# The Short- and Long-Run Determinants of Unskilled Immigration into U.S. States

Nicole B. Simpson *
Colgate University

Chad Sparber †
Colgate University

August 24, 2010

**Abstract**

This paper uses a gravity model of migration to analyze how income differentials affect the flow of immigrants into U.S. states. We add to existing literature by decomposing income differentials into short- and long-term components and by focusing on newly arrived unskilled immigrants between 2000-2008. Our sample is unique in that 95 percent of our observed immigrant flows equal zero. The trade literature has advocated using the Eaton and Tamura (1994) threshold Tobit model in similar settings, and we are the first to apply the methodology to analyze the determinants of immigration. We find that recent U.S. immigrants positively respond to differences in long-term (or trend) GDP between origin countries and U.S. states. When appropriately accounting for the zero values, we also find that differences in GDP fluctuations significantly affect the flow of unskilled immigrants. In addition, we find that short-run GDP fluctuations pull unskilled immigrants into certain U.S. states, whereas GDP levels push unskilled immigrants out of their countries of origin.

Keywords: Immigration, Macroeconomics, GDP, Gravity

JEL Codes: F2, E01, J61

* Corresponding Author, Department of Economics, 13 Oak Drive, Hamilton, New York, 13346; (315)228-7991, Fax: (315)228-7033, nsimpson@colgate.edu.

† Department of Economics, 13 Oak Drive, Hamilton, New York, 13346; (315)228-7967, Fax: (315)228-7033, csparber@colgate.edu.

# 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 et al (2009), Mayda (2010), and Peri and Ortega (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 a particular group of immigrants, namely newly arrived unskilled U.S. immigrants. We add to this literature in three ways: (1) by analyzing recent inflows of unskilled immigrants into U.S. states between 2000-2008; (2) by decomposing GDP into short- and long-run components; and (3) by employing the Eaton and Tamura (1994) threshold Tobit method of estimation.

First, we analyze the flow of new immigrants into U.S. states between 2000-2008 using U.S. Census and American Community Survey (ACS) data. Our work complements the literature that focuses on the locational choice of new immigrants based on state-specific factors such as the generosity of welfare programs, unemployment rates, and the existing stock of immigrants (for example, Bartel, 1989; Zavodny, 1997; Borjas, 1999). These papers often assess the demographic characteristics of immigrants as a potential determinant 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 study the flow of newly arrived unskilled immigrants: immigrants who legally or illegally arrived in the U.S. within the past year (from the survey date) and have a high school degree or less. We focus on this group of immigrants since the immigration debate in the U.S. is especially contentious about them; according to Mayda (2006), unskilled immigrants likely generate the largest negative political reactions by the U.S. public.

Second, we decompose GDP differentials into short- and long-term components to study their differential effects on unskilled immigrant flows. We are aware of no other study that does this. Variation across 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 immigrant flows to macroeconomic cycles (exceptions include 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 et al (2008) and Mayda (2010). This allows us to assess whether unskilled immigrants leave countries who 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 a gravity model of immigration in the spirit of Karemera et al. (2000), Mayda (2010), Lewer and Van den Berg (2008), and Ortega and Peri (2009). However, we employ a number of techniques, including the Eaton and Tamura (1994) threshold Tobit model – a method that, to our knowledge, has not yet been used to analyze the determinants of immigration.[1] The use of this model is necessitated by unique features of our data. Specifically, we observe annual bilateral gross flows of unskilled workers into each U.S. state from 120 different source countries. More than 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. However, this problem of zero flows in a gravity model is quite common in the trade literature and we therefore appeal to that literature for alternative estimation techniques. We first estimate our gravity model using a standard OLS regression that drops all observations with zero immigrant flows. We next add one to each observed flow to include all of the observations in OLS. Lastly, we apply the Eaton and Tamura (1994) method and compare the set of significant predictors across the various techniques.

Our results indicate that long-run GDP differentials between origin countries and U.S. states are significant determinants of unskilled immigrant flows, and this is robust to model specification and technique. We also find that fluctuations in GDP positively affect immigrant flows but only when the entire sample of nonnegative immigrant flows are considered. The results from the Eaton-Tamura technique confirm this finding. Our results are robust to different samples (i.e., whether or not to include female immigrants) and sets of control variables (including time, origin and destination fixed effects). We extend our analysis by decomposing the effects into push and pull factors and and find that long-run GDP trends push unskilled immigrants out of their origin country. We also document a cyclical effect such that recent booms in U.S. states attract unskilled immigrants from abroad.

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 describe our data in detail. Finally, we present the results and discuss how they add to existing literature.

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

# 2 Empirical Strategy

## 2.1 Theoretical Motivation

The canonical theoretical model of migration consists of an income maximization problem where the potential immigrant $i$ from origin country $o$ chooses the destination $d$ based on the relative returns to migrating to that country after factoring out migration costs. Assume there is a continuum of agents of type $i$, a discrete number of origin countries $o = \{1, 2, ..., O\}$, and a discrete number of destination countries $d = \{1, 2, ..., D\}$. The origin country $o$ is in the set of destination countries since the agent may choose to remain in their country of origin.

The immigrant's decision is based on the income differential between the destination and origin country $(Y_{i,d} - Y_{i,o})$ net of migration costs $(C_{i,d,o})$. Specifically, the utility of agent $i$ from origin country $o$ migrating to destination $d$ is:

$$U_{i,d,o} = u(Y_{i,d} - Y_{i,o} - C_{i,d,o})$$

where $u(\cdot)$ is a strictly increasing, continuous function. 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), and costs that are individual-specific (i.e., family members left back home).

The agent chooses the destination $d$ that maximizes his utility:

$$\underset{d=\{1,...,D\}}{max}\{U_{i,d,o}\} \tag{1}$$

or equivalently,

$$max\{U_{i,1,o}, U_{i,2,o}, ..., U_{i,D,o}\}.$$

Since there is a continuum of agents, the problem in equation (1) for agent $i$ can be rewritten as:

$$\underset{\lambda_{i,d,o}}{max}\{\sum_{d=1}^{D} \lambda_{i,d,o} U_{i,d,o}\} \tag{2}$$

where $\lambda_{i,d,o}$ represents the probability that agent $i$ from country $o$ migrates to destination $d$. Agents randomize over the set of possible destinations in equilibrium so that the probability that agent $i$ from country $o$ migrates to destination $d$ is equivalent to the fraction of that type living in destination $d$. Since each type of agent has to live somewhere, the sum of these probabilities is 1, or $\sum_d \lambda_{i,d,o} = 1$. Let $\bar{U}_{i,\bar{d},o}$ represent the level of utility that agent $i$ from country $o$ receives when living in his/her optimal destination $\bar{d}$.

