

Analyzing The Economic Impact of Illegal Immigration

Daniel Liang
New York, USA
dzl217@nyu.edu

Peter Peirce
New York, USA
pp1844@nyu.edu

Abstract—

By comparing estimates of illegal immigration at a county level to select economic indicators over a five year period, the beginnings of a relationship between illegal immigrant population levels and economic growth in the United States can be determined. The Pearson correlation coefficient was calculated for each sector to determine if there was any correlation between estimates of illegal immigration populations and individual workforce sectors. There are a few weak positive correlations between the number of illegal immigrants and the economic growth in specific sectors, but much of the data suggests that there is no correlation. While this exploration is preliminary, it should provide a general gauge on the direction of the relationship for future research.

Keywords—illegal immigration, economic growth, Pearson correlation coefficient

I. INTRODUCTION

The goal of this analytic is to use estimates of illegal immigrant populations, compared against county-level economic indicators between 2010 and 2014 to determine a possible correlation between the two. Yearly data for each county starting in 2010 and going through 2014 will be used to build a picture of the economic growth of that county over that time period. Then, illegal immigration data will be used to determine if higher populations of illegal immigrants are correlated with higher, lower, or are not correlated at all to economic growth. We expect policymakers who want to make informed decisions on immigration policy will benefit from this analytic. While there are multiple view points of the impact of unauthorized immigrant populations on the economy, the correlation produced will reinforce a particular view by using methods and techniques from similar studies.

II. MOTIVATION

Unauthorized immigration has always been a topic of concern, but much more so recently with new policies and increased media coverage within the United States on the issue. Some recent and upcoming policies from the government that would affect illegal immigrants include the travel ban on several Muslim countries, a plan to build a border wall between Mexico and the United States, and the recent child separation policy. These are just some of the largest and most covered policies amongst the many obscure ones that are being

introduced to greatly affect the illegal immigrant population residing or coming into the United States. While there is much action being done about changing how illegal immigrants can come in or stay in the United States, there's surprisingly little research on the overall impact of their population on the United States.

There are other countries that have extensive research on the effect of foreign labor on their economies. Canada is an example of one such country; researchers were able to look at characteristics of their immigrants and create meaningful work to provide feedback for policymaking. One study looks over relevant aspects of the Canadian immigration system and individuals' immigration classes, such as humanitarian immigrants or refugees, to compare the earnings and employment outcomes each specific immigrant classes and see if other classes have specific advantages over the others. [8] Similarly, another study looks at the wages of recent and earlier arriving immigrants to see how important this is to the declining assimilation of immigrants in Canada. [6] Both of these studies have lead to more literature being created and increased awareness and literacy of the topic. This also lead to Canada being one of the most receptive countries for immigrants because of the government's and public's opinions on immigrants [9], which significantly impacted Canadian immigration policy.

This analytic was inspired by recent events as well as these relevant resources. It is important to fully consider the effects that illegal immigrants have on all aspects of a country prior to making important decisions and policies. Users of the analytic may include economists, demographers, and policy-makers, who can all create new literature or more educatedly make decisions. This would benefit both the users themselves, the immigrant population, and the entire country as a more informed decision should lead to better policies being made.

III. RELATED WORK

Many resources were referred to when compiling ideas about this topic. One paper which inspired much of the idea behind the topic was a paper by multiple authors which provides an overview of the state of the use of big data in academic research on immigration from a Canadian perspective in 2016. Data from census surveys, social media and internet sources, financial service transaction data, and more all provide researchers with huge amounts of data that would previously have been unmanageable. The paper reported nearly 60% of studies used the Statistics Canada census data. It

also found that the majority of studies were interested in labor market and social integration research and that the use of big data in published studies concerning immigration has grown over time. [1]

There are many potential uses of big data in future immigration research. Realtime data with geolocation can help humanitarian organizations efficiently allocate resources to help a greater number of people. One problem facing non-government researchers is the lack of access they have to administrative-level data in more than thirty current surveys. While researchers with access to this data have been able to use it to perform innovative studies like the GIS Mapping Project which tracks recent immigrants in western and northern Canada. [1]

