the popularity of the methods, we perform traditional SOLS and two-part estimation. Motivated by recent evidence in Martin and Pham (2008), however, we also perform the Eaton and Tamura procedure.

2.3.1 Eaton and Tamura

Scaled ordinary least squares, though extremely common, is an inherently bias model. There is little justification for adding a scalar of one to each value before taking logs. Eaton and Tamura (1994) introduced a threshold tobit model to overcome this limitation. When analyzing Japanese and American trade patterns with a sample of countries in the late 1980s, the authors were confronted with a dataset in which many trade flows equaled zero. Rather than simply adding one to each value before taking logs, they added λ , a value to be statistically estimated.⁶

Let the flow of immigrants ($M_{t,d,o}$) to destination state d from origin country o in year t be defined by:

$$M_{t,d,o} = \max \left\{ 0, \tilde{M}_{t,d,o} \right\} \quad (10)$$

The latent variable $\tilde{M}_{t,d,o}$ is a function of several determinants of migration ($X_{t,d,o}$), a mean-zero normally-distributed error term ($\varepsilon_{t,d,o}$), and a threshold value (λ) that the function of explanatory variables must achieve before positive migration flows occur.⁷

$$\tilde{M}_{t,d,o} = -\lambda + \exp(\alpha + \beta \cdot X_{t,d,o} + \varepsilon_{t,d,o}) \quad (11)$$

By substituting equation (11) into equation (10), rearranging, and taking natural logs, we derive equation (12). Eaton and Tamura (1994) provide the density function for $\tilde{M}_{t,d,o}$ and the necessary log-likelihood function for maximum likelihood estimation.⁸ Thus,

$$\ln(\lambda + M_{t,d,o}) = \begin{cases} \alpha + \beta \cdot X_{t,d,o} + \varepsilon_{t,d,o} & \text{if } \tilde{M}_{t,d,o} > 0 \\ \ln(\lambda) & \text{if } \tilde{M}_{t,d,o} \leq 0 \end{cases} \quad (12)$$

The Eaton and Tamura model is not altogether unfamiliar to immigration economists; Head and Ries (1998) and Rauch and Trindade (2002) used the methodology in their influential analyses of immigration's role in promoting international trade. To our knowledge,

⁶Head and Ries (1998) note that one problem with adding one to each observation is that results will be sensitive to the units of measurement, whereas the Eaton and Tamura method overcomes this limitation.

⁷Head and Ries (1998) interpret λ as undermeasurement.

⁸We are indebted to Cong S. Pham for kindly providing Stata code for the procedure.

however, we are the first to apply the technique to a gravity model of the determinants of immigration. The model presents two limitations, however. First, since it is a non-linear model estimated by maximum likelihood, it is possible that it will fail to converge to a solution. We do not encounter this problem in our analysis. Second, it can be difficult to interpret coefficient estimates, as is the case with the common SOLS solution of adding one to zero values. Strictly speaking, coefficients do not represent percentage changes of the dependent variable, though we follow the often-used convention of interpreting them in this manner.

2.3.2 Two-Part Model

The two-part model consists of first estimating a probit model with a latent variable formulation.⁹ If $M_{t,d,o}$ is the flow of immigrants to destination state d from origin country o in year t , then let the indicator $M^*=1$ if $M_{t,d,o} > 0$, and $M^*=0$ otherwise. As before, the regressors are $X_{t,d,o}$. The two-part model for $M_{t,d,o}$ is then

$$f(M_{t,d,o}/X_{t,d,o}) = \left\{ \begin{array}{l} = Pr(M^* = 0|X_{t,d,o}) \text{ if } M_{t,d,o} = 0 \\ = Pr(M^* = 1|X_{t,d,o})f(M_{t,d,o}|M^* = 1, X_{t,d,o}) \text{ if } M_{t,d,o} > 0 \end{array} \right\} \quad (13)$$

The two-part model consists of: (1) estimating a probit on M^* , and (2) estimating the truncated OLS for $M_{t+1,d,o}$ specified in Equation 9 if $M_{t,d,o} > 0$.

3 Data

Since policy debate often focuses on immigrants with little educational attainment, we focus our analysis on foreign-born workers with a high-school degree or less education. We consider only those who are in the U.S. labor force at the time of survey and are between 18 and 89 years of age. We first analyze the flow of male immigrants, but then incorporate female immigrants into our analysis in Section 4.3.

We limit our analysis to the 2000-2009 period. Though this is a relatively short time series for analyzing short- and long-run GDP differences, it was a decade of considerable volatility in GDP, both in the U.S. and abroad (particularly when compared to the Great Moderation of the 1990s). One advantage of this short time series is that U.S. immigration policy was relatively unchanged during the period (with a few notable exceptions, including changes in the number of H-1B visas for college-educated workers). However, this decade witnessed the largest inflow of new immigrants in U.S. history, with approximately 14 millions new (legal and illegal) immigrants (Camarota, 2011). In addition, this decade experienced the

⁹When the dependent variable exceeds zero, the model is a hurdle (or threshold) model.

most significant dispersion of new immigrants across the U.S., with fewer immigrants going to traditional U.S. destinations than in previous decades (Camarota, 2011). As a result, we think it is important to understand how GDP differentials between origin countries and U.S. states affected the flow of new immigrants during this period.

We use IPUMS data from the 2000-2009 Census and ACS surveys to obtain annual estimates for the gross inflow of new less-educated immigrants in the labor force from each origin country to each U.S. state. The value of this dataset is that it provides good measures of both legal and illegal immigrants residing in the U.S. by country of origin, and immigrants are identified by state of residence. Prior to the 2000 Census and the inception of annual ACS surveys (beginning in 2001), researchers were not able to use Census data to generate accurate measures of newly arrived U.S. immigrants since the surveys did not report the exact years of entry, but instead provided a range of years that included the year of arrival. To our knowledge, we are the first to use Census/ACS to generate annual gross inflow data for the U.S. and measure its response to state-level economic conditions. Note that gross inflows include both new and return migrants.