Following the work of Grogger and Hanson (2008) and Ortega and Peri (2009), we assume a linear utility function. Thus, the optimal utility level for agent $i$ from country $o$ is: $\bar{U}_{i,\bar{d},o} = Y_{i,\bar{d}} - Y_{i,o} - C_{i,\bar{d},o}$. Given a distribution of agents $i \in f(i)$, we can derive an aggregate utility function for the migration decision by integrating over $i$. Aggregate utility $U$ in this economy is:

$$U = \int_i (Y_{i,\bar{d}} - Y_{i,o} - C_{i,\bar{d},o})di \tag{3}$$

Since the mass of each type $i$ individual is 1, the number of individuals from country $o$ that migrate to each destination $d$ can be represented by $M_{d,o}$, where:

$$ln(M_{d,o}) = Y_d - Y_o - C_{d,o}. \tag{4}$$

Thus, immigrant flows depend on the aggregate income differential between the destination and origin net of moving costs. It is important to note that researchers use either aggregate measures of income (i.e., GDP) or micro-level measures of income (i.e., wages) in models of immigration. We choose the former since we are assessing how unskilled 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. For a recent discussion of this issues, we refer the reader to Rosenzweig (2007).

## 2.2 Empirical Specification

Equation 4 motivates our empirical specification – a gravity model of immigration, similar to Karemera et al. (2000), Mayda (2010), Lewer and Van den Berg (2008), and Ortega and Peri (2009). More generally, 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, $Y_{t,d} - Y_{t,o}$. Migration costs are proxied using time-invariant (e.g., geographic) controls ($X_{d,o}$) and factors that vary over time and origin country ($W_{t,d,o}$). Year fixed effects ($FE_t$) account for time trends, $\varepsilon_{t,d,o}$ is the error term, and $\alpha$, $\beta$, $\delta$, and $\eta$ are the coefficients to be estimated. The empirical specification is:

$$\begin{aligned} \ln(M_{t+1,d,o}) = \alpha + \beta \cdot (Y_{t,d} - Y_{t,o}) \\ + \delta \cdot X_{d,o} + \eta W_{t,d,o} + FE_t + \varepsilon_{t,d,o} \end{aligned} \tag{5}$$

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

$$Y_{t,c} = a_c + b_c \cdot t + e_{c,t} \tag{6}$$

where $e_{c,t}$ is an error term.

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, equation 5 can be written as:

$$\begin{aligned} \ln(M_{t+1,d,o}) = \alpha + \beta_1 \cdot (\hat{Y}_{t,d} - \hat{Y}_{t,o}) + \beta_2 \cdot (\Delta Y_{t,d} - \Delta Y_{t,o}) \\ + \delta \cdot X_{d,o} + \eta W_{t,d,o} + FE_t + \varepsilon_{t,d,o} \end{aligned} \tag{7}$$

Our dataset includes immigrant flows between 120 origin countries and the 48 contiguous U.S. states. We follow the literature in identifying control variables. First, we include the natural log of the distance between the origin country's capital city and the state's geographic center ($Dist$). State-specific time-invariant controls include dummy variables if the destination state borders Mexico ($SBorder$) or Canada ($NBorder$), and a vector of U.S. regional dummy variables ($Region$). Country-specific factors include indicators for those bordering the U.S. ($Border$), countries whose primary language is English ($English$), and those that were colonized by the U.S. ($Colony$). Time-variant factors include the natural log of the stock of the of immigrants 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, $Pop_{t,o}$.

$$\begin{aligned} \ln(M_{t+1,d,o}) = {} & \alpha + \beta_1 \cdot (\hat{Y}_{t,d} - \hat{Y}_{t,o}) + \beta_2 \cdot (\Delta Y_{t,d} - \Delta Y_{t,o}) \\ & + \delta_1 \ln(Dist_{d,o}) + \delta_2 SBorder_d + \delta_3 NBorder_d + \delta_4 Region_d \\ & + \delta_5 Border_o + \delta_6 English_o + \delta_7 Colony_o \\ & + \eta_1 \ln(Stock_{t,d,o}) + \eta_2 \ln(Pop_{t,d}) + \eta_3 \ln(Pop_{t,o}) + FE_t + \varepsilon_{t,d,o} \end{aligned} \tag{8}$$

Note that we do not directly control for U.S. immigration policy, as Clark, Hatton and Williamson (2007) advocate. U.S. immigration policy is decided at the national level so that our study does not require destination-specific policy variables. However, changes in U.S. immigration policy over time warrants the inclusion of time fixed effects in our specification. We are therefore indirectly controlling for immigration policy.

## 2.3 Estimation Technique

Gravity models were first used by trade economists to analyze bilateral export and import flows. Immigration economists have since adopted them to help identify the determinants of migration. The characteristics and limitations of the gravity model are therefore shared by these two fields. We turn to developments in the trade literature for a solution to an issue confronting our dataset – how to select and estimate a model when a large proportion of observed flow values equal zero. Evidence from international trade suggests that the Eaton and Tamura (1994) threshold Tobit model is ideal for such a scenario. We believe we are the first to employ the model to estimate the determinants of immigration.

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. Estimation problems arise when country pairs experience zero trade flows since log values are undefined. This is a nontrivial issue in trade. 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).

One simple attempt to overcome this limitation is to estimate a truncated model (i.e., drop observations of zero flows). Another method – sometimes called 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 and censoring log-values less than zero. More complicated solutions include the Poisson Pseudo Maximum Likelihood (PPML) estimator, the Heckman two-step model, and the Eaton and Tamura threshold Tobit.

Given the importance of zero flows in the trade literature, economists have begun to evaluate the efficacy of alternative strategies. Martin and Pham (2008) provide a recent and thorough comparison of these models when zero values are frequent. 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.[2]

[2] Moreover, 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

A well-known paper by Santos Silva and Tenreyro (2006) advocates the 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. 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).

