

# U.S. Migration Patterns as a Predictor for State Patterns and Sentiments

Richard Holt\*

May 9, 2019

## 1 Introduction

The purpose of this research is to assess the relationship between immigration and economic growth. As economies around the world develop more diverse and structured markets, questions about immigration policies arise regularly. Economists have long debated the merits of an open border policy with regards to a country's comparative advantages. It is intuitive to think that a correlation would exist between fluid travel and economic growth, but the goal of this research is to determine which variable is leading. Past research bears out the relationship between immigration and economic variation, but it is almost impossible to determine which variable depends on the other. The important question that this research will attempt to answer is whether the relationship between the two at the national level is a good model for the relationship at the state level. If it is not a good model, and if states do deviate significantly, the goal of the analysis is to

---

\*Department of Economics, University of Oklahoma. E-mail address: holt@ou.edu

understand whether the relationship clearly mirrors or follows the sentiments about immigration and immigration policy that the state holds generally.

## **2 Literature Review**

The literature on the topic of immigration's relationship with economic patterns is extensive, but it was important to focus on studies related to the sentiments and policies towards immigration and how they could be explained through economic indicators. Poul Schou conducted a study on the economic sustainability of immigration based on the overall costs of integration. Schou used a model known as DREAM to compute the general equilibrium of immigrant households over the course of seventeen generations. The DREAM model helped Schou understand how immigration changes the demographics of an area and how the initial costs of immigration compare to the long run costs. Schou's study concluded that the fiscal sustainability of immigration is affected differently depending on what type of country the immigrant is coming from—decreasing when the immigrant is from an underdeveloped country and slightly increasing when they are from a developed country.[4] Understanding this relationship provides context for the sentiments that incumbent citizens hold about immigrants based on where they are from, and notably their reason for immigrating initially. The reason for immigrating is a factor built into the model of this research.

Fiscal sustainability aside, case studies indicate that immigration tends to have a positive effect on the income of the host country. The costs of integration are the largest detriment to fiscal sustainability, and alleviating them often results in large gains for the host country. Mete Feridun studied this relationship in The Case of Norway, using the Augmented Dickey-Fuller unit root

test and Granger causality test to understand correlation and causality of the time series data and test how stationary the data is. The Granger test provided better results for the stable data in this experiment as opposed to the VAR model, which would incorrectly assess the causal relationship for data with a common trend. Since the goal of this research is to understand the correlation between the sentiments and the deviations in data, a regular OLS model was sufficient, because the deviations did not follow a common thread. Feridum concluded that the positive effect of immigration could be observed so long as integration costs were minimal or ignored.[2] Again, understanding localities and demographics that would cause variance in the integration costs is important for the purpose of this research, as Feridum points out that low-skill immigrants from under-developed countries tend to have higher integration costs.

Two studies were referenced to understand how labor flows have changed and affected the U.S. economy. Hanson assessed the labor flows in their article, *The Rise and Fall of U.S. Low-Skilled Immigration*. Low-skilled immigration flows over extended periods of time have several important characteristics that determine the sentiments of the native populations as well as the short-term public policy to address immigration. Low-skilled workers, they found, tend to become specialized in specific types of work as well as having higher labor force participation rates than native populations. Immigration ebbs and flows with the growth of the national economy or the state economies, and high immigration is easily associated with high U.S. incomes. The study found that changes to the U.S. immigration policy necessarily effected the economies of other nations, given that the mobility of immigrants in and out of the countries would be influential on how the other countries developed accordingly.[3] When large, immigrant-heavy industries in the U.S. are doing well, it should not be surprising to see immigration increase, but it can be surprising to see how the public and the short-term policy makers react.

The second study referenced on the topic of labor skill was conducted by Nicole Simpson and Chad Sparber. The two researched the costs of low-skilled immigrant labor compared to high-skilled labor, using census data between 2000 and 2009. The relationship was modeled using an SOLS model, an Eaton and Tamura threshold tobit, and two part models. By exploring the demographics of the immigrant populations in addition to the fluctuations in GDP on the national level, the study was able to model the relationship between the skill-level of immigrants and the relative costs. The study found that fluctuations in GDP tend to positively affect the immigration of less-educated men and that long and short-term trends affected the patterns in different ways. Long-run patterns, they found, act as a push for immigrants. While potential immigrants do not necessarily focus much on the long-run trends of GDP in the country they are going to, long-run decline in their domestic economy acts as a major impetus for immigration away. Short-run trends, on the other hand, act as a pull for immigrants, drawing them into the growing economy. While people from poor economies want to immigrate to the United States, the short-term fluctuations in their domestic economy are unlikely to affect their decision as much as a short-term boom in the destination economy.[5] Knowing these temporal factors provides a basis for assessing data across short periods of time. It is unnecessary, for example, to model the relationship over the course of 50 to 100 years, when data may not be available. Rather, the relationship seen over the course of a decade would be sufficient to accommodate short-term booms and busts and the accompanying immigration flows. The important element of the data for this research is the conditions within the distribution, the nature of smaller state economies with ties and correlations to the national economy, and it supports the proposition that the model will effectively provide context for sentimental deviations.