The dataset is not without its limitations, however. First, the ACS/Census data is better at measuring the immigrant stock (or net change in immigrant stock) than at capturing the gross flow of new immigrants entering the country. Unfortunately, both stock and net change measures are inconsistent with theoretical models of bilateral migration. We believe a reasonable proxy for U.S. gross inflows is the number of foreign-born residents in each state who arrived in the U.S. within the last year. Second, our gross inflow measure does not include immigrants in the labor force who previously arrived in the U.S. but recently moved to a new state in the U.S.¹⁰ Third, there is likely a lag between arrival in the U.S. and being enumerated in the ACS, and this lag may lead to a downward bias in immigrant flows which vary by country of origin and/or destination state. This issue may be especially salient in the case of less-educated, illegal immigrants (which is the focus of our study). Finally, the ACS is administered monthly, but the Census only provides information aggregated to the annual level. An economic shock in period t might have a larger effect on potential migrants at the beginning of period $t + 1$ than at the end, but the ACS will not allow us to identify such a distinction. While there is little we can do to directly address these issues, we believe that they will not significantly bias our analysis.

¹⁰Our analysis is related to, but not directly comparable with, the work of Borjas (2001) and others that analyze how newly arrived immigrants (those who have been in the U.S. less than five years) respond to wage differentials within the U.S.

3.1 Recent Immigration Trends

According to the ACS data, approximately 11.26 million new immigrants entered the U.S. between 2000 and 2009. This includes both legal and illegal immigrants, as measured by the Census/ACS. Table 1 shows the breakdown of the sample by gender, education level, and employment status. Of the 11.25 million new immigrants, approximately 53% are male, 59% are less-educated (having a high school degree or less), and 63% are in the labor force at the time of survey. The largest subgroup of immigrants is male, less-educated, and in the labor force, which represents 28% of the sample, or 3.14 million people.

In this paper, we are attempting to measure the extent to which economic conditions drive immigrant flows, so we analyze only immigrants who are in the labor force (those who report to be employed or unemployed but looking for work at the time of survey). In addition, female immigrants in the labor force represent a relatively small share of all new immigrants (19%), and hence are not the focus of our analysis (although we will provide results for this group). We focus on less-educated immigrants since they represent a majority of all new immigrants and the immigration debate about them is the most contentious, though Section 4.4 explores effects for well-educated immigrants.

The first column of Table 2 reports the primary regions of origin of all new immigrants between 2000 and 2009. Clearly, Latin America has provided the majority (55.7%) of new immigrants. The second largest sending region is Asia, representing 26.1% of all new U.S. immigrants. Approximately 11.4% of all new immigrants originate from Europe. Not surprising, the distribution of sending regions for male, less-educated immigrants in the labor force (reported in column 2) is quite different, with 86% coming from Latin America. Only 7.1% and 3.9% of these types of immigrants originate from Asia and Europe, respectively.

As evident in Table 3, more than one-third of all new immigrants and two-thirds of all new male, less-educated immigrants in the labor force are from Mexico. India is the next largest sender, representing 7.2% of all immigrants. Immigrants from the Philippines and China represent approximately 3.7% of all U.S. immigrants. Notice, however, that the distribution of male, less-educated immigrants in the labor force is more skewed to Latin American countries, with a disproportionate share coming from Guatemala, El Salvador, Brazil and Honduras (in the second column).

Table 4 provides a snapshot of where the new immigrants are moving within the U.S. For brevity, we categorize the U.S. states into six regions that are consistent with U.S. Census regions.¹¹ Table 4 shows that approximately one-quarter of all new immigrants live in the Pacific and Southeast regions each. Approximately 20% of new immigrants live in the

¹¹In the empirical specifications, our analysis is at the state level, however.

Northeast, 13% in the Midwest and South Central, and 8% in the Mountain region.¹² The distribution of the locational choices of male, less-educated immigrants in the labor force is very similar (in column 2) to that of all immigrants (in column 1). In column 3, we report the mean trend Gross State Product (GSP) in each region and find no clear correlation between trend GSP and recent immigrant flows.

3.2 Summary Statistics

Recall from Equation 8 that the trend component is estimated based on regressions (using data from 2000-2009) of $\hat{Y}_{t,c} = \hat{a}_c + \hat{b}_c \cdot T$ for $c = \{o, d\}$ where T is the time trend. After obtaining the trend component $\hat{Y}_{t,c}$, we then compute the short-run component as $\Delta Y_{t,c} = Y_{t,c} - \hat{Y}_{t,c}$, where $Y_{t,c}$ is the current year per capita GDP (or GSP). We use per capita real GDP (in 2010 dollars) for origin countries using data from the World Development Indicators and per capita state GSP data (in 2010 chained dollars) from the Bureau of Economic Analysis. We include the 48 contiguous states in the analysis and drop the District of Columbia (which has an exceptionally high GSP per capita). We have nine years of data since we lag all of the independent variables. During this period, we have complete data on 112 different source countries. Therefore, we have a total of 48,374 observations for our state-level analysis (48 X 9 X 112). However, only 2,609 observations have non-zero male less-educated immigrant flows.

Table 5 reports the mean and standard deviation for each variable. The first two columns are for the entire sample and the last two columns represent the sample of non-zero less-educated male immigrant flows. The average bilateral flow of less-educated males between a country and a state is 1,100 per year among observations with positive flows, but just 58 when including the entire sample. We report unweighted means to be consistent with our regression analysis that follows. Both samples exhibit tremendous variation in migrant flows.

GDP differentials, both long-term and short-run, are the independent variables of interest. Average trend per capita GSP of U.S. states is \$44,508 (in 2010 dollars) for the entire sample (with very little variation), while the average GDP for origin countries is \$13,282 (with high variation). Average fluctuations in per capita GDP would equal zero by construction if we used the entire time series since fluctuations are defined as the difference between current and trend GDP. Since the independent variables are lagged, however, we lose the last year of data. The resulting averages are \$65 for origin countries and \$179 for U.S. states, leading to a \$114 gap in GDP fluctuations between the destination and origin of immigrants on average. Variation in GDP fluctuations is very high relative to the mean, with more variation coming

¹²The South Central region includes Arkansas, Louisiana, Oklahoma and Texas.

from state fluctuations than from origin countries.