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.[3] Some eschew the gravity model and instead measure flows or emigration rates in levels (not logs).[4] 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.

In our dataset of immigrant flows from origin countries to U.S. states, we encounter values of zero in roughly 95% of the observations. In this regard, our data is not unlike the Baldwin and Harrigan (2007) HS10 U.S. import data. Our problem is perhaps more severe than that experienced by most immigration economists, but quite similar to that confronted in trade. Thus, we turn to the trade literature for a solution. Motivated by Martin and Pham (2008), we advocate the Eaton and Tamura methodology.

Eaton and Tamura (1994) introduced the threshold Tobit model to analyze 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 adopt the common solution of adding one to each value before taking logs, they added $\lambda$, a value to be statistically estimated.[5]

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

$$M_{t,d,o} = \max\left\{0, \tilde{M}_{t,d,o}\right\} \tag{9}$$

The latent variable $\tilde{M}_{t,d,o}$ is a function of several determinants of migration ($X_{t,d,o}$), a

among observations with positive values. It is difficult to imagine such a variable.

[3]See Lewer et. al. (2009), Lewer and Van den Berg (2008), Peri and Ortega (2009), or Falck et. al. (2010) for recent examples.

[4]See Adsera and Pytlikova (2010), Pedersen et. al. (2008), Zavodny (1997), or Mayda (2010).

[5]Head 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.

mean-zero normally-distributed error term ($\varepsilon_{t,d,o}$), and a threshold value ($\lambda$) that the function of explanatory variables must achieve before positive migration flows occur.[6]

$$\tilde{M}_{t,d,o} = -\lambda + \exp\left(\alpha + \beta \cdot X_{t,d,o} + \varepsilon_{t,d,o}\right) \tag{10}$$

By substituting equation (10) into equation (9), rearranging, and taking natural logs, we derive equation (11). Eaton and Tamura (1994) provide the density function for $\tilde{M}_{t,d,o}$ and the necessary log-likelihood function for maximum likelihood estimation.[7] Thus,

$$\ln\left(\lambda + M_{t,d,o}\right) = \begin{cases} \alpha + \beta \cdot X_{t,d,o} + \varepsilon_{t,d,o} \text{ if } \tilde{M}_{t,d,o} > 0 \\ \ln\left(\lambda\right) \text{ otherwise} \end{cases} \tag{11}$$

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 analysis of immigration's role in promoting international trade. To our knowledge, however, we are the first to apply the technique to a gravity model of the determinants of immigration.

The model presents only two limitations. 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 state-by-country analysis. Second, it can be difficult to interpret coefficient estimates. The literature has responded to this by adopting a number of strategies. Eaton and Tamura (1994) note that because of the intercepts, slope coefficients converge to elasticities only asymptotically. Head and Ries (1998) make a similar point, but argue that coefficients do not differ much from true elasticities as long as the dependent variable is positive. They therefore choose to interpret magnitudes as true elasticities. Rauch and Trindade (2002), in contrast, decline to offer any magnitude interpretation, and instead limit their discussion to whether effects are positive, negative, or insignificant. It is worth noting that the common SOLS solution of adding one to zero values also prohibits a strict elasticity interpretation of coefficients, though it typically does not prevent economists from doing so. In our analysis, we will continue to interpret magnitudes as elasticities, despite the previously cited caveats.

[6] Head and Ries (1998) interpret $\lambda$ as undermeasurement.

[7] We are indebted to Cong S. Pham for kindly providing Stata code for the procedure.

## 3 Data

We focus our analysis on foreign-born workers with a high-school degree or less; they represent 60% of new U.S. immigrants in the 2000-2009 period (based on Census/ACS data). We consider only those who are employed in the U.S. at the time of survey.[8] We first analyze the flow of male immigrants, but then incorporate female immigrants (who are employed) into our analysis (in Section 4.2).

We use data from the 2000-2008 Census and ACS surveys (IPUMS) to obtain annual estimates for the gross inflow of new foreign workers in 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., and immigrants are identified by state of residence. The dataset is not without its limitations, however. First, it 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 first arrived in the U.S. within the last year.[9] 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.

For our main explanatory variables, $(\hat{Y}_{t,d} - \hat{Y}_{t,o})$ and $(\Delta Y_{t,d} - \Delta Y_{t,o})$, we use per capita real GDP (in 2000 dollars) for origin countries using data from the U.S. Department of Agriculture.[10] Per capita state GDP data (in 2000 chained dollars) are 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 GDP per capita). We have eight years of data since we lag all of the independent variables. During this period, U.S. immigrants came from 120 different countries. Therefore, we have a total of 46,080 observations for our state-level analysis (48*8*120). However, only 2,234 observations have non-zero immigrant flows.

Our control variables include geographic indicators, destination and origin populations,

[8] However, all of our results are robust to including unskilled immigrants who are not employed.

[9] It is important to note that our gross inflow measure does not include foreign workers who previously arrived in the U.S. but recently moved to a new state in the U.S. Thus, our 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.

[10] http://www.ers.usda.gov/Data/Macroeconomics/

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 include males and females). The geographic indicators include the distance between world capitals and the geographic centers of states, using the Haversine distance formula and latitude/longitude data from the CEPII Research Center (for cities) and the U.S. Census (for state centers).[11] U.S. region indicators include the West Coast, Gulf Coast, East Coast and former confederate states (which are not mutually exclusive) and are defined in the Appendix. Population estimates are from the U.S. Department of Agriculture (for countries) and the U.S. Census (for U.S. states). We report the mean and standard deviation of each variable in Table 1. The average inflow of unskilled males is 1,159 per year for each state (for observations with positive immigrant flows) and exhibits tremendous variation. When including the observations with zero flows, average immigrant flows are 56 unskilled males per year in each state.

GDP differentials, both long-term and short-run, are the independent variables of interest. Average trend per capita GDP of U.S. states is \$34,211 (with very little variation), while the average value for origin countries is \$8,902 (with high variation). Average fluctuations in per capita GDP would equal zero by construction if we used the entire time series. Since the independent variables are lagged, however, we lose the last year of data. The resulting averages are negative for origin countries (-0.3%) and positive for U.S. states (4.1%), leading to a 4.4% gap in GDP fluctuations between the destination and origin of immigrants. Variation in GDP fluctuations is very high relative to the mean, with more of the variation coming from state fluctuations than from origin countries.

