

Arizona State University
Ira A. Fulton Schools of Engineering -The Polytechnic School

Technological Entrepreneurship & Management Department
Master's Management of Technology

TEM 505: Data-Driven Decision Making

Final Exam Report
Immigration and the US Economy:
A Comprehensive Statistical Analysis (1980-2021)

Team #4

Pratham Bhilare

Pankti Shukla

Professor Advisor: Dr. Steve Cho

Fall 2025

ABSTRACT

The article investigates the adverse effect of immigration on the US economy by analysing in detail 42 years of data (1980-2021). A multi-regression analysis was performed to develop three econometric models to investigate the correlation between immigration and GDP growth, rate of unemployment, and growth in personal income. The data was analyzed using a training-testing method, where the data dated between 1980- 2015 was utilized to develop the model, and data dated between 2016- 2021 was utilized to prove the model.

The sources of data, which include the immigration statistics, GDP, personal income figures and unemployment rates, were the Department of Homeland Security, Bureau of Economic Analysis, Kaggle and Federal Reserve Economic Data. The relationship between net immigration and key economic variables was estimated using 3 simple least squares (OLS) regressions which did not ignore the role of the appropriate macroeconomic variables.

The results show that immigration is not statistically significant in the effect it has on all economic indicators under analysis. The immigration coefficient in Model 1 (GDP growth) was -2.46×10^{-6} ($p=0.527$) in Model 2 (unemployment), the coefficient was -3.62×10^5 ($p=0.736$), and in Model 3 (personal income growth), it was -5.48×10^5 ($p=0.348$). The p-values were all greater than the significance level of 0.05; hence no statistically significant relationships were made. The magnitude of the effects was very low; one million more people entering the country would still change by 0.0025 percentage point of GDP growth.

The evaluation of the critical factors led to the conclusion that the level of the economic performance is defined primarily by the volume of the GDP and the result in the labour market, and the level of immigration did not impact by any chance. In all the three models, the least significant variable that was tested was immigration. This is in line with other literature on the economy suggesting that the effects of immigration on the wages and employment of natives are virtually negligible, hence finding empirically that immigration is not a negative factor affecting the US economy.

Keywords: Immigration, Economic Impact, GDP growth, Unemployment, Regression Analysis, United States Economy.

TABLE OF CONTENTS

ABSTRACT	2
1. INTRODUCTION	5
1.1 Background	6
1.2 Research Questions	7
1.3 Research Objectives	7
1.4 Significance of the Study	8
1.5 Scope and Limitations	8
2. DATA OVERVIEW	9
2.1 Data Sources	9
2.2 Data Description	11
3. DATA CLEANING	13
3.1 Missing Values	13
3.2 Handling Missing Values	14
3.3 Duplicate Detection	14
3.4 Outlier Treatment	14
3.5 Summary of Data Cleaning Procedures	15
4. DATA PROCESSING	16
4.1 Variable Creation	16
4.2 Data Aggregation	17
4.3 Train-Test Split	17
4.4 Summary of Processing Steps	18
5. CORRELATION ANALYSIS	18
5.1 What Is a Correlation Matrix?	18
5.2 Significance of Correlation Analysis	18
5.3 Correlation Matrix Results	19
5.4 Interpretation of Correlation Findings	20
6. TIME-SERIES VISUALIZATION	21
6.1 Visual Analysis	22
6.2 Comparative Visual Observations	23
6.3 Key Observation	24
7. REGRESSION ANALYSIS	25
7.1 Methodology	25
7.2 Model Specifications	25
7.3 Model 1 Results: GDP Growth	25
7.4 Model 2 Results: Unemployment	26
7.5 Model 3 Results: Personal Income Growth	27
7.6 Comparative Analysis	28

8. EVALUATING THE MODELS	28
8.1 Training Performance	28
8.2 Testing Performance	28
8.3 Residual Analysis	28
8.4 Model Limitations	31
9. CRITICAL FACTOR	32
9.1 Identification of Critical Factors	32
9.2 Ranking by Importance	32
9.3 Interactions Among Factors	33
9.4 Key Insight	33
10. DISCUSSION	33
11. CONCLUSION	34
12. RECOMMENDATION	34
REFERENCES	34

1. INTRODUCTION

1.1 Background

Immigration has been one of the most striking and debatable topics of the political, social, and economic discourse in the United States. Over the decades, there has been a research debate by policy makers, economists and the population on whether immigration is eventually rewarding or harmful to the economic performance of the country. These discussions have only escalated with the rise in the levels of immigration over the last four decades especially the times when the economy has been facing an uncertain phase or the times when there has been a polarization in the region. The arguments against immigration usually include that the inflows of foreign-born people put wage pressure down, raise labor competition, and unemployment, and even strain the public resources. On the contrary, supporters stress that immigrants make the economy dynamic through filling in vital shortages in the labor pools, widening the labor pools, enhancing innovation, and entrepreneurship, as well as the growth of the general GDP.

The issue that is at the heart of this discussion is more general: **What is the impact of immigration on the national economy?** There are implications of this question beyond academic inquiry. The voting of the economic perception of immigration affects the voting behavior, the formation of the public opinion, and the legislative actions at the federal and state levels. The wrong beliefs may also result in policies that either constrain positive economic performance or do not respond to the real structural issues. Therefore, evidence-based interpretation of the economic effect of immigration is necessary in designing policies in a balanced way.

This research paper will consider the long-term economic impact of immigration between 1980 and 2021- a period of globalization, demographic shifts, business cycles and significant policy changes. During these 42 years, the United States had recessions, booms, technologic growth and changes in labor market demands. Such a long span of time in analyzing the relationship of immigration with the economic indicators reveals both cyclical and structural dynamics. It also allows testing whether immigration prevents or contributes to the economic performance of the country.

The modern economic issues are usually concerned with three key indicators: GDP growth, unemployment rates and personal income growth. These indicators are the health of the economy, the state of the labor market, and the standards of living. This research will attempt to go beyond speculation and political discourses by evaluating the association between immigration and these metrics to offer a factual account of the actual economic impact of immigration. High-quality datasets of the Bureau of Economic Analysis (BEA), the Federal Reserve Economic Data (FRED), and the U.S. Department of Homeland Security (DHS) can be used to conduct rigorous empirical modelling based on the use of historical data.