Our control variables include geographic indicators, destination and origin populations, and immigrant stocks in each state. Immigrant stock is calculated by measuring the number of foreign-born people in each state from each country (of all education levels and including males and females). The mean immigrant stock is 4,766. The geographic indicators include the distance between world capitals and the U.S. state geographic centers, using the Haversine distance formula and latitude/longitude data from the CEPII Research Center¹³ and the U.S. Census. Population estimates are from the World Development Indicators (for countries) and the U.S. Census (for U.S. states).

4 Results

4.1 Less-Educated Male Immigrant Flows

We model the flow of immigrants from origin country o to destination state d in year t as specified in Equation 9, using independent variables that are lagged one year. Recall that we focus on less-educated male immigrants in the labor force. Table 6 presents the baseline results, which includes a full set of fixed effects for the 48 destination states and 112 origin countries in addition to time fixed effects. Each specification controls for immigrant stocks (by taking the natural logarithm of the immigrant stock plus one in each case).¹⁴

In column 1, we include all immigrant flows by adding one to the flow variable before taking the natural log and then employing OLS (i.e., SOLS); in column 2, we use the Eaton and Tamura technique (as described in section 2.3.1); in columns 3-4, we use the two-part model (as described in section 2.3.2) which first estimates a probit (in column 3) and then estimates the non-zero immigrant flows using truncated OLS (in column 4). The sample size is much smaller in column 4 compared to columns 1-3 since observations with zero immigrant flows are dropped.¹⁵

Table 6 presents the results of the four different estimation strategies. First, SOLS (in column 1) suggests that both trend GDP and GDP fluctuations are significant determinants of the flows of less-educated immigrant men in the U.S. Our results indicate that a \$1,000 differential in GDP fluctuations between the destination state and origin country leads to a 2.2% increase in less-educated male immigration flows (significant at the 1% level). Similarly,

¹³<http://www.cepii.fr/anglaisgraph/bdd/distances.htm>

¹⁴We do this to maximize the number of observations in each regression. All of our results are comparable when we do not add one to the immigrant stock variable.

¹⁵In the probit specification, there are two countries (Antigua-Barbuda and Finland) where the dummy variables perfectly predict the zero outcome. Hence these observations are dropped in the estimation, leading to 864 fewer observations (2 countries X 48 states X 9 years).

a \$1,000 increase in the trend GDP differential between the destination state and origin country induces a significant 1.2% increase in immigration. This is particularly striking given that the model is estimated with a full set of country and state fixed effects. The coefficient is identified only by differences in trend growth rates across states and countries ($b_c \cdot T$ in the Equation 8 construction of our trend variable), not by differences in permanently high levels of per capita GDP (a_c in Equation 8). We should also note that our array of fixed effects would absorb all of the variation in trend GDP if we had restricted growth rates (b_c) to be equal across states and countries. Thus, the GDP trend coefficient in Table 6 is only identifiable because we allow for state and country-specific trends.

These baseline results are consistent with those of previous studies. Ortega and Peri (2009), for example, find a significantly positive relationship between GDP differentials and bilateral immigrant flows using OECD data. Their OLS specification is similar to our SOLS specification in that they add one to both immigrant flows and immigrant stocks, while also including observations with zero flows. Their magnitudes are not directly comparable to ours since they use a different database, cover a cross-section of source countries, and do not distinguish between trend and cyclical effects. Nonetheless, they find that a \$1,000 GDP differential (in levels) leads to a 10 percentage point increase in bilateral immigration flows across OECD countries.

Our control variables have the expected signs when significant. Distance is negatively associated with higher flows of less-educated immigrant men, and larger origin countries send more immigrants. Both results are consistent with the literature (i.e., Karemera et al., 2000; Lewer and Van den Berg, 2008; Ortega and Peri, 2009). We also find that immigrant stocks are highly positively correlated with immigrant flows. This network effect has been frequently documented in the literature (Bartel; 1989; Zavodny 1997; Clark, Hatton and Williamson, 2007; Grogger and Hanson, 2011; Mayda, 2010).

Column 2 uses Eaton and Tamura threshold tobit estimation. The SOLS method of adding one to the dependent variable before taking logs, though common, is inherently biased in the sense that there is no reason to prefer an added scalar of one to any other value. We prefer the Eaton and Tamura method because it instead allows this scalar to be estimated by the data itself. Our results from employing these two strategies are similar but with important differences. First, effects from short-run GDP differentials increase somewhat. A \$1,000 differential in GDP fluctuations between the destination state and origin country leads to a 3.4% increase in less-educated male immigration flows. More interestingly, however, the coefficient on trend GDP loses significance. As with our SOLS specification, interpretation of this result must come with the caveat that much of the immigration effect from differences in long-term GDP are absorbed in the fixed effects.

The two-part model in columns 3 and 4 separates the likelihood of a country sending any immigrants with the magnitude of the immigrant flow response among existent bilateral immigrant routes. The probit model in column 3 suggests that long-run and short-run GDP differentials both matter in determining which countries send positive less-educated male immigrants to the U.S. Neither income measure, however, is important in determining the size of the flows in the truncated OLS model of column 4. The two-part model therefore suggests that the significant coefficients in the SOLS and Eaton and Tamura specifications are driven by the discrete jump of going from zero to positive flows, not in changing the magnitude of flows within existent bilateral migration channels.

Another interpretation of our results is that GDP fluctuations are associated with the flow of less-educated immigrant men as long as zero flow values are included in the estimation. This is robust across three different estimation techniques (SOLS, Eaton and Tamura, and probit). All three suggest that an increase in short-run GDP will lead to an increase in the flow of less-educated immigrant men in that state. The evidence for long-run GDP differentials is less robust, but fixed effects likely absorb much of the variation in long-term GDP that drives migration decisions.