As for other control variables, the average distance between origin countries and U.S. states is 5,354 miles. Approximately 30% of the sample represents countries that speak English, and just under 2% were a former U.S. colony. Approximately 8% of the observations are destination states that border Mexico (i.e., four of the 48 states border Mexico), while 25% of the sample (12 states) borders Canada. The region dummies are not mutually exclusive, and thus do not sum to one. Origin countries have an average of 4,693 residents in each state (at all education levels) who have been in the U.S. for more than one year, with substantial variation across states (standard deviation is 53,749). States have on average a population of 4.3 million while origin countries have a population of 47.1 million.

[11] http://www.cepii.fr/anglaisgraph/bdd/distances.htm

Table 1: Summary Statistics

| **Variables** | **Mean** | **St Dev** |
|---|---|---|
| Migrant Flows $M_{d,o} \geq 0$ | 56.18 | 902.32 |
| Migrant Flows $M_{d,o} > 0$ (2,234 observations) | 1158.77 | 3939.92 |
| $\hat{Y}_{t,d}$: State trend (in thousands) | 34.211 | 6.216 |
| $\hat{Y}_{t,o}$: Country trend (in thousands) | 8.902 | 12.214 |
| $\hat{Y}_{t,d} - \hat{Y}_{t,o}$: GDP trend differential (in thousands) | 25.309 | 13.661 |
| $\Delta Y_{t,d}$: State fluctuations (in thousands) | 0.041 | 0.517 |
| $\Delta Y_{t,o}$: Country fluctuations (in thousands) | -0.003 | 0.244 |
| $\Delta Y_{t,d} - \Delta Y_{t,o}$: GDP fluctuations differential (in thousands) | 0.044 | 0.544 |
| $Dist_{d,o}$ (in miles) | 5,354 | 2,234 |
| Origin country is Mexico or Canada | 0.017 | 0.128 |
| *English* | 0.308 | 0.462 |
| *Colony* | 0.017 | 0.128 |
| State borders Mexico | 0.083 | 0.276 |
| State borders Canada | 0.250 | 0.433 |
| West Coast | 0.063 | 0.242 |
| Gulf Coast | 0.104 | 0.305 |
| East Coast | 0.292 | 0.455 |
| Former Confederate | 0.229 | 0.420 |
| $Stock_{t,d}$ | 4,693 | 53,749 |
| $Pop_{t,o}$ (in millions) | 4.330 | 4.590 |
| $Pop_{t,d}$ (in millions) | 47.110 | 153.685 |
| Observations | 46,080 | |

# 4 Results

We model the flow of immigrants from origin country $o$ to destination state $d$ in year $t$ as specified in equation 8. Recall that we focus on unskilled male immigrants who are employed. Table 2 presents the baseline results. The specification in the first three columns does not control for immigrant stocks, while the last three columns do. Within each specification, we consider three estimation techniques: in column 1, we consider only non-zero immigrant flows and estimate equation 8 using (truncated) OLS; in column 2, we include all immigrant flows by adding one to the the flow variable before taking the natural log and then employing OLS (i.e., SOLS); in column 3, we use the Eaton-Tamura technique as described in section 2.3. The sample size is much smaller in column 1 compared to columns 2-3 since observations with zero immigrant flows are dropped. The same comparisons are made in columns 4-6 but here we control for immigrant stocks.[12]

Evidence from the truncated regression (column 1) favors a significant effect on immigration flows due to differences in long-run GDP – a \$1,000 increase in the long-run GDP per capita gap correlates with an 2.4% increase in the flow of unskilled immigrants. GDP fluctuations, in contrast, demonstrate no effect on gross immigrant flows. This suggests that for countries with positive immigrant flows, long-term GDP differentials matter for immigrant flows, but fluctuations do not. Put differently, individuals are motivated to leave poor countries for richer U.S. states, but short-term fluctuations in state and country per-capita GDP are not relevant for migration decisions. The remaining control variables have the expected sign when significant and are consistent with the literature (i.e., Karemera et al., 2000; Lewer and Van den Berg, 2008; Ortega and Peri, 2009). Countries that are farther away from the U.S. send fewer immigrants to the U.S., larger countries send more immigrants, and larger U.S. states attract more immigrants from abroad.

The remaining columns attest to the importance of accounting for country-state pairs with zero immigrant flows. Inclusion of these observations results in finding that short-run fluctuations in GDP per capita also determine immigrant flows. In column (2), SOLS suggests that a \$1,000 increase in the cyclical gap between state and origin country GDP per capita induces a 1.7% increase in unskilled immigrant flows (significant at the 10% level). In column (3), the Eaton-Tamura technique provides even stronger evidence that short-run fluctuations in state GDP affect immigrant flows: the coefficient is larger in magnitude and significant at the 5% level. Thus, it seems as if both short- and long-run GDP differentials between the sending country and the destination state determine unskilled immigrant flows into U.S. states so long as regressions include observations with zero flows.

Our results are similar when controlling for immigrant stocks (columns 4-6). Long-term GDP differences are robust determinants of immigrant flows across all three techniques. Short-run fluctuations remain significant when applying the Eaton-Tamura method, but not in other specifications. In addition, immigrant flows from countries that have a larger immigrant stock in the U.S. state send more people to that state. This result has been frequently documented in the literature as the "friends and relatives effect" (Bartel; 1989; Zavodny 1997; Clark, Hatton and Williamnson, 2007; Grogger and Hanson, 2008; Mayda, 2010). Since immigrant stocks are important determinants of flows, we include them in the analysis that follows.

In sum, our initial findings indicate that appropriately dealing with the large quantity of observations with zero immigrant flows by using the Eaton-Tamura method is important in

[12] In the cases when one is added to the independent variable, we add one to the immigrant stock variable before taking natural logs.

identifying which factors determine immigrant flows. When the zero values are not considered, only long-run differences in GDP are positively correlated with immigrant inflows in the U.S. However, when zero values are included, both long- and short-run GDP differentials are positively associated with the flow of unskilled male immigrants into U.S. states.