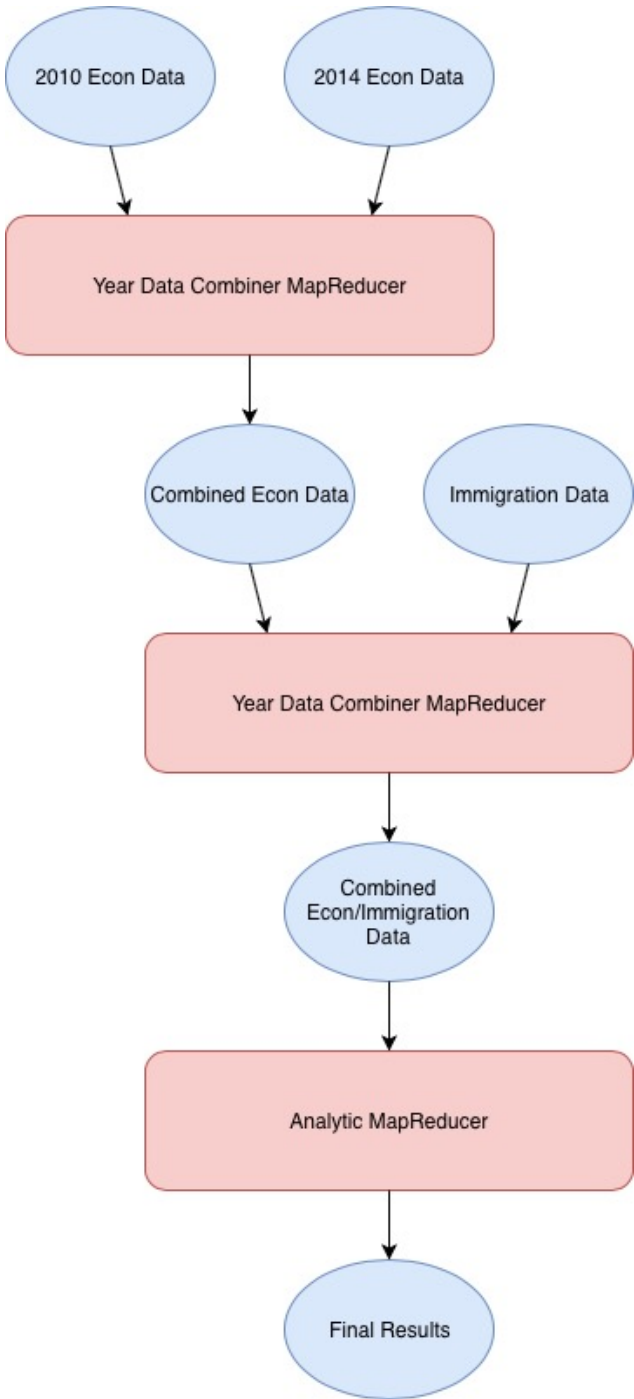
Despite the widespread and rapidly growing popularity of Big Data, researchers have yet to agree on what the concept entails, what tools are still needed to best interrogate these data, whether or not Big Data’s emergence represents a new academic field or simply a set of tools, and how much confidence we can place on results derived from Big Data. Despite these ambiguities, most would agree that Big Data and the methods for analyzing it represent a remarkable potential for advancing social science knowledge. Demographers however have generally collected and analyzed Big Data in a small way, by parsing out the points of information that we can manipulate with familiar models and restricting analyses to what typical computing systems can handle or restricted-access data disseminators will allow. In order to better interrogate existing data, the culture of demography needs to change to treat demographic micro-data as Big Data. This includes shaping the definition of Big Data, changing how we conceptualize models, and re-evaluating how we silo confidential data. [5]

Big Data has great potential, but in terms of social sciences and demography it is used in small ways. The data usually analyzed is exhaust data, which is data not created for the purpose of research and analysis. The parsing of data into small easy to use chunks can fail to account for the entirety of the data. In order for this to change, the author asserts that Big Data needs to be redefined in terms that advantages Demography, models that aren’t deductive need to be embraced, and confidential data practices need to be rethought. Standards of statistical clustering need to be raised and allowing the chance to view random patterns in large datasets is key to looking at the data with a larger perspective. These are critical changes that will push the boundaries of Big Data in Demography. [5]

The papers referred to above as well as other resources have inspired and influenced the direction of how we produced results on this topic either through this current analytic or future works.