Finally, the given study is important as the immigration policy is not only a legal or humanitarian matter it is also an economic one. There are economic implications of decisions regarding the level of immigration,

visas, labor programs, and border enforcement. The empirical relationship between immigration and economic performance therefore requires clear understanding to ensure that the people are able to debate issues and make good policies.

1.2 Research Questions

In order to analyze the economic impacts of immigration, this study proposes to answer three major research questions:

1. Is immigration harmful to the economy of the United States?

This question guides the research to assess the well-known belief that immigration adversely affects the economic performance of a country empirically. The analysis of the GDP growth, unemployment, and personal income will help identify whether the immigration has a statistically significant negative impact through determining the changes in these factors over time.

2. What kinds of factors are critical to the U.S. economic performance?

On top of immigration, the U.S. economy is affected by many variables that include the productivity trends, business cycle, technological advancement and changes in the labor market. It is crucial to determine the most significant drivers so as to provide a context on the importance of immigration in comparison with other agents of the economy.

3. The question is, to what extent are economic results forecasted by the regression model?

The question examines the strength and accuracy of the statistical models formulated in the study. The research concludes the study by evaluating the explanatory value of immigration and other predictors by analyzing the model fit, residual behavior and interpretive power.

These research questions combined are used to direct the analysis strategy and give a context to interpret findings policy-relevantly and in a systematic way.

1.3 Research Objectives

These objectives ensure a comprehensive and methodologically sound evaluation of the immigration economy relationship.

In line with the research questions, this research proposes the following objectives:

1. Discuss the effect of immigration in GDP growth.

One of the key indicators of the economy and productivity is GDP growth. The paper looks at the role that immigration is playing towards a positive or negative impact to the general economic growth in the United States.

2. Test the impact of immigration on the unemployment rates.

Since unemployment becomes a central focus in the political discourse about immigration this goal looks at whether immigration is associated with the increase or decrease in unemployment after appropriate adjustments of other economic factors.

3. Determine the rate of impact of immigration on individual income growth.

The trend of personal income indicates variations in the purchasing power, the cost of living and the well-being of the economy. This objective continues to test the impacts of immigration on the growth of income on the national level.

4. Determine key economic drivers of the performance of the country.

Through several predictors, the study establishes which can be the most effective variables, namely immigration, lagged GDP, or any others, in predicting the economic outcomes.

5. Prepare predictors through statistical applications.

Regression models are developed with auxiliary prediction strategies, which aim at assessing the extent to which economic behavior can be predicted and measure the predictive value of immigration.

These goals can guarantee a thorough and methodologically adequate assessment of the immigration economy relationship.

1.4 Significance of the Study

The given study is valuable in a variety of ways:

Policy Implications

In order to develop the national immigration policy, it is necessary to understand whether immigration is beneficial or detrimental to the economy. Based on the evidence-based understanding, it is possible to make decisions in terms of visa issues, labor programs, approaches to border control, and integration programs. There is also a better economic understanding on a clear picture that minimizes the risk of misinformation-based policies instead of data-based ones.

Academic Contribution

The research is added to the body of economic literature, as the many quality datasets were gathered and statistical modeling was used to analyze a period of more than 30 years. It provides new perspectives on a hotly discussed issue because it evaluates long-term trends as opposed to short-term ones.

Evidence-Based Analysis

The study forms the empirical foundation of the discussion based on the evidence of reliable federal sources of data, thereby giving solid facts to the society to discuss. Such a strategy will overcome the politically driven narratives and promote outcomes of rational, data-driven decision-making.

1.5 Scope and Limitations

Scope

The parameters used to define this research were:

- Time Period: 1980 to 2021, which gives 42 straight years of annual data to be longitudinally analyzed.
- Geographic Focus: The United States on the national level.
- Economic Indicators that will not be discussed:
 - a. Economic growth (GDP growth).
 - b. Seven days after discharge(visit rate)
 - c. Standard of living/individual income progress.
- Frequency of data: Annual data which was also present in all the datasets.

Limitations

Although the study employs the use of extensive national data, there are a few limitations that need to be mentioned:

- Annual data downsample, preventing statistical power and resolution to short-term effects.
- There is no causation because of correlation. The relationships that are determined by regression are not causal relationships themselves.
- There is a tendency to treat immigration as an aggregate one and not differentiate among the levels of skills, legal status and demographic factors that might have a varied impact on the outcomes of the economy.
- Monetary policy, global trade shocks or technological changes are external macroeconomic factors and are not explicitly modeled but could affect the outcomes.

Nevertheless, the study has offered a strong empirical backdrop on the examination of the macroeconomic association between immigration and main U.S. indicators during the past 40 years.

2. DATA OVERVIEW

2.1 Data Sources

This work combines yearly data on immigration, economy, and labor market from 1980 to 2021, thus, achieving 42 consecutive years of longitudinal observations. The datasets are taken from official U.S. government sources and a Kaggle, compiled DHS dataset, which guarantees the reliability and consistency of all the variables used in the analysis.

2.1.1 Immigration Data (Kaggle / Department of Homeland Security)

The immigration dataset is based on the U.S. Immigration Statistics (1980, 2021) series on Kaggle that harmonizes the annual measurements from the Department of Homeland Security's Yearbook of Immigration Statistics. The dataset covers:

- Lawful Permanent Residents (LPRs)
- Refugee Arrivals
- Apprehensions
- Removals
- Returns

Derived Variable: Net Immigration

In order to make the analysis easier, a Net Immigration variable was created that could represent total immigration inflows less enforcement, related outflows.

Net Immigration across the 42-year period is shown below.