Table 2: Baseline Results

| Independent variables | Without Stocks $ln(M_{d,o})$ | Without Stocks $ln(1+M_{d,o})$ | Without Stocks Eaton-Tamura | With Stocks $ln(M_{d,o})$ | With Stocks $ln(1+M_{d,o})$ | With Stocks Eaton-Tamura |
|---|---|---|---|---|---|---|
| $\hat{Y}_{t,d}-\hat{Y}_{t,o}$: GDP trend | 0.024 | 0.008 | 0.015 | 0.020 | 0.010 | 0.016 |
| | (0.003)*** | (0.001)*** | (0.002)*** | (0.002)*** | (0.001)*** | (0.002)*** |
| $\Delta Y_{t,d}-\Delta Y_{t,o}$: GDP fluctuations | -0.044 | 0.017 | 0.050 | -0.048 | 0.014 | 0.050 |
| | (0.036) | (0.010)* | (0.024)** | (0.037) | (0.010) | (0.024)** |
| $\ln(Dist_{t,s})$ | -0.456 | -0.383 | -0.836 | -0.225 | -0.284 | -0.530 |
| | (0.065)*** | (0.039)*** | (0.070)*** | (0.064)*** | (0.037)*** | (0.053)*** |
| $\ln(Pop_{t,o})$ | 0.086 | 0.096 | 0.258 | -0.010 | 0.048 | 0.125 |
| | (0.020)*** | (0.008)*** | (0.022)*** | (0.021) | (0.008)*** | (0.016)*** |
| $\ln(Pop_{t,d})$ | 0.434 | 0.188 | 0.486 | 0.190 | 0.083 | 0.182 |
| | (0.053)*** | (0.012)*** | (0.042)*** | (0.058)*** | (0.012)*** | (0.031)*** |
| $\ln(Stock_{t,d})$ | — | — | — | 0.235 | — | — |
| | — | — | — | (0.024)*** | — | — |
| $\ln(1+Stock_{t,d})$ | — | — | — | — | 0.058 | 0.183 |
| | — | — | — | — | (0.003)*** | (0.019)*** |
| Observations | 2,234 | 46,080 | 46,080 | 2,152 | 46,080 | 46,080 |
| $R^2$ | 0.46 | 0.18 | — | 0.51 | 0.20 | — |

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Dummy variables for contiguity, English language, US colony, border country, US regions and time fixed effects are included in all specifications but their estimated coefficients are not reported.

## 4.1 Destination and Origin Fixed Effects

Up until now, we have not included destination or origin fixed effects. We now include a full set of fixed effects for both the 48 destination states and 120 origin countries in addition to time fixed effects (which were previously included). The results are provided in Table 3.

Evidence for an effect of long-run per capita GDP differentials is mixed. When only the nonzero immigrant flows are considered, state and country indicators eliminate all effects from long-run GDP. The same is true with the Eaton and Tamura method (column 3). Only the SOLS model (column 2) continues to find any evidence for the importance of long-run GDP gaps in the presence of state and country fixed effects.

The disparity in the estimates of this coefficient across Tables 2 and 3 could reflect the component of long-term GDP that drives migration decisions. In Table 3, the coefficient is identified only by differences in trend growth rates across states and countries ($b_c$ in equation

(6) construction of our trend variable). Fixed effects absorb differences in permanently high levels of GDP per capita ($a_c$ in equation 6).[13]If long-term GDP levels are more important than growth rates in determining immigration, the estimates in Table 2 would be larger than those in Table 3.

In contrast to the evidence for long-run GDP effects, coefficient estimates for short-run GDP differentials when including state and country fixed effects are quite comparable to baseline results (in Table 2, columns 4-6). While the truncated OLS model continues to argue against an effect on immigrant flows (in column 1), models that include observations with zero immigrant flows (columns 2-3) find that short-run GDP fluctuations are positively correlated with immigrant flows into U.S. states. In the case of the SOLS model (column 2), the estimated effect actually becomes significant (it was insignificant in the baseline case). Together, these results are consistent with our earlier conclusion that the choice of how to account for zero values can have significant impact on the set of variables that are significant predictors of cross-country immigrant flows.

Table 3: Destination/Origin Fixed Effects

| Independent | Dest/Origin FE | | |
|---|---|---|---|
| variables | $ln(M_{d,o})$ | $ln(1+M_{d,o})$ | Eaton-Tamura |
| $\hat{Y}_{t,d} - \hat{Y}_{t,o}$: GDP trend | -0.011 | 0.018 | -0.008 |
| | (0.025) | (0.006)*** | (0.014) |
| $\Delta Y_{t,d} - \Delta Y_{t,o}$: GDP | -0.017 | 0.019 | 0.044 |
| fluctuations | (0.036) | (0.009)** | (0.022)** |
| $\ln(Dist_{t,s})$ | -0.435 | -0.692 | -0.551 |
| | (0.147)*** | (0.128)*** | (0.119)*** |
| $\ln(Pop_{t,o})$ | 1.343 | 0.432 | 1.048 |
| | (1.236) | (0.172)** | (0.554)* |
| $\ln(Pop_{t,d})$ | -2.371 | -0.004 | -0.140 |
| | (1.154)** | (0.246) | (0.580) |
| $\ln(Stock_{t,d})$ | 0.248 | — | — |
| | (0.029)*** | — | — |
| $\ln(1+Stock_{t,d})$ | — | 0.032 | 0.122 |
| | — | (0.003)*** | (0.016)*** |
| Observations | 2,152 | 46,080 | 46,080 |
| $R^2$ | 0.63 | 0.32 | — |

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Dummy variables for contiguity, English language, US colony, border country, US regions and time fixed effects are included in all specifications but their estimated coefficients are not reported.

[13]We should also note that time trends would account for each observation's GDP growth if we had restricted growth rates (bc) to be equal across states and countries. Thus, the GDP Trend coefficient in Table 3 is only identifiable because we allow for state and country-specific trends.

## 4.2 Male and Female Migrant Flows

Thus far we have considered only male immigrant flows. This is standard in the immigration literature when trying to isolate those immigrants that move for economic purposes (recall that we consider only employed immigrants). However, recent discussions on the topic indicate that women are increasingly migrating for economic reasons compared to earlier cohorts (United Nations, 2004). We now include women so that the dependent variable is the total number of employed men and women migrating to the U.S. This increases the number of nonzero observed flows from 2,152 to 2,905. Table 4 reports the results, state and country fixed are not included in columns 1-3 but are included in columns 4-6.