IV. DESIGN

The design diagram to the right shows a high-level overview of the movement of data through the cleaning and analytic stages of the project. Three original data sources are used, both the 2010 and 2014 county-level economic indicators, as well as the county-level illegal immigrant population numbers obtained through web scrapping. The



economic data is combined in the first MapReduce program. Then, the output of that job is combined with the immigration data through the second MapReduce program, which provides the final version of the data which will be used in the analytic. After the analytic MapReduce job runs, the final results are obtained.

V. RESULTS

Our experiment takes robust county-level economic data provided by the US Census Bureau, combines it with the best estimates of illegal immigrant county populations numbers that we were able to find, and uses that data to provide novel

	# Establishments	# Employees	Q1 Payroll	Annual Payroll
Total for all sectors	0.23	0.12	0.09	0.10
Agriculture, forestry, hunting	0.03	0.12	0.32	0.22
Mining, oil, gas	0.23	0.24	0.20	0.20
Utilities	0.18	0.20	0.15	0.13
Construction	0.23	0.11	0.11	0.11
Manufacturing	-0.05	-0.16	-0.03	-0.06
Wholesale trade	0.28	0.04	0.11	0.12
Retail trade	0.14	0.15	0.09	0.11
Transportation and warehousing	0.13	0.12	0.15	0.16
Information	0.18	0.18	0.10	0.08
Finance and insurance	0.31	0.05	0.01	0.02
Real estate	0.15	0.13	0.08	0.09
Professional, scientific, and technical services	0.22	0.05	0.08	0.08
Enterprise Management	0.07	-0.03	0.09	0.08
Waste management services	0.19	0.08	0.05	0.07
Educational services	0.20	0.11	0.07	0.09
Health care and social assistance	0.04	0.11	0.07	0.08
Arts, entertainment, and recreation	0.16	-0.05	0.04	0.03
Accommodation/ Food services	0.21	0.10	0.06	0.08
Other	0.18	0.18	0.08	0.08
Not listed	-0.13	0.32	0.24	0.23
average	0.15	0.10	0.10	0.10
min	-0.13	-0.16	-0.03	-0.06
max	0.31	0.32	0.32	0.23

insight into the potential relationship between the two. It's important to make clear that our results do not claim to prove causation, these numbers are purely correlative. Future work to prove causation would need to be more experimental in nature, testing the result of changing only one variable on another while holding as many other variables constant as possible.

The primary output of our analytic was the Pearson coefficient using the change in economic indicators between the years 2010 and 2014, and the percentage of the population made up of illegal immigrants in that same time frame. The Pearson correlation coefficient measures the linear correlation between two variables. It's value can range between -1 and 1, with -1 showing a perfect negative relationship, 1 showing a perfect positive relationship, and 0 showing no relationship at all. A negative relationship is one in which increasing the value of one variable results in the other variable decreasing in value. A positive relationship is one in which increasing the value of one variable results in the other variable increasing in value. [7]

Our results include the Pearson coefficient between the change over time of four different economic indicators (number of establishments, number of employees, first quarter payroll, yearly payroll) and the illegal immigrant population for twenty-three different industries as well as the total for all industries combined. After the final results were calculated in MapReduce, the average, min, and max across all industries for each economic indicator was also computed.

Our original hypothesis, that we would find a strong positive relationship between illegal immigrant population numbers and economic indicators, did not hold. Instead, we found that there is somewhere between no relationship and a weak relationship between these variables. The number of establishments variable had a range of Pearson coefficients between -0.13 and 0.31, with a mean value of 0.15. The number of employees variable had a range of Pearson coefficients between -0.16 and 0.32, with a mean value of 0.10. The first quarter payroll variable had a range of Pearson coefficients between -0.03 and 0.32, with a mean value of 0.10. The annual payroll variable had a range of Pearson coefficients between -0.06 and 0.23, with a mean value of 0.10.

These all suggest a very slight positive relationship between the variables.