Our method for calculating trend GDP using data from the 2000-2009 period is not the only available option. We also considered an alternative that estimates the trend component by projecting 1990-1999 GDP growth onto the subsequent decade (this also reduces the number of observations available). The results (which we do not report) continue to find that short-run GDP fluctuations are significantly correlated with immigrant flows. Moreover, regressions using this method robustly find that the GDP trend differential is a significant predictor of immigrant flows. Nonetheless, we prefer the results from our reported trend construction, in part because we feel that the loss in observations is too costly for the rest of the analysis.

In sum, the truncated OLS model that drops all zero values suggests that neither long-run or short-run GDP differentials have an effect on immigrant flows. However, models that include the observations with zero immigrant flows find strong evidence that short-run GDP fluctuations are positively correlated with immigrant flows of less-educated men into U.S. states. These regressions also provide some evidence that long-run GDP matters. This indicates that the inclusion of the observations with zero immigrant flows is essential to accurately estimate the effects of GDP differentials on immigrant flows and discredits methods that drop the zeros. In addition, our results suggest that simply adding a one to the dependent variable may be biasing the results for GDP differentials. When using the Eaton and Tamura method, which estimates a scalar to add to the dependent variable, long-run GDP differentials are no longer significant.

4.2 Immigrants from Mexico

Mexicans represent 67% of all new less-educated male immigrants, so it is important to understand how the results are affected by their inclusion or exclusion. In this section, we briefly discuss methods we used to determine if the reaction of Mexican immigrants is different from the reaction of immigrants from other countries.¹⁶

First, to test if Mexican immigrants react to GDP differentials differently than immigrants from other countries, we include an interaction term of a dummy variable for Mexico and the two GDP variables. We find that the estimated coefficients on GDP trend and GDP fluctuations are entirely comparable to those in the baseline case, both in magnitude and significance. In addition, the interaction term of Mexico with GDP fluctuations is weakly significant and positive. This provides some support to the idea that male immigrants from Mexico tend to react to GDP fluctuations in a way that is structurally different than immigrants from other countries. In addition, it reinforces the findings of Hanson and Spilimbergo (1999), which indicate that the flow of (illegal) Mexican immigrants are highly sensitive to changes in Mexican wages.

Since immigrants from Mexico represent a large share of flows and seem to react differently to macroeconomic factors, we drop Mexico from our sample of destination countries to test for robustness. Our results are almost exactly as the baseline results reported in Table 6, with strong evidence that GDP fluctuations are positively correlated with immigrant flows and limited evidence that GDP trends matter. The remaining control variables have the same signs and significance as in the baseline case. Thus, we find that the inclusion of Mexican immigrants in our full sample is not driving our results.

4.3 Less-Educated Female Immigrant Flows

Thus far we have considered only less-educated male immigrants in the labor force. This is standard in the immigration literature when trying to isolate immigrants who move for economic purposes. However, recent discussions indicate that women are increasingly migrating for economic reasons compared to earlier cohorts (United Nations, 2004). We now change the dependent variable so that it represents the log of less-educated women in the labor force migrating to the U.S. Table 7 reports the results. Note that state, country, and time fixed effects are included in all of the specifications.

SOLS, Eaton and Tamura, and probit results for women echo those for men, but with muted and sometimes insignificant effects from the GDP variables. For example, the SOLS regression (in column 1) finds that a \$1,000 increase in GDP trends will lead to a 0.8%

¹⁶For brevity, we do not include the full set of results, but they are available upon request.

increase in immigrant women, compared to 1.2% for immigrant men. Similarly, a \$1,000 increase in GDP fluctuations leads to a 1.6% increase in immigrant women and a 2.2% increase in immigrant men. In the Eaton and Tamura model for women, the insignificant coefficient on GDP trend (0.005) is half the size of the effect from the male regression. The also-insignificant coefficient on GDP fluctuations is less than one third of the size. Probit coefficients are similarly between one-half and one-third the size in female regressions. Altogether, these results suggest that female migration decisions are less sensitive to economic conditions than male decisions are. Strangely, the truncated model departs from this regularity by finding a large and significant coefficient on trend GDP for women, and a negative but insignificant coefficient on GDP fluctuations. However, the small number of observations in this specification, coupled with the consistent effects from the other models, persuade us to conclude that the results largely support the finding from past literature that female migration decisions are less sensitive to economic conditions. That is, the response of female immigrant flows to long-run GDP differentials is smaller than male flows as long as the observations with zero values are included. Similar to the results for male immigrants, our results for immigrant women indicate that methods that include the zero values deliver more robust findings.

4.4 Well-Educated Immigrant Men

The focus of the paper thus far has been on less-educated immigrant men in the labor force, who constitute 28% of all immigrants and 52.2% of all male immigrants in the sample. For comparison, we now analyze the effects for male labor with some college or more education (15.8% of all immigrants). The economic motives behind the migration decision likely differ for less-educated and well-educated immigrants, and our results should add insight into the extent of these differences.

Table 8 displays the results. We find very little evidence that long-run GDP differences matter for well-educated immigrant men, with the exception of the truncated OLS in column 4 when the observations with zero values are dropped. The relationship between short-run GDP fluctuations and flows of well-educated males, however, is similar to that of less-educated males, but the point estimates are less precise. The SOLS (in column 1) regression delivers a highly significant estimate for GDP fluctuations, the Eaton and Tamura estimate is nearly significant (p-value of 0.103), and the probit delivers an estimate that is significant at the 10% level. The size of the estimated coefficients for GDP fluctuations are similar to those of less-educated male immigrants (in Table 6). Thus, differences in short-run GDP differentials have a somewhat similar, but less precise, effect on flows of well-educated

immigrant men compared to less-educated immigrant men, while long-run GDP differentials have seemingly little effect.