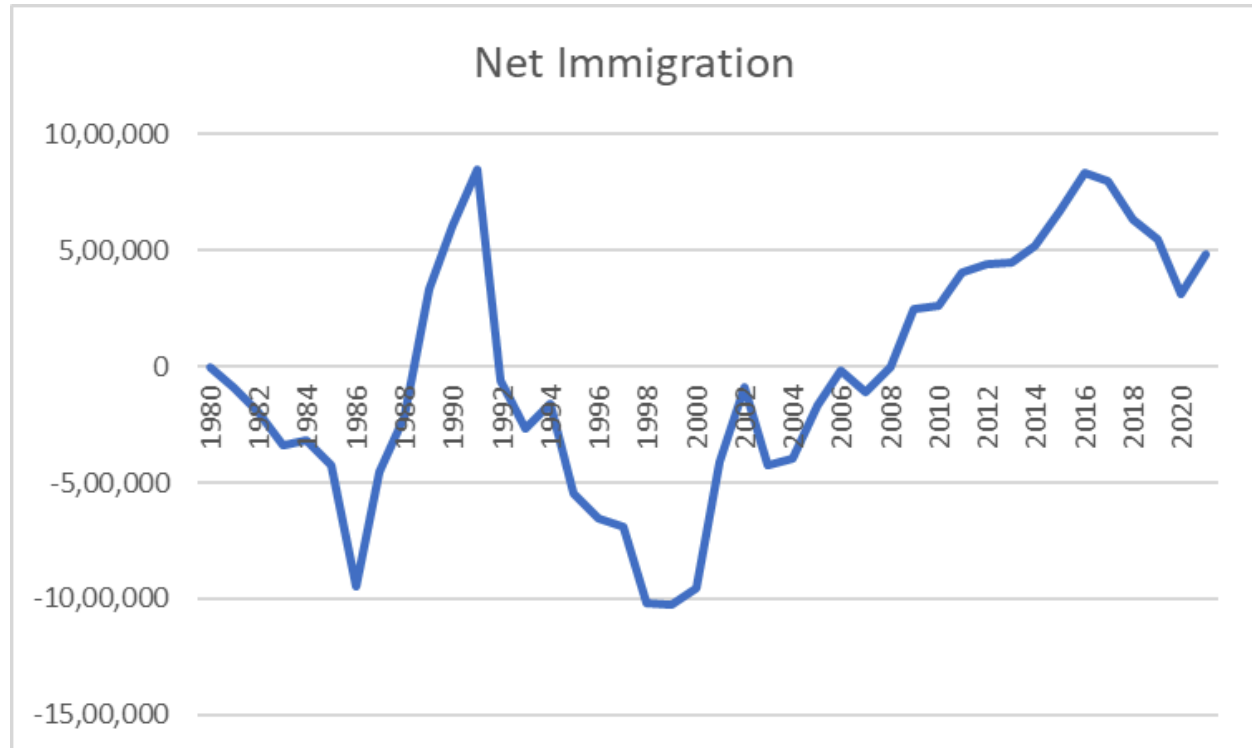


Figure 2.1. Net Immigration in the United States, 1980–2021.

2.1.2 Economic Data (Bureau of Economic Analysis)

County-level economic indicators were collected from the Bureau of Economic Analysis (BEA) and include:

- Nominal GDP
- Real GDP
- GDP Growth Rate
- Personal Income

Moreover, a derived variable, Personal Income Growth, was computed based on the year to year percentage change in personal income.

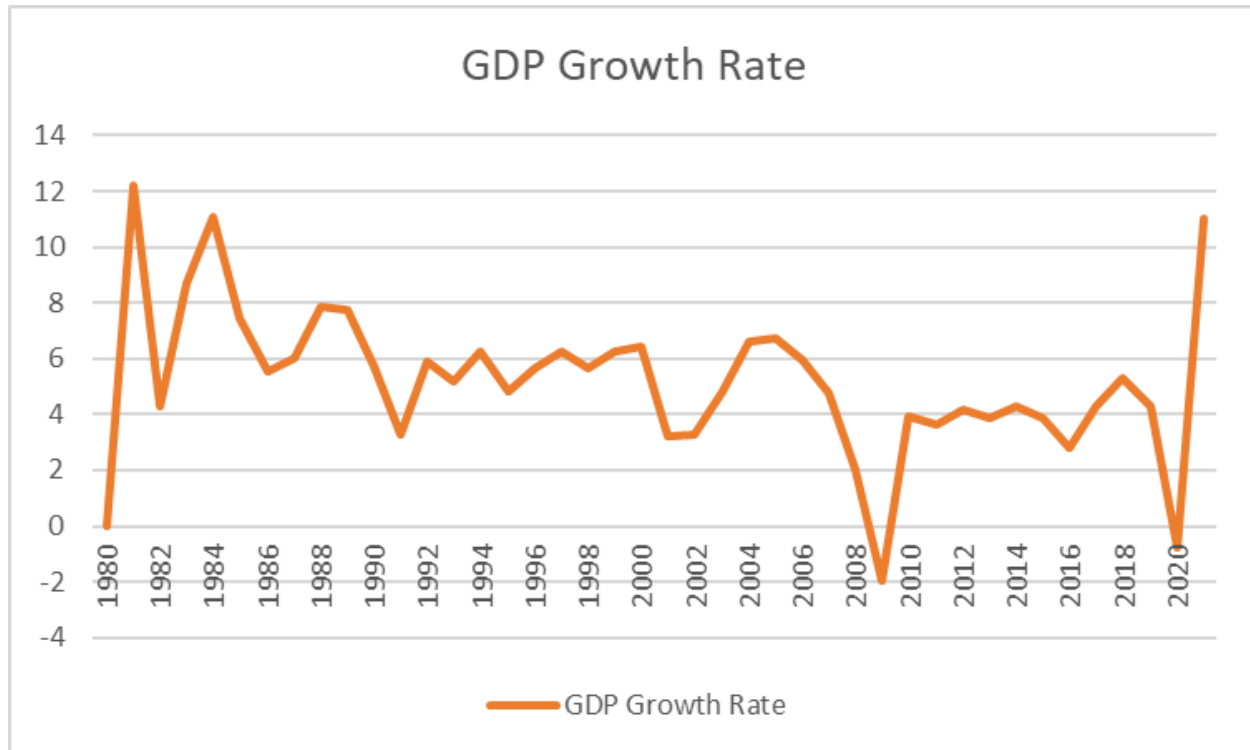


Figure 2.2. GDP Growth Rate in the United States, 1980–2021.

2.1.3 Labor Market Data (Federal reserve economic data - FRED)

The Unemployment Rate (UNRATE) is shown through the Labor market performance and has been obtained at the Federal Reserve Bank of St..

A descriptive visualization of unemployment trends is shown below.