More information is needed regarding the interactions of immigrants within certain sectors of

the economies, as it is difficult to generalize economies and their relationships to immigration sentiments even at the state level. To alleviate the concern of generalization, the different states were modelled against the national economy in the same way to ensure that those with uniquely strong and weak sectors would be balanced by differences between the states. For example, if some states had strong energy sectors at a time when immigration was more attuned to construction, or vice versa, the imbalance reflected will be diluted by the inverse effect occurring in the other states. If many states reflect the same strong or weak sectors, the impact should be universal regardless. Rafaela Dancygier and Michael Donnelly studied the relationship between immigration and sectoral economies to try to understand the significance of the distinctions that can be made between sectors. Their hypothesis was that the native populations would oppose immigration when there was a downturn in their own sector. They found that skill-based wage effects alone are insufficient to explain attitudes about immigration and that the cultural factors tend to be more important. The importance of cultural factors aligns well with the understanding that integration costs are significant determinants of immigration sentiments. The research found that immigrants tend to cluster in certain sectors of the economy and that high and low-skilled workers respond to the changes in the economy in similar ways, meaning that sectors only have to be understood as separate, but not unique.[1] The universality of the responses within sectors substantiates using the state-level modelling as quasi-sectors within the national economy.

### **3 Data**

Modelling the national relationship between economic growth and immigration required simple, time series data on GDP growth rates and immigration patterns. Data for GDP growth since 1947

was obtained from the St. Louis Federal Reserve Economic Data website. The data for economic growth did not require extensive cleaning or sorting, but the model could not be established using direct nominal GDP as the state levels would be small proportions. Rather, the changes in the annual rate of growth were used as the economic growth indicator. Data for national immigration was obtained from the U.S Census. Two types of data were used to assess the relationship. The first provided the aggregate numbers for individuals migrating and the nature of their migration. The data differentiated based on who was moving permanently and who was not, who was moving between countries or states as opposed to within them, and who was a U.S. migrant. The important types of immigration for the purposes of this research include those moving permanently and those moving between countries. The second type of data was data regarding the purpose of the migration. Reasons include looking for work, moving to begin an already acquired job, changing households, and other family and personal reasons. The important reasons for the sake of this research are those looking for new work and those with new jobs already acquired. The immigration data has similarly been modelled using the rates of immigration instead of the nominal figures. Assessing the changes within the greater distribution allows the model to assess variation as it relates in-flows and out-flows as opposed to the bulk of immigrants coming and going.

Applying the national model to the state data did not require data from as far back as 1947. In fact, the literature on the matter indicated that shorter time periods of data may be better assessed in short intervals with only a few booms or busts. State GDP data was obtained from the St. Louis Federal Reserve Data website as far back as 1997. The GDP data was broken into individual states, and a few were selected for the purposes of this modelling including Texas, Vermont, Ohio, and California. The state GDP data was treated the same as the national data to maintain consistency in the model. State migration data was more complicated to collect and was obtained from the

U.S. Census Bureau. The Census collects state migration data per yer between all 50 states, and the data was also manipulated to represent rates instead of aggregate numbers.

## 4 Methods

The OLS regression model was used to model the relationship between the national immigration rates and the national GDP rates. The OLS model is appropriate for this data as opposed to the Granger model because the data does not have similar patterns within it. While the OLS model can sometimes lead to spurious correlation and false causation, the Granger analysis is only important when looking at a wide number of variables and determining the causality between the variables. The purpose of this research is only to identify deviance between relationships as they exist, whether they are causal relationships or not, and to then test the correlation between the predictions and the sentiments and how they deviate from the national scale to the state scale. It is also appropriate to assume a linear relationship in the data given how close economic growth and immigration are intuitively. Whether one has a causal relationship with the other, it is true that growth will attract immigrants and vice versa, while it is also true that labor flows necessarily affect economic progress. The following equations were used in establishing the model.