The results are very similar to the baseline regressions reported in Tables 2 and 3. We find strong evidence that differences in long-run GDP positively affect immigrant flows into U.S. states and some evidence that short-run deviations in GDP determine flows. The only difference when including female immigrants is that GDP fluctuations are now significant at the 10% level using SOLS estimation (column 2) when fixed effects are not included (when controlling for immigrant stocks); it was not significant when only male flows were considered.

With fixed effects, the results are qualitatively identical to Table 3. Only the SOLS method finds a significant coefficient on long-run GDP differentials, while both the SOLS and Eaton and Tamura specifications uncover large effects from short-run (i.e., cyclical) differentials. In sum, our main conclusions are insensitive to whether or not the immigrant flow variable includes employed women. This indicates that male and female immigrants react to income differentials similarly.

Table 4: Male and Female Flows

| Independent variables | no FE $ln(M_{d,o})$ | no FE $ln(1+M_{d,o})$ | no FE Eaton-Tamura | FE $ln(M_{d,o})$ | FE $ln(1+M_{d,o})$ | FE Eaton-Tamura |
|---|---|---|---|---|---|---|
| $\hat{Y}_{t,d}-\hat{Y}_{t,o}$: GDP trend | 0.022 | 0.011 | 0.016 | 0.019 | 0.023 | -0.009 |
| | (0.002)*** | (0.001)*** | (0.002)*** | (0.021) | (0.007)*** | (0.013) |
| $\Delta Y_{t,d}-\Delta Y_{t,o}$: GDP fluctuations | -0.051 | 0.021 | 0.055 | -0.025 | 0.029 | 0.048 |
| | (0.031) | (0.011)* | (0.022)** | (0.031) | (0.011)*** | (0.020)** |
| $\ln(Dist_{t,s})$ | -0.239 | -0.331 | -0.523 | -0.383 | -0.795 | -0.710 |
| | (0.055)*** | (0.039)*** | (0.047)*** | (0.125)*** | (0.139)*** | (0.119)*** |
| $\ln(Pop_{t,o})$ | -0.009 | 0.069 | 0.143 | 1.060 | 0.473 | 0.845 |
| | (0.019) | (0.009)*** | (0.015)*** | (0.981) | (0.213)** | (0.522) |
| $\ln(Pop_{t,d})$ | 0.131 | 0.116 | 0.194 | -1.254 | -0.069 | -0.339 |
| | (0.048)*** | (0.013)*** | (0.028)*** | (0.929) | (0.296) | (0.528) |
| $\ln(Stock_{t,d})$ | 0.268 | — | — | 0.274 | — | — |
| | (0.021)*** | — | — | (0.027)*** | — | — |
| $\ln(1+Stock_{t,d})$ | — | 0.076 | 0.186 | — | 0.041 | 0.111 |
| | — | (0.004)*** | (0.016)*** | — | (0.003)*** | (0.012)*** |
| Observations | 2,905 | 46,080 | 46,080 | 2,905 | 46,080 | 46,080 |
| $R^2$ | 0.51 | 0.21 | — | 0.61 | 0.31 | — |

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Dummy variables for contiguity, English language, US colony, border country, US regions and time fixed effects are included in all specifications but their estimated coefficients are not reported.

## 4.3 Push and Pull Factors

In the specification in equation 8, we assumed that the coefficients on the destination and origin country GDP are the same. Empirically, it is not necessary to impose this restriction. We now separate these effects in the specification below:

$$\begin{aligned}\ln(M_{t+1,d,o}) &= \alpha + \beta_1\hat{Y}_{t,d} - \beta_2\hat{Y}_{t,o} + \beta_3\Delta Y_{t,d} - \beta_4\Delta Y_{t,o} \\ &+ \delta_1 \ln(Dist_{d,o}) + \delta_2 SBorder_d + \delta_3 NBorder_d + \delta_4 Region_d \\ &+ \delta_5 Border_o + \delta_6 English_o + \delta_7 Colony_o \\ &+ \eta_1 \ln(Stock_{t,d,o}) + \eta_2 \ln(Pop_{t,d}) + \eta_3 \ln(Pop_{t,o}) + FE_t + \varepsilon_{t,d,o}\end{aligned} \tag{12}$$

We can now differentiate between push and pull factors for both long-term and short-run GDP variables. Estimated coefficients for $\beta_2$ and $\beta_4$ will determine if income in the origin countries are pushing immigrants out, while estimates for $\beta_1$ and $\beta_3$ will identify whether income in the destination states are pulling immigrants in. The results (for male and female immigrants) are reported in Table 5. Columns 1-3 do not include state and country fixed effects, while the results in columns 4-6 include them.

Tables 2-4 presented mixed evidence on the influence of GDP differentials on migration decisions. Table 5 demonstrates that if such an effect exists, it is clearly driven by long-term GDP in the origin country, and not in the destination state. The coefficient on state GDP trend  is insignificant in all six specifications. The coefficient on origin country GDP per capita trend is negative and significant in all but the Eaton and Tamura regression with state and country fixed effects. It therefore appears that rich countries send fewer immigrants to the U.S. than poor countries do, but rich U.S. states do not necessarily attract more immigrants than poor states.

Short-run fluctuations have a different effect on immigrant flows. As in Tables 2-4, regressions in Table 5 uncover effects only if they include observations with zero flows. In those specifications, U.S. states that have experienced an economic boom attract immigrants from abroad. Conversely, GDP per capita fluctuations in origin countries do not push immigrants out of their home country.

Thus, long-run (or trend) GDP determines which countries send immigrants to the U.S., while short-run fluctuations determine which U.S. state they decide to move to, suggesting that immigrants are being pushed out of poor 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.