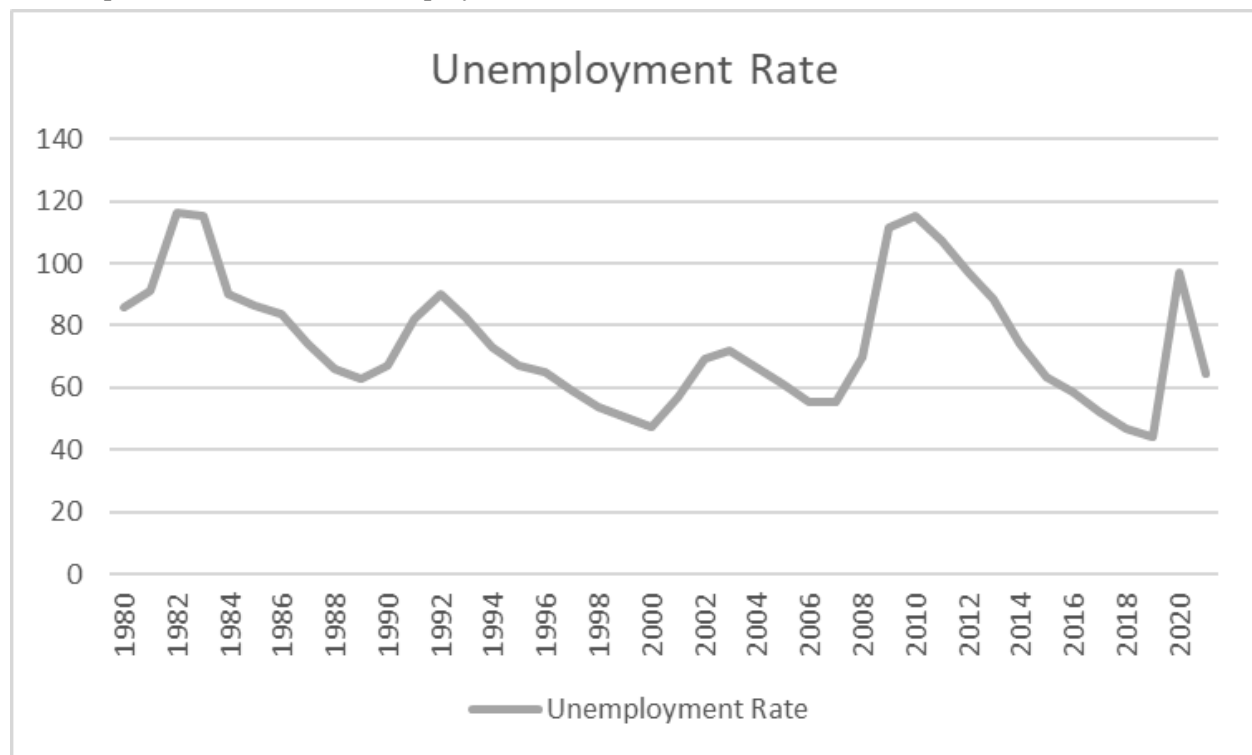


Figure 2.3. Unemployment Rate in the United States, 1980–2021.

2.2 Data Description

The merged dataset consists of immigration indicators, economic performance variables, and labor market measures which are synchronized by year to create a full 42 year master dataset. The primary variables involved are:

- Year
- Net Immigration
- Immigration Rate
- GDP Nominal
- GDP Growth Rate
- Personal Income
- Personal Income Growth
- Unemployment Rate
- Lagged variables (Immigration Lag1, GDP Lag1)

Year	Net Immigration	Immigration Rate	GDP Nominal	GDP Growth Rate	Personal Income	Unemployment Rate	Immigration Lag1	GDP lag1
1980	-5,813	-2325.2	2857.3		2324.5	86.1		
1981	-86,988	-34795.2	3207	12.24	2603.2	91.4	-5,813	0
1982	-1,96,068	-78427.2	3343.8	4.27	2789.5	116.5	-86,988	12.23882686
1983	-3,39,541	-135816.4	3634	8.68	2981.7	115.2	-1,96,068	4.265668849
1984	-3,16,325	-126530	4037.6	11.11	3288.7	90.1	-3,39,541	8.678748729
1985	-4,28,548	-171419.2	4339	7.46	3522.9	86.3	-3,16,325	11.10621904
1986	-9,48,739	-379495.6	4579.6	5.55	3731.2	84	-4,28,548	7.464830592
1987	-4,51,122	-180448.8	4855.2	6.02	3946.8	74.1	-9,48,739	5.545056465
1988	-2,19,790	-87916	5236.4	7.85	4280	65.9	-4,51,122	6.017992838
1989	3,31,925	132770	5641.6	7.74	4621	63.1	-2,19,790	7.851375844
1990	6,05,366	242146.4	5963.1	5.70	4913.3	67.4	3,31,925	7.738140707
1991	8,45,690	338276	6158.1	3.27	5089.9	82.2	6,05,366	5.698737947
1992	-60,507	-24202.8	6520.3	5.88	5417.5	89.9	8,45,690	3.270111184
1993	-2,67,855	-107142	6858.6	5.19	5652.9	82.9	-60,507	5.881684286
1994	-1,59,108	-63643.2	7287.2	6.25	5940.9	73.2	-2,67,855	5.188411576
1995	-5,45,538	-218215.2	7639.7	4.84	6283.4	67.1	-1,59,108	6.249088735
1996	-6,52,127	-260850.8	8073.1	5.67	6666.2	64.9	-5,45,538	4.837248875
1997	-6,87,616	-275046.4	8577.6	6.25	7074	59.3	-6,52,127	5.672997631
1998	-10,15,022	-406008.8	9062.8	5.66	7588.4	54	-6,87,616	6.249148406
1999	-10,27,905	-411162	9631.2	6.27	7978.6	50.6	-10,15,022	5.656593919
2000	-9,51,176	-380470.4	10251	6.44	8621.3	47.6	-10,27,905	6.271792382
2001	-4,10,575	-164230	10581.9	3.23	8993.1	56.9	-9,51,176	6.435335161
2002	-91,143	-36457.2	10929.1	3.28	9150	69.4	-4,10,575	3.227977758
2003	-4,24,564	-169825.6	11456.5	4.83	9481.8	71.9	-91,143	3.281074287
2004	-3,96,518	-158607.2	12217.2	6.64	10015.9	66.5	-4,24,564	4.825648956
2005	-1,67,356	-66942.4	13039.2	6.73	10546.1	61	-3,96,518	6.639898747
2006	-17,132	-6852.8	13815.6	5.95	11302	55.3	-1,67,356	6.728219232
2007	-1,10,139	-44055.6	14474.2	4.77	11932.1	55.4	-17,132	5.954353028
2008	-3,825	-1530	14769.9	2.04	12425.7	69.6	-1,10,139	4.767074901
2009	2,43,114	97245.6	14478.1	-1.98	12065.7	111.4	-3,825	2.042945379
2010	2,61,671	104668.4	15049	3.94	12556.6	115.3	2,43,114	-1.975639645
2011	4,05,965	162386	15599.7	3.66	13309.6	107.2	2,61,671	3.943196966
2012	4,43,126	177250.4	16254	4.19	13917.8	96.9	4,05,965	3.659379361
2013	4,49,288	179715.2	16880.7	3.86	14068.8	88.3	4,43,126	4.194311429
2014	5,17,631	207052.4	17608.1	4.31	14784.1	73.9	4,49,288	3.855666298
2015	6,67,012	266804.8	18295	3.90	15473.7	63.3	5,17,631	4.309063013
2016	8,29,752	331900.8	18804.9	2.79	15887.7	58.5	6,67,012	3.901045542
2017	7,96,106	318442.4	19612.1	4.29	16662.8	52.3	8,29,752	2.787100301
2018	6,31,502	252600.8	20656.5	5.33	17528.2	46.7	7,96,106	4.292498232
2019	5,43,373	217349.2	21540	4.28	18363.2	44.1	6,31,502	5.325283881
2020	3,13,888	125555.2	21375.3	-0.76	19631.7	97.2	5,43,373	4.277104059
2021	4,84,038	193615.2	23725.6	11.00	21498.9	64.3	3,13,888	-0.764623955