General Model

$$Y_{GDP} = \beta_0 + \beta_1 * MigrationRate + \beta_2 * DifferentState + \beta_3 * DifferentCountry + \epsilon \quad (1)$$

### Skilled Labor Model

$$Y_{GDP} = \beta_0 + \beta_1 * MigrationRate + \beta_2 * NewJob + \beta_3 * DifferentCountry + \varepsilon \quad (2)$$

### Unskilled Labor Model

$$Y_{GDP} = \beta_0 + \beta_1 * MigrationRate + \beta_2 * LookWork + \beta_3 * DifferentCountry + \varepsilon \quad (3)$$

The variation within the migration variable, coupled with it being the variable of interest for sentiment purposes, made it the appropriate independent variable for this regression. The deviation by state away from the national GDP predictor is appropriate as it should result in no significant difference regarding the leading and following effects of the variables.

There were a couple of variables that were important to include regarding the migration data including the reason for migration and the type of migration. The important data for the purpose of this research were those immigrants travelling between countries and who already had jobs or who were specifically looking for work. The use of two models allows a distinct look at the effects of and sentiments towards skilled labor as opposed to unskilled labor. Controlling for this is important to ensure that the effects, if significant, are not lost in the data, given the importance stressed by the literature. While responses to economic factors should be the same between the two generally, the condition itself is an important qualifier for which it is best to control.

The most difficult part of data collection was the collection of state migration patterns. While



future research would do well to include more in-depth information about the state migration patterns and the reasons behind them, it was sufficient for the sake of these initial findings to focus on the numbers and the variance from the prediction. The U.S. Census data was also extremely chaotic, and the formatting of the data would change from year to year in a way that was difficult to clean with any universality. While more time spent with the cleaning process would likely have yielded at least 15 good years of time series data, the literature review was clear that findings are appropriate in the short term as well. For this reason, the data from 2005-2009 that was easily organized was the focus of this research.

Once the aggregate data from 2005-2009 had been collected, the strategy was to organize a table by state-to-state movement and look at the totals for each year. The data was then reduced down to the annual rates so as to be compared to the national rates and the rates of GDP growth. Ignoring the "intra-state" migration which ballooned the nominal figures, the data was able to be organized and assessed across a simple time series. Since the state data was not broken down in terms of reasons for movement, it was impossible to determine whether the people moving were doing so because of work, including whether they were looking for work or if they already had a job set up.

Finally, to provide a baseline for the variance within the skilled and unskilled models, the general national data was used without the intentions as a controlling variable in the model. Instead, only interstate and international flow were used to control against the general rates. Doing this provides a few different observations of the predictive value of the data to make up for some of the shortage of total data.

## 5 Findings

The surface level relationship between GDP and migration rates was largely intuitive, but nonetheless interesting in considering how the two variables are related. Figure 1 shows the graphical relationship between the migration of bona fide movers as a percent of total migrants and GDP growth over the course of the past 60 years. Much of the relationship shown likely reflects an upper limit on what that portion of immigration will ultimately be, but it also reflects a temporal shift in the sentiments towards immigration generally.

Figures 2 and 3 show the relationship broken into the movement between countries and states, respectively. The relationship is similar, which is expected, but again reinforces the likelihood that the relationship is more asymptotic than anything else. It would not be an easy jump to say that shifting attitudes towards immigrants from other countries would translate to a nativism between states, but perhaps that is a relationship to be explored by future research. Ultimately, for the purpose of this research, the relationship shown in the graphs confirms that, at the national level, immigration as a whole responds to the national economy in similar ways, be that between countries or states. For this reason, interstate migration data can be appropriately used as the test data to understand the migration relationship between national and state data.

Figures 4 and 5 show the relationships between migrants who have jobs in place before their move as opposed to those who are coming specifically to find work, respectively. These groups have been defined as skilled workers and unskilled workers, respectively. The relationships here are very interesting in that, over the shorter time period of about 20 years, the two variables are inverse. Skilled labor actually hit its valley as a share of total migration just before the 2008 economic crash while unskilled labor peaked just before 2008 and then dropped sharply afterward.

The findings for unskilled labor align with the literature review's understanding of push and pull effects, but the skilled labor seems to be just the opposite. Obviously, these findings are just over a short period of immigration, and more research could be done to assess the marginal relationship around migration before and after economic crises.

The general model was run using the aforementioned equation and the time series data for the past 70 years. The model provided the optimal parameters based on the general OLS regression. The optimal parameters of the skilled and unskilled models are also determined using the OLS regression because it is appropriate for the research and it maintains consistency between the three models. Table 1 shows the optimal parameters and outcomes for the three different models. It was important to note the significant difference in the relationship between the skilled and unskilled optimal values for movement between country and state as opposed to the general model which did not account for the different. In general, the general model seemed to aggregate the relationships fairly well and in interesting ways, showing that the optimal value for general migration was negative despite the small, positive optimal values for the interstate and international movement. The unskilled and skilled breakdown revealed the underlying connection, showing the strong positive optimal value for unskilled movers overall, which was reasonable given the relationship between unskilled migration and the economic crisis shown in Figure 5. However, the positive relationship that existed was almost entirely tied to interstate movement, while international movement for both was weakly negative. These results likely speak to a relationship between intra-national migrants and a different effect that they have than international migrants, and this is something for which future research should account.