4.5 Push and Pull Factors: Less-Educated Male Immigrants

The specification in Equation 9 assumes that the coefficients on the destination and origin country GDP are the same. Empirically, it is not necessary to impose this restriction. By disaggregating the trend and fluctuation components of GDP by destination and origin, Equation 14 can help clarify the source of correlations and tell a more precise story about immigrant flows.

$$\begin{aligned} \ln(M_{t+1,d,o}) = & \alpha + \beta_1 \hat{Y}_{t,d} - \beta_2 \hat{Y}_{t,o} + \beta_3 \Delta Y_{t,d} - \beta_4 \Delta Y_{t,o} \\ & + \delta \ln(Dist_{d,o}) + FE_d + FE_o \\ & + \gamma \ln(Stock_{t,d,o}) + \eta \ln(Pop_{t,d}) + \mu \ln(Pop_{t,o}) + FE_t + \varepsilon_{t,d,o} \end{aligned} \quad (14)$$

This specification differentiates between push and pull factors for both long-term and short-run GDP variables. Estimated coefficients for β_2 and β_4 will determine if income in origin countries pushes immigrants out, while estimates for β_1 and β_3 will identify whether income in destination states pulls immigrants in. The results (for less-educated male immigrants) are reported in Table 9.

Tables 6-8 presented mixed evidence on the influence of long-term GDP differentials on migration decisions. Table 9 demonstrates that if such an effect exists, it is clearly driven by long-term GDP in the origin country and not by income in the destination state. The coefficient on state GDP trend is insignificant in all four specifications, while the coefficient on origin country GDP per capita trend is negative and significant in all but the truncated OLS specification (in column 4). According to the Eaton and Tamura method in column 2, a \$1,000 increase in origin country trend GDP leads to a 4.4% reduction in less-educated male immigrants. Recall that fixed effects absorb much of the trend GDP variation so that coefficients are being identified by differences the growth portion of trend GDP (the $b_c \cdot T$ term in Equation 8). Thus, we see that countries experiencing more long-term GDP growth send fewer immigrants to the U.S. than slow-growth countries do, but state trends are not a determinant in attracting immigrants.

Different mechanisms appear to govern immigration's relationship with short-run GDP fluctuations. Unlike with trend GDP, state-level GDP fluctuations attract immigrants, but fluctuations in origin-country income have no effect. Similar to the baseline regressions in Table 6, the results in Table 9 argue for effects from short-run GDP only in specifications

accounting for zero flow values (columns 1-3). However, we now see that such effects arise primarily because economic booms in U.S. states will attract immigrants from abroad. For example, the Eaton and Tamura results in column 2 argue that a \$1,000 increase in short-run GDP in a particular U.S. state will lead to a 4.8% increase in male immigrants to that state. Coefficients on origin-country GDP fluctuations are always insignificant.

Altogether, it appears that long-run (or trend) GDP determines which countries send immigrants to the U.S., while short-run fluctuations determine which U.S. state they move to, suggesting that immigrants are being pushed out of poor (or slow-growth) countries and pulled into states that have experienced recent booms. This result is broadly consistent with Mayda (2010) who finds that pull factors (i.e., per capita GDP) are positively associated with higher emigration rates for a panel of OECD countries.

5 Conclusion

This paper adds to the literature on the determinants of immigrant flows in three ways. First, we use variation in income across U.S. states to uncover how newly arrived less-educated immigrants respond to income differentials. Second, we decompose income differentials into short- and long-run components. Third, we employ several estimation techniques, including the threshold tobit and two-part models to appropriately account for the large number of zero values for immigrant flows in our dataset.

We study U.S. immigration between 2000 and 2009. This period is an interesting case-study because the U.S. experienced the largest gross inflow of new immigrants in its history, and those immigrants were more dispersed across the U.S. compared to recent cohorts. In addition, the U.S. witnessed a severe recession in the latter half of this decade. Not only does this time period provide a great deal of macroeconomic variation, but it will also appeal to policy makers interested in the extent to which differences in trend GDP and GDP fluctuations are correlated with immigrant flows.

We find that both long-term and short-run GDP differentials significantly determine the flow of newly arrived less-educated male immigrants into U.S. states. However, the evidence for long-term GDP differentials is mixed, which is most likely due to the inclusion of state and country fixed effects. In addition, the evidence for short-run differentials requires that observations of zero flow values are included in the regression. For example, a truncated OLS specification which drops the observations with zero values (representing 95 percent of the sample) suggests that neither differences in GDP trends nor GDP fluctuations between the source country and destination state affect the flow of less-educated male immigrants into the U.S. However, specifications that include zero values suggest otherwise, most notably in

recognizing a positive relationship between GDP fluctuations and immigrant flows.

We document important differences in the response of recent immigrant flows to short-run and long-run GDP components based on gender, education level, and country of origin. For example, the flows of less-educated female immigrants into the U.S. are generally less responsive to differences in GDP fluctuations than their male counterparts. In addition, Mexican migration is not driving our results, as Mexican migrants respond to economic conditions much like residents from other countries do. Our results for well-educated men are less robust, but suggest that this group also responds to differences in GDP fluctuations more than GDP trends.

We also augment the immigration literature attempting to disentangle push and pull effects. We find that less-educated immigrants are pushed out of their countries by long-run GDP trends, and are pulled into U.S. states by short-run upswings in economic activity. Not surprising, short-run fluctuations in the origin country do not lead to an increase in less-educated immigrant flows to the U.S. It is not difficult to imagine a story consistent with these findings. People from poor countries want to immigrate to the U.S., but short-term fluctuations in their country of origin are largely irrelevant for the decision to stay or leave. When deciding upon a new destination, however, an individual is likely to be enticed by a booming location and the associated promise of available jobs. From the perspective of a potential new worker, states with recent economic growth look more attractive than states with stagnant economic activity.

Our results also shed some light on the importance of empirical specification when studying immigrant flows. The truncated OLS model estimates the determinants of migration conditional upon a bilateral country-by-state observation recording positive flows. It should not be taken as representative, however, for those wishing to understand all potential flows, since the sample excludes 95% of all possible observations. The probit model, while useful, only estimates a dichotomous effect. That is, it identifies whether GDP differentials affect the probability that a country-by-state migration channel records positive flows. This may or may not be interesting to the policymaker. SOLS is a simple method of using the entire sample of data to estimate the effects of GDP on the quantity of immigration flows, but it accomplishes this by arbitrarily adding one to all flow values before taking logs. As an alternative, the Eaton and Tamura model allows the added scalar to be a value that is estimated by the data itself. It therefore permits more flexibility than simple OLS. This flexibility should encourage researchers to prefer the Eaton and Tamura method to estimate the relationship between GDP components and immigrant flows.