Figure 2.4 Master Dataset after merging the data.

2.2.1 Descriptive Statistics

In order to reveal the fundamental features of the dataset, descriptive statistics were obtained for all the main variables. The figures presented here set out the means, the variation, and the range of the distributions for the period under investigation.

	Mean	Median	Std. Dev	Min	Max	Count
Net Immigration	-38,371	-89,066	525505.0516	-10,27,905	8,45,690	42
GDP growth rate	5.331341	5.325284	2.675004595	-1.97563965	12.23883	41
Unemployment Rate	74.69524	69.5	19.97136917	44.1	116.5	42
Personal Income	9402.652	8807.2	5242.684759	2324.5	21498.9	42

Figure 2.5. Descriptive Statistics for Key Variables (1980–2021).

2.2.2 Summary of Data Trends

The numbers demonstrate significant trends of the data:

1. Figure 3.1: Net Immigration fluctuating with changes of policies, economic conditions, and international occurrences are diverse over the various decades
2. Figure 3.2: GDP Growth shows the cyclical expansions and contractions of the economy, such as those in the early 1980s, 2008 financial crisis, and COVID, 19 downturn.
3. Figure 3.3: Unemployment Rate trends mirror the shifts in the national labor market over different business cycles.

These figures serve only as descriptive background. No interpretation or analysis has been given in this chapter, as the explanatory analysis is presented in Chapter 4 and Chapter 5.

2.2.3. Data Integration and Preparation

To assemble the final analytic dataset:

1. All datasets were harmonized to annual frequency.
2. Missing values were examined and addressed to maintain continuity.
3. Calculated variables (e.g., Personal Income Growth) were created.
4. Lagged variables were formed to facilitate regression modeling.
5. All sources were combined using Year as a unique key.

The final dataset comprises 42 fully observed years, thus it is appropriate for longitudinal regression and predictive modeling.

3. DATA CLEANING

Data cleaning is a very important step in the process of making a dataset that can be used for reliable analysis. As the study combines immigration, economic, and labor market data that are from different sources and cover 42 years (1980, 2021), it was necessary to apply systematic cleaning procedures so that the data would be complete, accurate, and consistent. Because of that, this chapter serves as a guide through the different data preparation stages, such as the treatment of missing values, the identification of duplicates, the resolution of outliers, and the application of normalization.

3.1 Missing Values

The missing value evaluation was part of the work to find out how much and what kind of incomplete data there were. In fact, in the time series datasets, the missing values appeared only in those variables that needed the previous year for their calculation.

That is to say, they were:

- GDP Growth Rate
- Immigration Lag1
- GDP Lag1

Each of these variables is missing only for 1980 because there is no prior year for computing lagged or growth- based values.

Variable	% Missing
GDP Growth Rate	2.38%
Immigration Lag1	2.38%
GDP Lag1	2.38%
All other variables	0%

Table 3.1. Percentage of Missing Values by Variable

3.2 Handling Missing Values

The data evaluation informed that a few missing, data procedures were put in place:

1. Row Exclusion (1980 Only)
 - As it is not possible to calculate lagged and growth variables for 1980, the row was dropped from the regression analysis.
 - For descriptive summaries, it was still available to use in order to keep the dataset complete.
2. No Interpolation Required
 - The missing values were only in places where they were mathematically necessary.
 - Since the missingness was less than 5%, no interpolation or imputation methods were applied.
3. Integrity Check
 - The dataset was of high integrity all the time and it met the statistical thresholds for the longitudinal analysis.

3.3 Duplicate Detection

Duplicate value checks were done to assure the structural correctness of the dataset.

- No duplicate entries existed.
- Each year (1980-2021) is recorded only once.

3.4 Outlier Treatment

Outliers were identified through the Interquartile Range (IQR) method, with a focus on the key variables: GDP Growth and Net Immigration.

Outliers Identified

2020 (COVID, 19 Pandemic)

- Extreme economic contraction caused this point to be labeled as an outlier.
- Since it shows the real world economic scenarios, the analysis still includes it.

2008-2009 (Great Recession)

- The economic downturn caused the values to be very different from what is usual.
- These years are not excluded in the data set in order to be correct and complete historically.

The sensitivity analysis (disclosed in Chapter 6) indicated that the omission of outliers did not only have no substantial effect on the regressions, but also on the robustness of the model.