After determining the optimal values, the models themselves showed the relationships between the variables at each level. The significance of these factors was not extremely important for

the purposes of this research, as only the predictive value and deviance from the predictions was strongly considered. Especially given the short period of time series data, it would not be appropriate even to assume that tests of significance were durable to future research, but the literature review suggests that it is appropriate for predictive purposes. The models are provided in Tables 2, 3, and 4 for the general model, skilled model, and unskilled model, respectively. Again, an analysis of the significance of these models is largely unnecessary given that the sentiments of people in these states is not related to actual effect on GDP, only on how the two change relative to each other.

Finally, the models for the individual states were run and tested against the state models, finding the deviance for each state used in the model. For this research, only a small number of states were used to present data, and these states were selected at random (that is, not based on the results the provided). Texas, Vermont, Ohio, California, New York, Florida, Wyoming, and Oklahoma were selected to sample the results, although all 52 states and territories could be used if they were added to the code. The deviance away from the general model for each state is shown in Tables 5 and 6, as well as the sentiments from Charts 1 and 2. These sentiment analyses were conducted by the PRRI in 2015 and were based on the approval of the cultural differences in newcomers (Chart 1) and whether natives believe that migrants should be able to attain citizenship after meeting certain requirements (Chart 2). At face value, for the test states over the course of a small period of time, the relationship for this sentiment data does not seem to be all that significant. While the sentiments between the two different analyses are generally consistent, they do not tend to correlate well with the deviance from the general model. The results are surprising especially considering the variation in the state deviations and how clear the differences were between each. However, even between states like Wyoming, California, and Ohio, which had very distinct deviations and

consistent sentiments, there was no clear relationship in terms of their sentiment deviation.

## **6 Conclusion**

The relationship described in this research is not easily established between the migration model's deviation from the national model and the sentiments about immigration at the state level. While there are relationships that can be established internally in terms of the deviations from the general model as well as the modelling of the national relationship based on different immigration factors.

There are several avenues that future research could take to address the issues underlying the results of this research. Rather than looking at immigration patterns as shares of total immigration, the data could be easily manipulated to determine the relationship between GDP rates and general migration rates to see if the results are significantly different. Data broken down at this level would be better able to account for immigration shocks, while continuing to control for the types of immigration allows a closer look at policy-based impacts. Since the United States likely does not have to worry about major policy shocks from things like quotas and caps on total immigration, there would not be many confounders involved with the aggregate data the way that there is with the proportional data.

Future research should also look to extend the time period associated with both the general model and the state models. Immigration data is only so accurate as it extends further into the past, but at least having data for a time period closer to 20 years would provide a more concrete look at the long-term relationship. Additionally, creating a time series with this kind of appropriate length could allow researchers to identify relationships on the margins of economic shocks like that observed at 2008 in the data here. Longer time series data would also allow researchers to

match the time series for the national model with the time series of the state-level model. Doing this will help increase the predictive value of the model since it will be accounting for the same shocks and levels in both.

The main issues to be addressed with the research here are the non-parallel, potentially insufficient time series data and the type of migration data used in the model comparison. For the purposes of drawing additional conclusions about the sentiments, it would be prudent to conduct further analysis on the sentiment data and potentially spend time developing independent data to be more specific. While the sentiment figures shown are just one example of a potential look at state-by-state sentiment, it is important to find sources that are as objective as possible. The sentiment analysis in this case is difficult to assess the general validity of, and the questions asked in the survey could likely be tailored better to have a more comprehensive understanding of where the sentiments are coming from and how related the sentiments are to movement.

Similarly, time series data for sentiments would be important to match with the time series of the GDP and migration data. The sentiment data presented was only for 2015, but sentiments should change with the changes in GDP and migration over the course of several years, especially given the short time period upon which this research relies. Ultimately, it will be important to find data that can be validated and consistently matched with the GDP and migration data across a time series.

The important takeaways from the research presented here are the methods for developing such models for testing the relationship between national and state patterns and using that information to assess future data. The shortcomings of the data presented here are easily addressed with more time spent finding, cleaning, organizing, and correlating data between the national level, the state level, and the sentiment data. It will also be important to further develop a model considering all

52 states and territories and creating an objective sentiment analysis framework.