References

- [1] Adsera, Alicia and Mariola Pytlikova. (2010). “The Role of Language in Shaping International Migration: Evidence from OECD Countries 1985-2006.” Mimeo.
- [2] Anderson, James E. (2011). “The Gravity Model.” *Annual Review of Economics*, 3: 133-160
- [3] Baldwin, Richard and James Harrigan. (2007). “Zeros, Quality, and Space: Trade Theory and Trade Evidence.” NBER Working Paper 13214.
- [4] Bartel, Ann P. (1989). “Where Do the New United States Immigrants Live?” *Journal of Labor Economics* 7 (October): 371–91.
- [5] Beine, M., Docquier, F. and C. Ozden. (2011). “Diasporas.” *Journal of Development Economics*, 95 (1), 30-41.
- [6] Borger, S. (2008). “Estimates of the Cyclical Inflow of Undocumented Migrants to the United States.” University of California San Diego Working Paper.
- [7] Borjas, George. (1999). “Immigration and Welfare Magnets.” *Journal of Labor Economics*, October 1999, pp. 607-637.
- [8] Borjas, George. (2001). “Does Immigration Grease the Wheels of the Labor Market?” Brookings Papers on Economic Activity, pp. 69-119.
- [9] Camarota, Steven. (2001). “A Record-Setting Decade of Immigration: 2000 to 2010,” Center for Immigration Studies Backgrounder, October 2011. <http://www.cis.org/articles/2011/record-setting-decade.pdf>
- [10] Clark, Ximena, Timothy Hatton, and Jeffrey Williamson. (2007). “Explaining U.S. Immigration, 1971-1998.” *Review of Economics and Statistics*, 89(2): 359-373.
- [11] Davis, S.J. and J. Haltiwanger. (1992). “Gross Job Creation, Gross Job Destruction, and Employment Reallocation.” *The Quarterly Journal of Economics*, 107(3): 819-863.
- [12] Dodson III, Marvin. (2001). “Welfare Generosity and Location Choices among New United States Immigrants,” *International Review of Law and Economics*, 21 (1), 47-67
- [13] Eaton, Jonathan and Akiko Tamura. (1994). “Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns.” *Journal of the Japanese and International Economies*, Vol. 8, pp. 478-510.

- [14] Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Sudekum. (2010). "Dialects, Cultural Identity, and Economic Exchange." IZA Discussion Papers 4743.
- [15] Grogger, Jeffrey and Gordon H. Hanson. (2011). "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 95(1), 42-57.
- [16] Hainmueller, Jens and Michael J. Hiscox. (2010). "Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review*, 104, 61-84
- [17] Hanson, G.H. and A. Spilimbergo. (1999). "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the U.S.-Mexico Border." *American Economic Review* 89(5), pp. 1337-57.
- [18] Hanson, Gordon H., Robertson, Raymond and Antonio Spilimbergo, 2002. "Does Border Enforcement Protect U.S. Workers From Illegal Immigration?," *The Review of Economics and Statistics*, 84(1), pp. 73-92.
- [19] Hatton, Timothy J. and Jeffrey G. Williamson. (2003). "What Fundamentals Drive World Migration?" UN World Institute for Development Economic Research, Discussion Paper No. 2003/23.
- [20] Head, Keith and John Ries. (1998). "Immigration and Trade Creation: Econometric Evidence from Canada." *Canadian Journal of Economics*, Vol. 31 (1), pp. 47-62.
- [21] Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. (2008). "Estimating Trade Flows: Trading Partners and Trading Volumes." *The Quarterly Journal of Economics*, Vol. 123 (2), pp. 441-487.
- [22] Karemera, David, Victor I. Oguledo, and Bobby Davis. (2000). "A Gravity Model Analysis of International Migration to North America." *Applied Economics*, 32 (13): 1745-55.
- [23] Kugler, Adriana and Mutlu Yuksel. (2008). "Effects of Low-Skilled Immigration on U.S. Natives: Evidence from Hurricane Mitch," NBER Working Paper 14293.
- [24] Lewer, Joshua J. and Hendrik Van den Berg. (2008). "A Gravity Model of Immigration." *Economics Letters*, April 2008, v. 99, iss. 1, pp. 164-67.
- [25] Lewer, Joshua J., Gail Pacheco and Stephani Rossouw. (2009). "Do Non-Economic Quality of Life Factors Drive Immigration?" IZA DP No. 4385.

- [26] Mandelman, Federico and Andrei Zlate. (2010). "Immigration, Remittances and Business Cycles," Federal Reserve Bank of Atlanta Working Paper.
- [27] Martin, Will and Cong S. Pham. (2008). "Estimating the Gravity Model When Zero Trade Flows are Frequent." Deakin University, Faculty of Business and Law, School of Accounting, Economics, and Finance Working Paper 2008_03.
- [28] Mayda, Anna Maria. (2006). "Who Is Against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants," *The Review of Economics and Statistics*, 88(3), pp.510-530.
- [29] Mayda, Anna Maria. (2010). "International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows." *Journal of Population Economics*, 23 (4): 1249-1274.
- [30] McFadden, D. (1984). "Econometric analysis of qualitative response models." In: Griliches, Z., Intriligator, M. (Eds.), *Handbook of Econometrics*, Volume 2. Elsevier/North- Holland, Amsterdam.
- [31] Orrenius, Pia M. and Madeline Zavodny. (2009). "Tied to the Business Cycle: How Immigrants Fare in Good and Bad Economic Times." Washington, DC: Migration Policy Institute.
- [32] Ortega, Francesc and Giovanni Peri. (2009). "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005." NBER Working Paper 14833.
- [33] Pedersen, Peder J., Mariola Pytlikova, and Nina Smith. (2008). "Selection and Network Effects - Migration Flows Into OECD Countries 1990-2000." *European Economic Review*, Vol. 52, pp. 1160-1186.
- [34] Rauch, James E. and Vitor Trindade. (2002). "Ethnic Chinese Networks in International Trade." *The Review of Economics and Statistics*, 84 (1), pp. 116-130.
- [35] Ravenstein, E. G. (1889). "The Laws of Migration." *Journal of the Royal Statistical Society*, 52: 241-305.
- [36] Rosenzweig, Mark. (2007). "Education and Migration: A Global Perspective." Yale University, Mimeo.
- [37] Şahin, Ayşegül, Joseph Song and Bart Hobijn. (2010). "The Unemployment Gender Gap During the Current Recession." *Federal Reserve Bank of New York Current Issues*, March, Volume 16 (2).