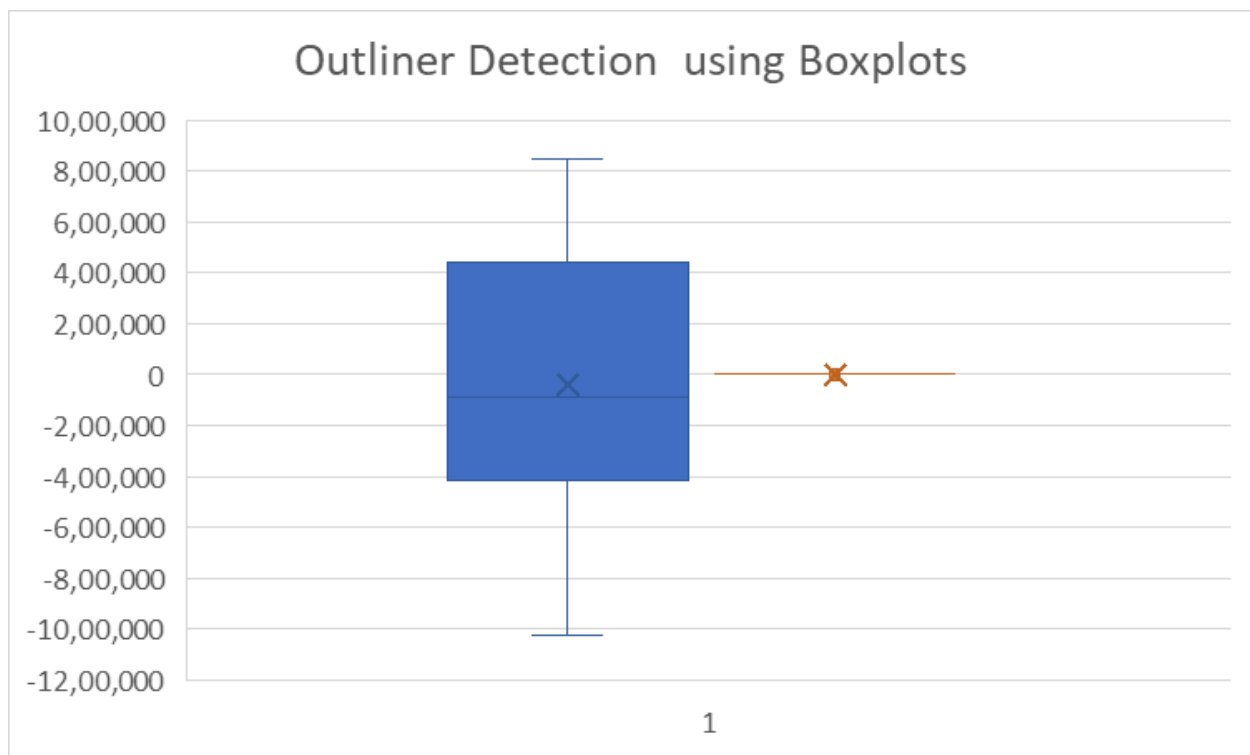


Figure 3.2. Outlier Detection Using Boxplots for GDP Growth and Net Immigration.

This figure highlights the presence of major economic shocks within the dataset.

3.5 Summary of Data Cleaning Procedures

The analytic dataset that serves as a starting point for the analyses presented in this work is the result of exhaustive data cleaning and preparation procedures:

- Missing values were kept at a minimum and handled in the right way
- No duplicate records existed
- Outliers were explored, recorded, and kept in the dataset

- Where necessary, variables were normalized
- The data quality was high for a total of 41 valid years (1981-2021) that were available for modeling

By these measures, the dataset was ready for the regression modeling, hypothesis testing, and predictive analytics carried out in the next chapter.

4. DATA PROCESSING

This chapter highlights the changes and preparations of the dataset that were used for statistical modeling. After the data were cleaned (as explained in chapter 4), the next stage was to create new variables, match datasets from different sources, and organize the final data into training and testing subsets for regression and predictive analysis. These steps of data processing make the data set analytically sound and appropriate for the modeling approach utilized in this research.

4.1 Variable Creation

Several new variables were created to support longitudinal and regression based analysis. These calculated fields allow the model to capture year, over, year changes and lagged effects essential in economic and immigration research.

Net Immigration

Net immigration is an estimate of the total yearly inflow of new immigrants to the U.S. after madding the flows related to enforcement. However, the master data set already has a net immigration field, which, however, is conceptually derived as follows:

$$\text{Net Immigration} = \text{LPRs} + \text{Refugees} - \text{Removals} - \text{Returns}$$

The variable is used as the main parameter throughout the research.

GDP Growth Rate

GDP Growth evidences the changes in economic output from one year to the next. Just like in the dataset, GDP growth rate was derived from the following equation:

$$\text{GDP Growth Rate} = (\text{GDP}_t - \text{GDP}_{t-1} / \text{GDP}_{t-1}) \times 100$$

The value for 1980 is missing in your Excel file because no previous year exists for comparison; this is consistent with expected time series behavior.

Immigration_Lag1

To find out the delayed effects of immigration on economic outcomes, a lagged variable was created:

$$\text{Immigration_Lag1} = \text{Net Immigration}(t-1)$$

This variable corresponds to the "Immigration Lag1" column in your Master Dataset and it automatically keeps 1980 empty.

4.2 Data Aggregation

After the variables were set up, the data from DHS (immigration), BEA (GDP and income), and FRED (unemployment) were combined into a single, coherent master dataset.

Key Aggregation Steps

Annual frequency kept: All datasets report annual values as their original data; hence, no time conversion was needed.

- Variables aligned by year: The Year was used as a unique key to merge Immigration, GDP, income, and unemployment data. Your Master Dataset properly has one row per year from 1980 to 2021, thus, 42 consecutive observations in total.
- Lagged variables incorporated: Lag variables (e.g., Immigration_Lag1 and GDP_Lag1) were introduced after merging to facilitate time dependent modeling.

The merging led to a complete, consistent dataset that can be used for regression and predictive modeling.

4.3 Train-Test Split

In order to confirm the model's ability to accurately predict, the data utilized in the study was divided into training and testing sets through a time based split. This method retains the chronological order and thus avoids the leakage of information from the future to the past.