## Figures and Tables

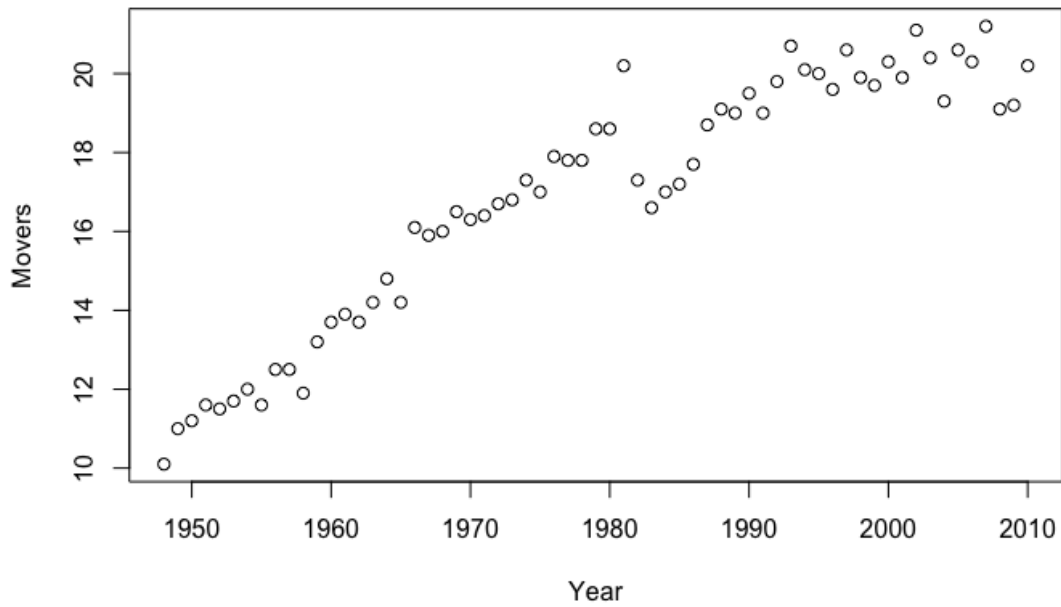


Figure 1: Percent migrants United States 1948-2017



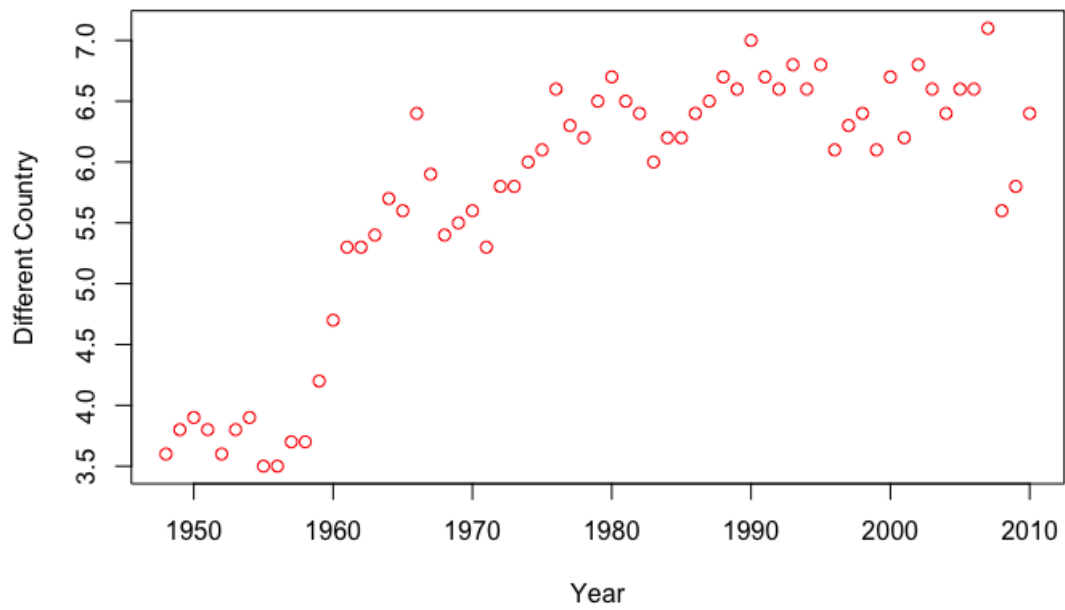


Figure 2: Percent international migration United States 1948-2017

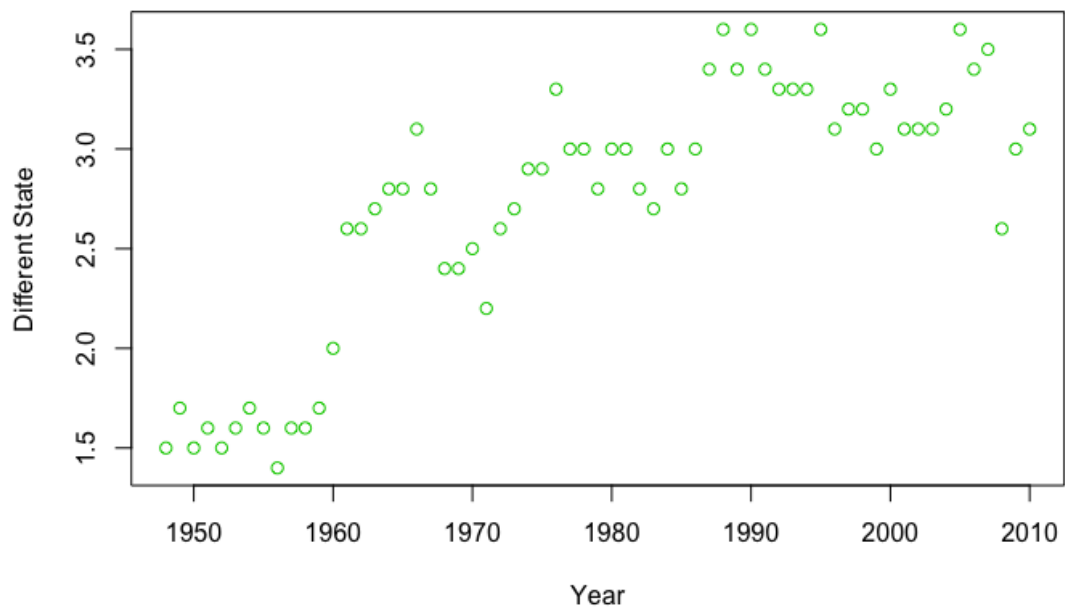


Figure 3: Percent interstate migration United States 1948-2017

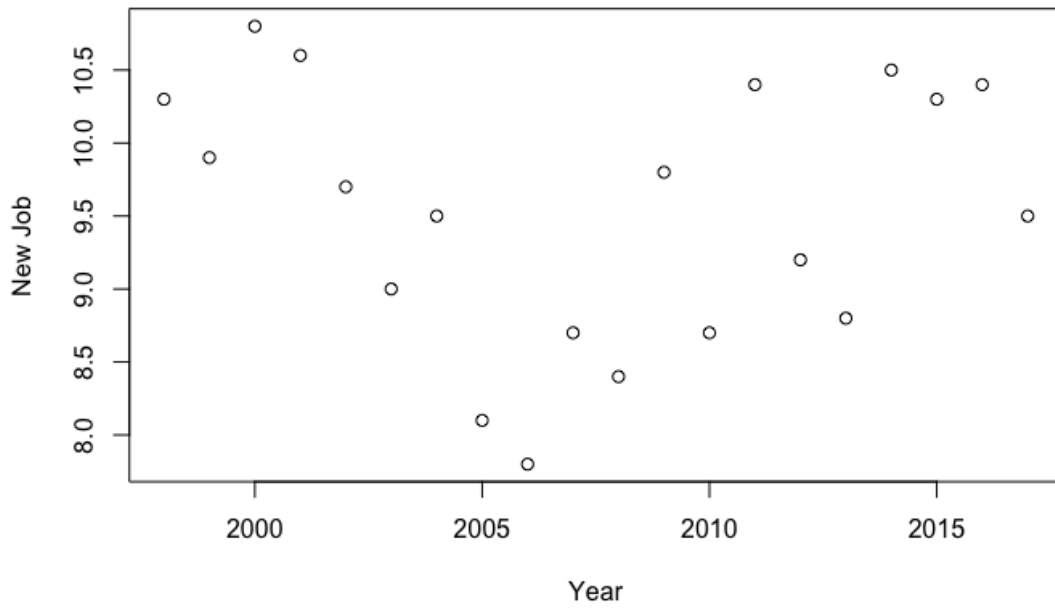


Figure 4: Percent skilled migrants United States 1998-2017

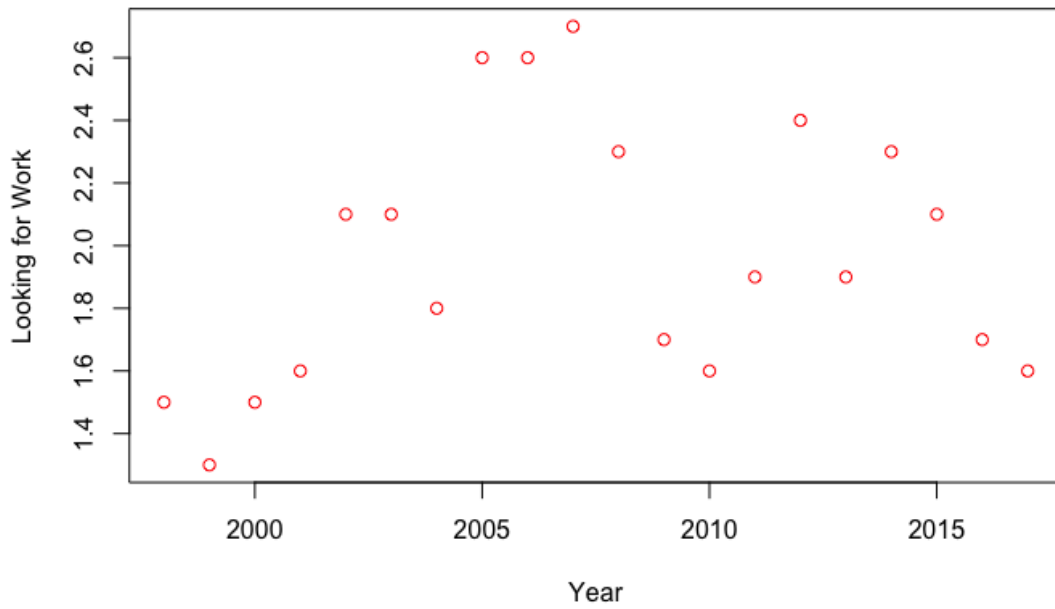


Figure 5: Percent unskilled migrants United States 1998-2017

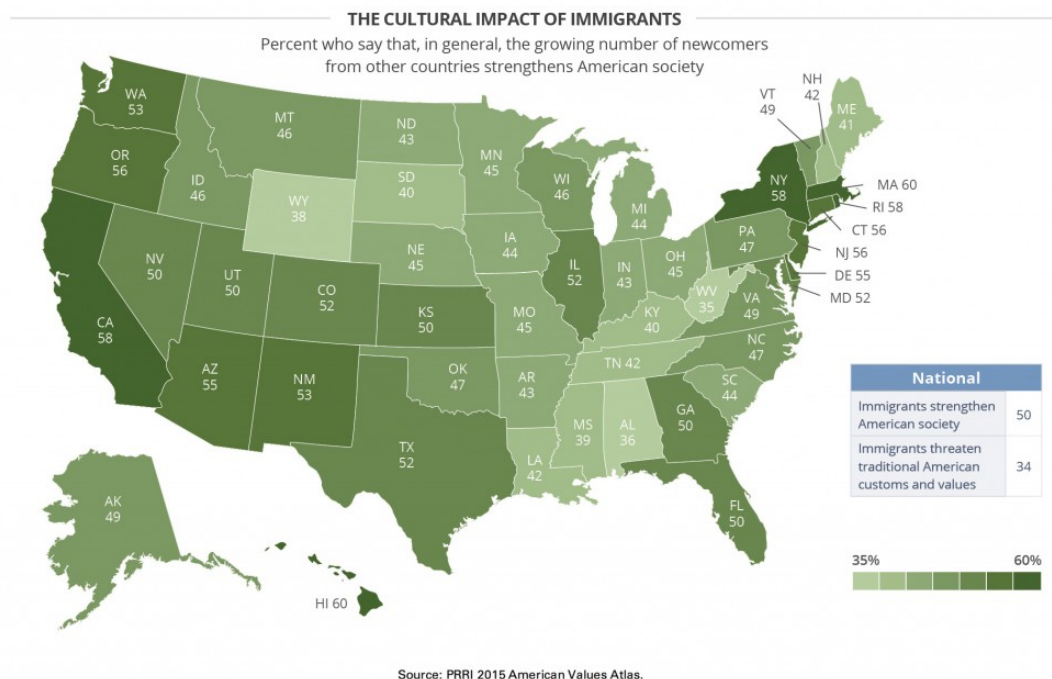


Figure 6: Immigration sentiment I

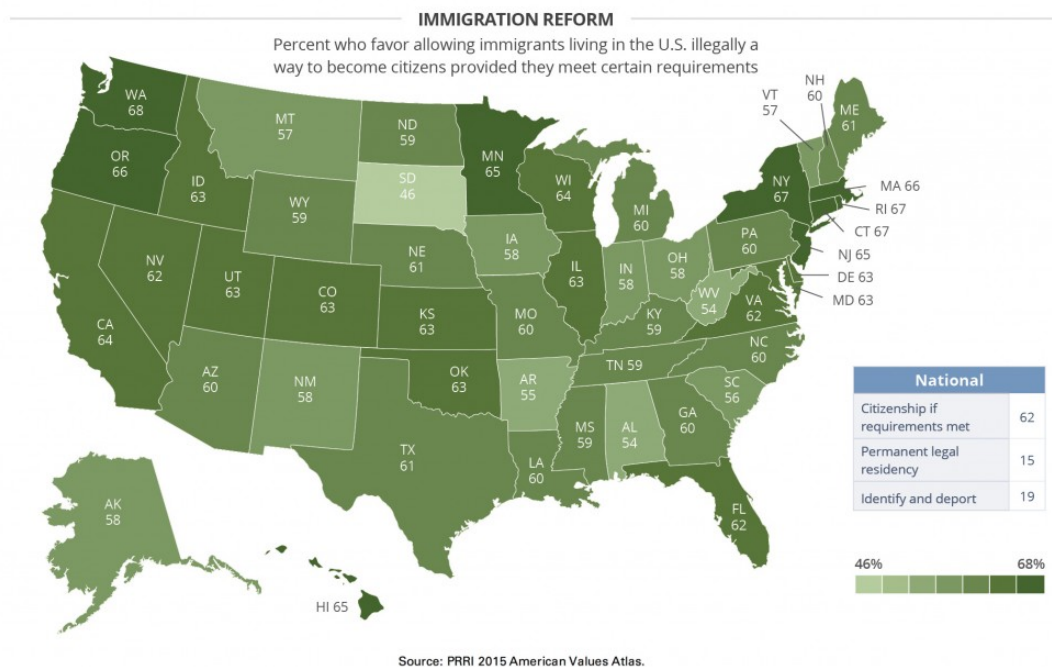


Figure 7: Immigration sentiment II

Table 1: Optimal Values

Model	General	Skilled	Unskilled