Table 5: Push and Pull Factors

| Independent variables | Push&Pull | | | Push&Pull w/FE | | |
|---|---|---|---|---|---|---|
| | $ln(M_{d,o})$ | $ln(1+M_{d,o})$ | Eaton-Tamura | $ln(M_{d,o})$ | $ln(1+M_{d,o})$ | Eaton-Tamura |
| $\hat{Y}_{t,d}$: Destination | -0.002 | 0.002 | 0.005 | 0.004 | 0.013 | -0.024 |
| GDP trend | (0.006) | (0.002) | (0.003) | (0.027) | (0.010) | (0.016) |
| $\hat{Y}_{t,o}$: Origin | -0.026 | -0.012 | -0.017 | -0.063 | -0.036 | -0.018 |
| GDP trend | (0.002)*** | (0.001)*** | (0.002)*** | (0.037)* | (0.008)*** | (0.021) |
| $\Delta Y_{t,d}$: Destination | -0.046 | 0.022 | 0.065 | -0.012 | 0.034 | 0.055 |
| GDP fluctuations | (0.033) | (0.013)* | (0.023)*** | (0.033) | (0.012)*** | (0.021)*** |
| $\Delta Y_{t,o}$: Origin | 0.159 | -0.021 | 0.015 | 0.157 | -0.003 | 0.015 |
| GDP fluctuations | (0.101) | (0.017) | (0.063) | (0.102) | (0.017) | (0.049) |
| $\ln(Dist_{t,s})$ | -0.220 | -0.327 | -0.512 | -0.386 | -0.796 | -0.712 |
| | (0.054)*** | (0.039)*** | (0.046)*** | (0.126)*** | (0.139)*** | (0.119)*** |
| $\ln(Pop_{t,o})$ | -0.015 | 0.066 | 0.138 | 0.586 | 0.355 | 0.215 |
| | (0.019) | (0.009)*** | (0.015)*** | (1.073) | (0.214)* | (0.022)*** |
| $\ln(Pop_{t,d})$ | 0.144 | 0.121 | 0.201 | -1.393 | -0.116 | -0.427 |
| | (0.046)*** | (0.013)*** | (0.028)*** | (0.932) | (0.295) | (0.531) |
| $\ln(Stock_{t,d})$ | 0.274 | — | — | 0.273 | — | — |
| | (0.020)*** | — | — | (0.027)*** | — | — |
| $\ln(1+Stock_{t,d})$ | — | 0.079 | 0.189 | — | 0.041 | 0.110 |
| | — | (0.004)*** | (0.016)*** | — | (0.003)*** | (0.012)*** |
| Observations | 2,905 | 46,080 | 46,080 | 2,905 | 46,080 | 46,080 |
| $R^2$ | 0.51 | 0.21 | | 0.61 | 0.31 | |

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Dummy variables for contiguity, English language, US colony, border country, US regions and time fixed effects are included in all specifications but their estimated coefficients are not reported.

# 5 Immigration into the U.S.

Next, we generalize our model to consider immigrant flows into the U.S., but not into particular U.S. states. Thus, the only destination is the U.S. (i.e., $d = US$), turning equation 8 into:

$$\begin{aligned}\ln(M_{t+1,US,o}) = \alpha &+ \beta_1 \cdot (-\hat{Y}_{t,o}) + \beta_2 \cdot (-\Delta Y_{t,o}) \\ &+ \delta_1 \ln(Dist_{US,o}) + \delta_5 Border_o + \delta_6 English_o + \delta_7 Colony_o \\ &+ \eta_1 \ln(Stock_{t,US,o}) + \eta_3 \ln(Pop_{t,o}) + FE_t + \varepsilon_{t,US,o}\end{aligned} \tag{13}$$

Since time fixed effects are included and all immigrants go to the same destination, the terms $\hat{Y}_{t,US}$ and $\Delta Y_{t,US}$ drop out of the specification, as do the destination-specific control variables (i.e., *Region* and $Pop_{t,US}$). Thus, the model is able to identify push factors only.

All of the other variables are defined as above, with the exception of $Dist_{US,o}$ which now represents the distance between the origin country's largest city and the District of Columbia.

Table 6 reports the results for the same three specifications as above for our sample of unskilled male immigrants. Once again, we have 120 origin countries with 8 years of data, yielding a sample size of 960. Column (1) drops the zero values (which in this case represents 360 observations); column (2) adopts SOLS with the entire sample; and column (3) employs the Eaton-Tamura method. In all three cases, we find that long-term GDP differentials are robust determinants of immigrant flows into the U.S. A \$1,000 increase in the permanent GDP differential between the U.S. and the origin country is associated with a 3-4% increase in gross immigrant flows into the U.S., depending on the specification. This effect is identified by push factors, or equivalently, a \$1,000 decline in the origin country's GDP trend. GDP per capita fluctuations, in contrast, demonstrate no effect on gross immigrant flows. This suggests that individuals are motivated to leave poor countries for the U.S., but short-term fluctuations are not relevant for cross-country migration decisions. This is entirely consistent with the results from our state-by-country analysis.

Variations of the model were considered, including alternative combinations of the control variables, and the results are robust to these variations (but are not reported for brevity). One exception is the inclusion of country fixed effects, which yields insignificant coefficients on our variables of interest. We suspect that country fixed effects absorb too much data variation and inhibit the model's ability to identify any significant effects.[14] We also express the caveat that, unlike for the state-by-country analysis, the Eaton and Tamura method sometimes encounters difficulty in converging to a solution at the national level. We therefore express more confidence in the state-by-country analysis but feel this analysis is important to include since this is the strategy taken by most economists in estimating the determinants of U.S. immigration.

[14] In addition, when female immigrants are included, the results are inconclusive.

Table 6: National-Level Results

| Independent variables | No Dest/Origin FE $ln(M_{d,o})$ | $ln(1+M_{d,o})$ | Eaton-Tamura |
|---|---|---|---|
| $\hat{Y}_{t,s}$: GDP trend | 0.032<br>(0.006)*** | 0.044<br>(0.011)*** | 0.033<br>(0.006)*** |
| $\Delta Y_{t,s}$: GDP fluctuations | -0.135<br>(0.156) | 0.151<br>(0.314) | -0.110<br>(0.212) |
| $\ln(Dist_{t,s})$ | -0.457<br>(0.232)*** | -0.946<br>(0.318)*** | -0.323<br>(0.214)*** |
| $\ln(Pop_{t,o})$ | 0.071<br>(0.073) | 0.375<br>(0.100)*** | 0.063<br>(0.072)*** |
| $\ln(Stock_{t,d})$ | 0.500<br>(0.094)*** | —<br>— | —<br>— |
| $\ln(1+Stock_{t,d})$ | —<br>— | 1.140<br>(0.113)*** | 0.562<br>(0.100)*** |
| Observations | 600 | 960 | 960 |
| $R^2$ | 0.54 | 0.44 | — |

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are in parentheses. Dummy variables for contiguity, English language, US colony, and time fixed effects are included in all specifications but their estimated coefficients are not reported.