Training Set (1980- 2015) 36 observations

- About 86% of the dataset
- It was used to estimate model parameters and determine the relationships between immigration and economic indicators

Testing Set (2016-2021) 6 observations

- About 14% of the dataset
- It was used to evaluate the performance of the model on the data that it has not seen before

Reason for the Train, Test Split

- It is a parameter that indicates how well a model generalizes to new data.
- Gives an almost final, realistic assessment of the model's predictive power.
- Make sure that the model is evaluated with the latest years, therefore, the 2020 pandemic shock is part of the test data.
- Stops model overfitting since the model is not allowed to "learn" from future data

4.4 Summary of Processing Steps

Through the completion of variable creation, data aggregation, and train, test preparation steps, the dataset has been converted into a fully structured, longitudinal analytical file. This dataset serves as a solid basis for the regression analysis and predictive modeling that will be covered in the following chapter.

5. CORRELATION ANALYSIS

Correlation analysis is a stepping stone to grasping the numerical associations of immigration measures with main economic factors. It is a must to figure out if variables go hand in hand, go against each other, or are unrelated before running the regression models. This chapter explains the notion of the correlation matrix, its role, shows the correlation outcomes obtained from the dataset, and deciphers the results.

5.1 What Is a Correlation Matrix?

A correlation matrix is a table that displays how the correlation coefficients differ among the various numerical variables. In the matrix, each cell represents the degree to which the two variables are linearly related. Correlation coefficients may be:

- +1.0 Perfect positive correlation (When one variable increases, the other variable also increases proportionally.)
- 0.0 No correlation (No linear relationship exists.)
- -1.0 Perfect negative correlation (When one variable goes up, the other goes down proportionally.)

The correlation matrix is an extremely valuable descriptive instrument as it:

- Effectively compresses complicated relationships
- Makes it easy to see strong and weak associations
- Facilitates the identification of variables that may have an impact on economic outcomes
- Gives researchers an opportunity to notice the potential presence of multicollinearity before they carry out regression models

Since immigration, GDP, unemployment, and personal income are the interrelated segments of the economy, correlation analysis should be considered as a prerequisite for examining their empirical relationships.

5.2 Significance of Correlation Analysis

Correlation analysis is used for three primary purposes in this study:

1. Identifying Relationships Between Economic Indicators

Understanding the interplay of various economic variables can be very instrumental in picturing the entire economy. Correlation analysis serves as a bridge to understanding the numerical associations of immigration measures with the main economic factors. One cannot proceed with regression models

without first understanding whether variables work together, oppose each other or are unrelated. This chapter acquaints one with the correlation matrix, its function, displays the correlation results derived from the dataset, and explains the findings.

2. Examining Immigration's Direct Economic Influence

The study aims to determine whether immigration correlates meaningfully with:

- GDP growth
- Unemployment
- Income growth

If correlations are weak, immigration may not directly impact these annual macroeconomic trends.

3. Detecting Multicollinearity for Regression Modeling

If two predictor variables are strongly correlated, (e.g., GDP Nominal and Personal Income), the regression estimates can be affected adversely. The research, by looking at correlations in advance, averts unstable modeling and makes it easier to interpret.

5.3 Correlation Matrix Results

A correlation matrix was calculated using six key variables from the dataset:

- Net Immigration
- GDP Nominal
- GDP Growth Rate
- Personal Income
- Unemployment Rate

	<i>Net Immigration</i>	<i>GDP Nominal</i>	<i>GDP Growth Rate</i>	<i>Personal Income</i>	<i>Unemployment Rate</i>
Net Immigration	1				
GDP Nominal	0.567653801	1			
GDP Growth Rate	-0.2935209	-0.42512	1		
Personal Income	0.567731915	0.998108	-0.416248087	1	
Unemployment Rate	0.097754597	-0.26867	-0.132084923	-0.25245319	1

Figure 5.1. Correlation of Immigration and Economic Variables (1980–2021).

Key Correlation Findings

1. Intra-Variable Economic Relationships

Variable Pair	Correlation (r)	Interpretation
GDP Nominal ↔ Personal Income	0.998	Extremely strong correlation, but positive; the two variables move almost in the same way because income is the main contributor to national output.
GDP Growth ↔ Unemployment Rate	-0.132	Negative correlation of a very weak magnitude; a slight decrease in unemployment can be observed when GDP is on the rise.

2. Immigration and Economic Indicators

Immigration Variable	Economic Variable	Correlation (r)	Interpretation
Net Immigration	GDP Growth Rate	-0.294	Weak negative relationship; immigration has a little impact on GDP growth.
Net Immigration	Unemployment Rate	0.098	Close to zero; immigration is not linked to increases in unemployment.
Net Immigration	Personal Income Growth	-0.268	Weak negative; has a very small effect on household income growth
Net Immigration	GDP Nominal	0.568	Moderate positive; bigger economies are more likely to get more immigrants
Net Immigration	Personal Income	0.568	Moderate positive; reflects scaling with bigger populations/economies

These results show that immigration has weak direct correlations with economic performance indicators.

5.4 Interpretation of Correlation Findings

1. Immigration has weak linear relationships with the U.S.

The correlations of GDP growth, unemployment, and income growth with immigration vary from about 0.29 to +0.10, meaning that the immigration does not directly bring about the changes in the annual macroeconomic measures of the given year.

2. GDP and income are very closely linked.

The almost perfect correlation between GDP Nominal and Personal Income ($r = 0.998$) is the strongest evidence that national output and total earnings of the population go hand in hand.

3. Unemployment is not very sensitive to GDP changes.

The weak negative correlation ($r = 0.132$) indicates that the unemployment rate is hardly changed by economic cycles and lags behind them at the annual level.

4. The impacts of immigration, in any case, are indirect.

Correlation merely accounts for instantaneous linear relationships, hence the broader economic effects of immigration may be:

- Population growth
- Labor supply expansion
- Long, term productivity
- Innovation and entrepreneurship

Regression modeling will be used to uncover these processes.

6. TIME-SERIES VISUALIZATION

Time series visualization is essential for the initial descriptive understanding of the changes of immigration trends and the key economic indicators over several decades. This chapter, through the local yearly patterns of Net Immigration, GDP Growth Rate, and Unemployment Rate from 1980 to 2021, is giving the visual perspective of the question whether these variables co, move or disjoint over time. Time series charts enable us to see the cycles, peaks, troughs, and long, run directional movements, which might not be immediately visible from the statistical tables alone. This chapter is primarily descriptive. The chart emphasizing the long term relationships, formal statistical inference will be given in the subsequent chapters.