- [38] Santos Silva, J.M.C. and Silvana Tenreyro. (2006). “The Log of Gravity.” *The Review of Economics and Statistics*, Vol. 88 (4), pp. 641-658.
- [39] Taylor, E. (2006). “International Migration and Economics Development.” International Symposium on International Migration and Development, Population Division, Department of Economic and Social Affairs, United Nations Secretariat. http://www.un.org/esa/population/migration/turin/Symposium_Turin_files/P09_SYMP_Taylor.pdf
- [40] United Nations, (2004). “World Survey on the Role of Women in Development: Women and International Migration.” Economic and Social Affairs. <http://www.un.org/womenwatch/daw/public/WorldSurvey2004-Women&Migration.pdf>
- [41] Warin, T., and P. Svaton. (2008). “European Migration: Welfare Migration or Economic Migration?” *Global Economy Journal*, 8(3), 1-30.
- [42] Zaiceva, A., and K.F. Zimmermann. (2008). “Scale, Diversity, and Determinants of Labour Migration in Europe.” *Oxford Review of Economic Policy*, 24(3), 427-451.
- [43] Zavodny, Madeline. (1997). “Welfare and the Locational Choices of New Immigrants.” *Economic Review – Federal Reserve Bank of Dallas*, Second Quarter 1997, pp. 2-10.

Tables

Table 1: Characteristics of Immigrant Population, 2000-2009

Type of Immigrant	Flow	% of Total
Male, less-educated, labor force	3,136,560	28.0%
Male, less-educated, not in labor force	628,320	5.6%
Male, well-educated, labor force	1,777,795	15.8%
Male, well-educated, not in labor force	463,279	4.1%
Female, less-educated, labor force	1,167,661	10.4%
Female, less-educated, not in labor force	1,726,462	15.3%
Female, well-educated, labor force	101,6070	9.0%
Female, well-educated, not in labor force	1,339,901	11.9%
Total	11,256,048	
Total Male	6,005,954	53.4%
Total Less-educated	6,659,003	59.2%
Total in Labor Force	7,098,086	63.1%

Notes: Data is from the 2000-2009 ACS.

Table 2: Regions of Origin, 2000-2009

Regions	Flow as % of Total	Flow as % of Baseline Sample*
Latin America	55.72%	86.14%
Asia	26.11%	7.09%
Europe/Russia	11.42%	3.88%
Africa	3.43%	2.06%
Canada	2.55%	0.49%
Oceania	0.78%	0.34%

Notes: Data is from the 2000-2009 ACS. *The 'Baseline Sample' represents new male, less-educated immigrants in the labor force.

Table 3: Largest Countries of Origin, 2000-2009

Regions	Flow as % of Total	Flow as % of Baseline Sample*
Mexico	36.7%	66.81%
India	7.2%	1.02%
Philippines	3.7%	0.61%
China	3.6%	1.65%
Canada	2.5%	0.49%
Korea	2.5%	0.29%
Guatemala	2.3%	4.65%
El Salvador	2.1%	2.97%
Japan	2.0%	0.19%

Notes: *The 'Baseline Sample' represents new male, less-educated immigrants in the labor force. Note that Brazil and Honduras are among the top sending countries for the sample, representing about 1.9% each (but are not listed in the table).

Table 4: U.S. Destination Regions, 2000-2009

Regions	Flow as % of Total	Flow as % of Baseline Sample	Trend GSP
Pacific	24.18%	23.02%	\$50,360
Southeast	22.82%	24.77%	\$43,064
Northeast	19.85%	15.33%	\$53,814
Midwest	12.91%	11.24%	\$45,572
South Central	12.47%	15.59%	\$46,811
Mountain	7.76%	10.05%	\$44,492

Notes: Data is from the 2000-2009 ACS. *The 'Baseline Sample' represents new male, less-educated immigrants in the labor force. Trend and cyclical GSP reported in 2010 \$. The six regions are consistent with Census Bureau regions and divisions.

Table 5: Summary Statistics

Variables	$M_{d,o} \geq 0$		$M_{d,o} > 0$	
	Mean	St Dev	Mean	St Dev
Migrant Flows	58.06	931.71	1100.545	3913.226
$\hat{Y}_{t,d}$: State trend (in thousands of US 2010 \$)	44.508	7.845	47.724	7.057
$\hat{Y}_{t,o}$: Country trend (in thousands of US\$)	13.282	12.851	11.182	10.856
$\hat{Y}_{t,d} - \hat{Y}_{t,o}$: GDP trend differential (in thousands of US\$)	31.226	14.986	36.542	13.343
$\Delta Y_{t,d}$: State fluctuations (in thousands of US\$)	0.179	0.877	0.379	0.913
$\Delta Y_{t,o}$: Country fluctuations (in thousands of US\$)	0.065	0.462	0.060	0.344
$\Delta Y_{t,d} - \Delta Y_{t,o}$: GDP fluctuations differential (in thousands of US\$)	0.114	0.924	0.319	0.916
Distance between origin country and destination state (in miles)	5,345	2,168	4,187	2,527
Immigrant Stock	4,766	56,050	56,266	233,86
Country Population (in millions)	49.720	159.472	122.159	285.170
State Population (in millions)	6.051	6.465	11.835	9.528
Observations	48,384		2,609	

Notes: Reported summary statistics are unweighted.