# 6 Conclusion

This paper adds to the literature on the determinants of immigrant flows in three ways. First, we use variation in GDP across U.S. states to uncover how newly arrived unskilled immigrants respond to income differentials. Second, we decompose income differentials into short- and long-run components. Third, we employ several estimation techniques, including the Eaton and Tamura threshold Tobit model – a methodology from the trade literature appropriate for datasets containing a large number of zero values for immigrant flows. We find that this specification is important to accurately identify the primary determinants of unskilled immigration.

We find that both long-term and short-run GDP differentials are significant determinants of the flow of newly arrived unskilled immigrants into U.S. states. However, the evidence for long-term GDP differentials is mixed when state and country fixed effects are included. In addition, the evidence for short-run differentials requires that observations of zero flow values are included in the regression. We also add to the immigration literature that attempts to disentangle push and pull effects. We find that unskilled immigrants are being pushed out of their countries by long-run GDP trends, and are pulled into U.S. state by short-run upswings in economic activity. Not surprising, short-run fluctuations in the origin country do not lead to an increase in unskilled 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.


# Acknowledgements

The authors thank Piyush Chandra, Giovanni Peri, Cong S Pham, Todd Sorensen, and seminar participants at Colgate University and the 2010 WEAI in Seattle for detailed comments.


# Appendix

Table 7: List of States by Region

| **Region** | **States** |
|---|---|
| West Coast | CA, OR, WA |
| Gulf Coast | AL, FL, LA, MS, TX |
| East Coast | CT, DE, FL, GA, ME, MD, MA, NH, NJ, NY, NC, RI, SC, VA |
| Former Confederate | AL, AR, FL, GA, LA, MS, NC, SC, TN, TX, VA |

# References


[1] Adsera, Alicia and Mariola Pytlikova (2010). “The Role of Language in Shaping International Migration: Evidence from OECD Countries 1985-2006,” Mimeo.

[2] Baldwin, Richard and James Harrigan (2007) “Zeros, Quality, and Space: Trade Theory and Trade Evidence” NBER Working Paper 13214.

[3] Bartel, Ann P. (1989), “Where Do the New United States Immigrants Live?” *Journal of Labor Economics* 7 (October): 371–91.

[4] Borger, S. (2008). ”Estimates of the Cyclical Inow of Undocumented Migrants to the United States, ” University of California San Diego Working Paper.

[5] Borjas, George (1999). “Immigration and Welfare Magnets,” Journal of Labor Economics, October 1999, pp. 607-637.

[6] Borjas, George (2001). “Does Immigration Grease the Wheels of the Labor Market?” Brookings Papers on Economic Activity, pp. 69-119.

[7] Clark, Ximena, Timothy Hatton, Jeffrey Williamson. 2007. “Explaining U.S. Immigration, 1971-1998.” *Review of Economics and Statistics*, 89(2): 359-373.

[8] Davis, S.J. and Haltiwanger, J. (1992). “Gross Job Creation, Gross Job Destruction, and Employment Reallocation.” *The Quarterly Journal of Economics*, 107(3): 819-863.

[9] 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.

[10] Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Sudekum (2010), “Dialects, Cultural Identity, and Economic Exchange,” IZA Discussion Papers 4743.

[11] Grogger, Jeffrey & Gordon H. Hanson, 2008. ”Income Maximization and the Selection and Sorting of International Migrants,” NBER Working Papers 13821.

[12] 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.

[13] Head, Keith and John Ries (1998) “Immigration and Trade Creation: Econometric Evidence from Canada” *Canadian Journal of Economics*, Vol. 31 (1), pp. 47-62.

[14] 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.

[15] Karemera, David, Victor I. Oguledo, & Bobby Davis, 2000. "A Gravity Model Analysis of International Migration to North America." *Applied Economics*, 32 (13): 1745–55.

[16] Kugler, Adriana and Mutlu Yuksel (2008). "Effects of Low-Skilled Immigration on U.S. Natives: Evidence from Hurricane Mitch," NBER Working Paper 14293.

[17] Lewer, Joshua J. & Hendrik Van den Berg, 2008. "A Gravity Model of Immigration." *Economics Letters*, April 2008, v. 99, iss. 1, pp. 164-67.

[18] Lewer, Joshua J., Gail Pacheco and Stephani 2009. "Do Non-Economic Quality of Life Factors Drive Immigration?" IZA DP No. 4385.

[19] Mandelman, Federico and Andrei Zlate (2010. "Immigration, Remittances and Business Cycles," Federal Reserve Bank of Atlanta Working Paper.

[20] 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.

[21] Mayda, Anna Maria (2006). "Who Is Against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants," *The Review of Economics and Statistics*, MIT Press, vol. 88(3), pp.510-530.

[22] Mayda, Anna Maria, 2010. "International migration: A panel data analysis of the determinants of bilateral flows." *Journal of Population Economics,* 23 (4): 1249-1274.

[23] 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.

[24] Ortega, Francesc and Giovanni Peri, 2009. "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005." NBER Working Paper 14833.

[25] Rauch, James E. and Vitor Trindade (2002) "Ethnic Chinese Networks in International Trade" *The Review of Economics and Statistics*, Vol. 84 (1), pp. 116-130.

[26] Rosenzweig, Mark (2007). "Education and Migration: A Global Perspective." Yale University, Mimeo.

[27] 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.

[28] 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

[29] Warin, T., and Svaton, P. 2008. European Migration: Welfare Migration or Economic Migration? *Global Economy Journal*, 8(3), 1-30.

[30] Zaiceva, A., and Zimmermann, K. F. 2008. Scale, Diversity, and Determinants of Labour Migration in Europe. *Oxford Review of Economic Policy*, 24(3), 427-451.

[31] Zavodny, Madeline (1997) "Welfare and the Locational Choices of New Immigrants" *Economic Review – Federal Reserve Bank of Dallas*, Second Quarter 1997, pp. 2-10.