Objective Function	244.361	24.085	26.386
Movers	-1.885	-0.548	2.679
Country	0.304	-0.884	-0.429
State	0.509	4.914	2.183
Type	NA	-4.241	-0.992

Table 2: General OLS

	<i>Dependent variable:</i>
	GDPRate
PercentMovers	0.304 (0.218)
DifferentState	-1.122 (1.678)
DifferentCountry	0.509 (1.120)
Constant	-1.885 (1.791)
Observations	63
R <sup>2</sup>	0.148
Adjusted R <sup>2</sup>	0.105
Residual Std. Error	2.035 (df = 59)
F Statistic	3.423** (df = 3; 59)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3: Skilled Model

	<i>Dependent variable:</i>
	GDPRate
PercentMovers	−0.451 (0.636)
NewJob	0.408 (0.473)
DifferentCountry	1.570 (1.162)
Constant	−2.793 (6.923)
Observations	20
R <sup>2</sup>	0.447
Adjusted R <sup>2</sup>	0.343
Residual Std. Error	1.335 (df = 16)
F Statistic	4.303** (df = 3; 16)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4: Unskilled Model

	<i>Dependent variable:</i>
	GDPRate
PercentMovers	−0.338 (0.511)
Work	−1.503* (0.777)
DifferentCountry	1.432 (0.887)
Constant	3.208 (2.945)
Observations	20
R <sup>2</sup>	0.531
Adjusted R <sup>2</sup>	0.443
Residual Std. Error	1.229 (df = 16)
F Statistic	6.028*** (df = 3; 16)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 5: Deviance and Sentiments

State	Texas	Vermont	Ohio	California
Deviance	114.832	5.512	19.828	46.470
Sent. I	+2	-1	-5	+8
Sent. 2	-1	-5	-4	+2

Table 6: Deviance and Sentiments

State	New York	Florida	Wyoming	Oklahoma
Deviance	20.352	10.501	661.266	145.455
Sent. I	+8	0	-12	-3
Sent. 2	+5	0	-3	+1

## References

- [1] Rafaela M. Dancygier and Michael J. Donnelly. Sectoral economies, economic contexts, and attitudes toward immigration. *The Journal of Politics*, 75(1):17–35, 2012.
- [2] Mete Feridun. *Migrants and Markets*. Amsterdam University Press, 2008.
- [3] Chen Liu Gordon Hanson and Craig McIntosh. The rise and fall of u.s. low-skilled immigration. *Brookings Paper on Economic Activity*, pages 83–151, 2017.
- [4] Poul Schou. Immigration, integration and fiscal sustainability. *Journal of Population Economics*, 19(4):671–689, 2006.
- [5] Nicole B. Simpson and Chad Sparber. The short- and long-run determinants of less-educated immigrant flows into the u.s. states. *Southern Economic Journal*, 80(2):414–438, 2013.