VI. FUTURE WORK

The analytic currently only focuses on the estimates of the illegal immigrant populations within each county. This is by far the largest indication of any correlation, as the correlation is solely dependent on the growth of the sector via the number of working illegal immigrants. Looking at the entire population aggregates all the potential subtleties for correlations from all other factors of the county's illegal immigration population, so any possibility of an overall correlation would be found here. Given that there was very little to no correlation in any of the sectors overall, there seems to be little incentive in finding correlations on more subtle factors like the population's overall educational background or English proficiency. While this current analytic did not delve into these areas, future analytics that expand into finding these correlations could potentially produce interesting results.

Among difficulties experienced in producing the analytic, the one that affected the analytic the most was a dearth of data on illegal immigrant population numbers. This made it difficult to gather insights on how to initially determine a correlation and overall makes the impact of the analytic weaker. Unfortunately at this time robust data on illegal immigrants is nonexistent. Talk of adding a legal-status question to the 2020 Census brings hope of future studies with access to significantly more robust illegal immigrant population numbers. Using larger and more reliable data sources would be possible as more data on this topic is produced and accumulated in the future. This would increase the complexity and insightfulness of this analytic, and could produce different outcomes than what had been presented.

VII. CONCLUSION

The results of this analytic suggest that there is a negligible relationship between the unauthorized immigrant population numbers in the United States and the growth of specific

workforce sectors. The analytic is based on linear regressions between sets of variables to provide a number representing a correlation. This is a reasonable and effective way of measuring this particular correlation for a few reasons. Many of the related works for this topic have used linear regressions to model populations with the economy because changes with population numbers can generally suggest a constant change across a dependent variable [7], which in this case is the economy. In other words, it is reasonable to assume an increase in workforce population can in general lead to some linear change in the growth of an economy, as increased labor would lead to more production and a stronger economy. A population number would work well as an independent variable, as it is continuous and contains almost no anomalies since the population number would very rarely drastically change. [7] While an economy displays unpredictable behavior dependent on many different variables, using population numbers as the only independent variable can reasonably create a monotonous relationship with the economy.

The results found here were not definitive, but they lend themselves to interesting future work and interpretation. These results shows that illegal immigration population numbers alone do not show any correlations with the growth in the economy. The fact that there was in reality growth post-recession means that amongst other factors, it's possible that there lie other interesting characteristics from the unauthorized immigrant population that could affect the economy's growth. This also supports the statements made about treating demographics in a more abstract way with Big Data, so that other random patterns that could not be seen here, may appear and lead to unique results. With the increase in quality and quantity of data sources and an increasingly broader use for Big Data over the data, there can emerge a more definitive perspective over the effects of illegal immigration over the United States' overall economic growth.

ACKNOWLEDGMENT

This report would not be possible without the Dumbo Cluster provided by the High Performance Computing department at New York University. The cluster provided the necessary resources to run Hadoop MapReduce jobs that helped clean and profile the data, and run the analytic that produced the results of the report. Additionally, we thank Professor McIntosh for her support and feedback in providing insight and direction for creating this report.

REFERENCES

1. W. Ashton, P. Bhattacharyya, E. Galatsanou, S. Ogoe, L. Wilkinson. Emerging Uses of Big Data in Immigration Research. SSHRC. Oct, 2016.
2. H. Wang, D. Huo, Y. Xu. A big data analytics framework for border crossing transportation. IJBDI. Oct, 2017.
3. Y. Ahuja. Forecasting Trends of Migration to the United States of America: A Human Resource Perspective. San Jose State University. Jan, 2017.
4. A. Rekha. Big Data and National Cyber Security Intelligence. MBDIPS. 2016.
5. S. Bohon. Demography in the Big Data Revolution: Changing the Culture to Forge New Frontiers. PSPR. Mar, 2018.
6. Tondji, Jean-Baptiste, 2015. "Wage of Immigrants in the Canadian Labour Market," MPRA Paper 80783, University Library of Munich, Germany.

7. Schneider A, Hommel G, Blettner M. Linear regression analysis—part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010. 107;44:776–782.
8. Arthur Sweetman & Casey Warman, 2012. "The Structure of Canada's Immigration System and Canadian Labour Market Outcomes," Working Papers 1292, Queen's University, Department of Economics.
9. Abdur Rahim (2014). *Canadian Immigration and South Asian Immigrants*. Xlibris Corporation. p. 191.