Table 6: Baseline Results: Less-Educated Immigrant Men

Independent variables	Two-Part Model			
	$\ln(1 + M_{d,o})$	Eaton-Tamura	Probit	$\ln(M_{d,o})$
$\hat{Y}_{t,d} - \hat{Y}_{t,o}$: GDP trend	0.012 (0.003)***	0.010 (0.010)	0.0006 (0.0003)**	0.022 (0.019)
$\Delta Y_{t,d} - \Delta Y_{t,o}$: GDP fluctuations	0.022 (0.005)***	0.034 (0.014)**	0.0011 (0.0004)***	0.002 (0.023)
$\ln(Dist_{d,o})$	-0.710 (0.122)***	-0.548 (0.112)***	-0.0110 (0.0027)***	-0.564 (0.176)***
$\ln(Pop_{t,o})$	0.539 (0.170)***	1.317 (0.511)**	0.0314 (0.0131)**	1.059 (0.910)
$\ln(Pop_{t,d})$	-0.201 (0.271)	-0.455 (0.538)	-0.0033 (0.0142)	-1.752 (0.913)*
$\ln(1 + Stock_{t,d,o})$	0.037 (0.003)***	0.140 (0.016)***	0.0039 (0.0003)***	0.086 (0.021)***
Observations	48,384	48,384	47,520	2,609
R^2	0.326	—	—	0.584

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Results incorporate time, destination and origin fixed effects.

Table 7: Less-Educated Immigrant Women

Independent variables	Two-Part Model			
	$\ln(1 + M_{d,o})$	Eaton-Tamura	Probit	$\ln(M_{d,o})$
$\hat{Y}_{t,d} - \hat{Y}_{t,o}$: GDP trend	0.008 (0.003)***	0.005 (0.009)	0.0002 (0.0002)	0.076 (0.019)***
$\Delta Y_{t,d} - \Delta Y_{t,o}$: GDP fluctuations	0.016 (0.004)***	0.010 (0.012)	0.0004 (0.0003)	-0.017 (0.026)
$\ln(Dist_{d,o})$	-0.609 (0.114)***	-0.515 (0.107)***	-0.0097 (0.0024)***	-0.542 (0.180)***
$\ln(Pop_{t,o})$	0.240 (0.157)	0.625 (0.442)	0.0145 (0.0106)	-0.281 (0.908)
$\ln(Pop_{t,d})$	-0.030 (0.240)	-0.043 (0.450)	-0.0011 (0.0106)	1.526 (0.920)*
$\ln(1 + Stock_{t,d,o})$	0.027 (0.003)***	0.109 (0.017)***	0.0027 (0.0002)***	0.070 (0.018)***
Observations	48,384	48,384	47,520	2,159
R^2	0.240	—	—	0.482

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Results incorporate time, destination and origin fixed effects.

Table 8: Well-Educated Immigrant Men

Independent variables	Two-Part Model			
	$\ln(1 + M_{d,o})$	Eaton-Tamura	Probit	$\ln(M_{d,o})$
$\hat{Y}_{t,s}$: GDP trend	0.006 (0.004)	0.006 (0.006)	0.0007 (0.0005)	0.040 (0.013)***
$\Delta Y_{t,s}$: GDP fluctuations	0.025 (0.006)***	0.014 (0.008)	0.0014 (0.0007)*	-0.014 (0.017)
$\ln(Dist_{d,o})$	-0.525 (0.133)***	-0.356 (0.074)***	-0.0273 (0.0047)***	-0.180 (0.129)
$\ln(Pop_{t,o})$	0.645 (0.197)***	1.162 (0.285)***	0.0953 (0.0237)***	-1.154 (0.616)*
$\ln(Pop_{t,d})$	0.701 (0.313)**	0.544 (0.307)*	0.0722 (0.0258)***	-0.738 (0.703)
$\ln(1 + Stock_{t,d,o})$	0.016 (0.003)***	0.058 (0.007)***	0.0047 (0.0005)***	0.063 (0.013)***
Observations	48,384	48,384	46,656	4,350
R^2	0.286	—	—	0.416

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Results incorporate time, destination and origin fixed effects.

Table 9: Push and Pull Factors for Less-Educated Immigrant Men

Independent variables	Two-Part Model			
	$\ln(1 + M_{d,o})$	Eaton-Tamura	Probit	$\ln(M_{d,o})$
$\hat{Y}_{t,d}$: Destination GDP trend	0.004 (0.005)	-0.006 (0.013)	0.0000 (0.0003)	0.012 (0.023)
$\hat{Y}_{t,o}$: Origin GDP trend	-0.027 (0.005)***	-0.0442 (0.016)***	-0.0018 (0.0004)***	-0.053 (0.033)
$\Delta Y_{t,d}$: Destination GDP fluctuations	0.025 (0.006)***	0.048 (0.015)***	0.0015 (0.0004)***	0.015 (0.026)
$\Delta Y_{t,o}$: Origin GDP fluctuations	-0.001 (0.009)	0.030 (0.030)	0.0008 (0.0008)	0.062 (0.054)
$\ln(Dist_{d,o})$	-0.710 (0.122)***	-0.551 (0.112)***	-0.0109 (0.0027)***	-0.569 (0.176)***
$\ln(Pop_{t,o})$	0.374 (0.174)**	0.992 (0.539)*	0.0188 (0.0139)	0.821 (0.975)
$\ln(Pop_{t,d})$	-0.262 (0.270)	-0.647 (0.548)	-0.0095 (0.0141)	-1.936 (0.931)**
$\ln(1 + Stock_{t,d,o})$	0.037 (0.003)***	0.140 (0.016)***	0.0038 (0.0003)***	0.086 (0.021)***
Observations	48,384	48,384	47,520	2,609
R^2	0.326	—	—	0.584

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Results incorporate time, destination and origin fixed effects.