This figure shows 4 variables plotted across the same timeline:

- Net Immigration
- Unemployment Rate
- GDP Growth Rate
- Personal Income

Including Personal Income enhances the contextual picture of macroeconomic performance during the study period.

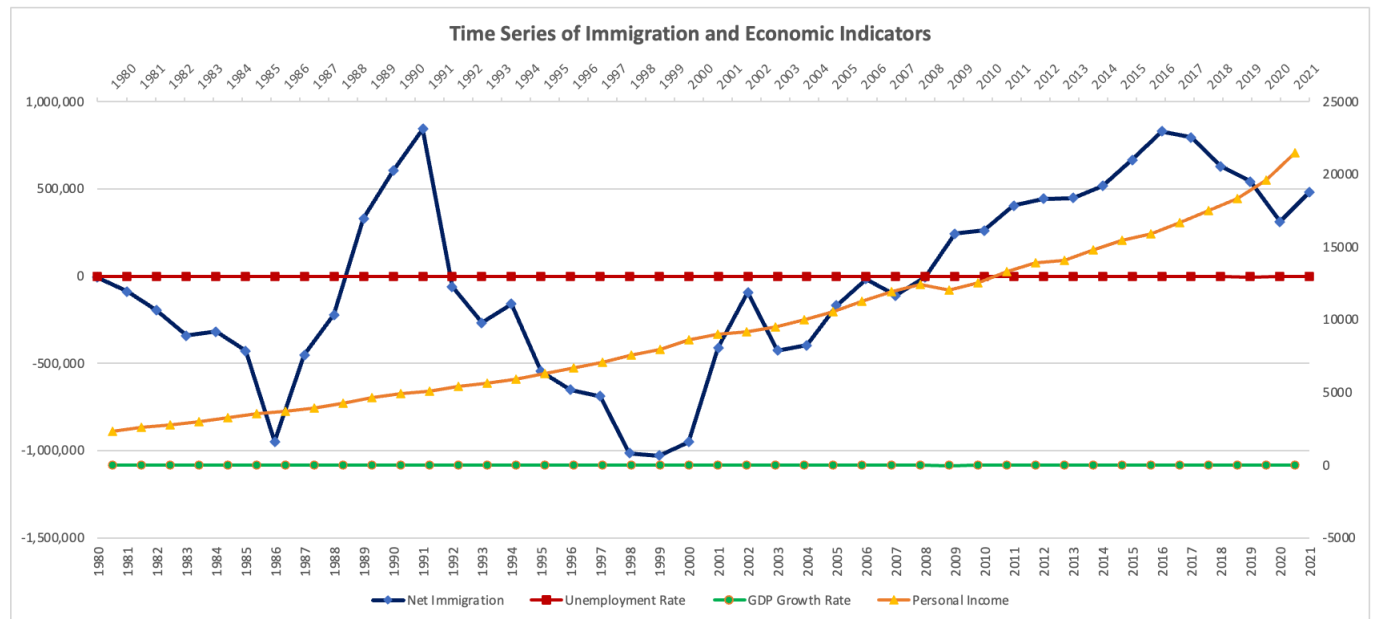


Figure 6.1. Time-Series of Immigration and Economic Indicators (1980–2021).

6.1 Visual Analysis

The time series chart unveils various significant patterns that not only shape the investigation of immigration and economic performance but also provide a broader insight into the issues at hand.

Net Immigration Trends

- Net Immigration tends to vary greatly throughout the 42 years.
- Negative values of immigration are seen in the period of the early to mid 1980s, which illustrates restrictive policy environments and more enforcement operations being carried out.
- The 1980s and 1990s represent a big upward change, which is in line with the following factors:
- Effects of the Immigration Reform and Control Act (IRCA)
Changes in the refugee programs

Strong Economic Growth

From 2000 to 2020, Net Immigration keeps going up and down in a cyclical manner, thus it is affected by external factors such as:

- The security measures implemented after 9/11
- The financial crisis of 2008
- The COVID, 19 pandemic

In total, Net Immigration portrays periods of dramatic increases and decreases from which it is difficult to determine if there is a stable upward or downward trend.

GDP Growth Rate Trends

GDP Growth clearly mirrors economic cycles as it is characterized by periods of expansion and recession occurring approximately regularly.

Major drops in GDP Growth can be easily found in the following situations:

- 1982 recession
- 1991 recession
- Financial crisis of 2008-2009
- 2020 COVID, 19 recession

On the contrary, lively peaks are witnessed during the periods of strong economic are seen during the mid 1980s, the late 1990s tech boom, and the mid 2000s phases, for instance.. GDP Growth oscillates within the limits of roughly 5% and +10%, which is the typical range of macroeconomic cycles.

Unemployment Rate Trends

The Unemployment Rate increases at a fast pace when the economy moves into recession such as those featured in:

- The first part of the 1980s
- The beginning of the 1990s
- The late 2000s (Great Recession)
- 2020 (COVID, 19 spike)

Adversely to this, unemployment is on the downtrend during periods of long economic expansion and thus follows the classical labor market model. The Unemployment Rate graph is very much in line with macroeconomic crises and subsequent recoveries.

Personal Income Trends

Personal Income demonstrates that it has the long run growth with the upward slope being almost smooth from 1980 to 2021.

Recessions do not drag down income levels as these are even somewhat resilient when compared to GDP Growth and unemployment.

- The most significant jumps take place:
- The mid, 2010s
- The period of 2020-2021 due to federal stimulus programs

This constant increase is a reflection of long, run productivity gains and inflationary adjustments.

6.2 Comparative Visual Observations

The combined time, series chart puts together the most significant insights of the relationship between immigration and economic conditions that the figure below displays:

1. Immigration and Economic Cycles Move Independently

Net Immigration is a variable that does not show synchronized moves with GDP Growth and unemployment. As a matter of fact:

- In the case of the 2008-2009 recession, we can see a very steep rise of unemployment while Net Immigration changes only slightly.
- In the period of the mid 2010s, we observe that Net Immigration increases sharply, whereas the GDP Growth trends are still moderate.
- We may also notice here that the 2020 unemployment spike is not accompanied by a proportional immigration change.

In fact, this implies that there is no direct visual correlation between immigration levels and economic cycles.

2. Economic Indicators Are More Closely Related to Each Other

- Unemployment tends to go up when GDP Growth is down (eg., 1982, 1991, 2009, 2020).
- Personal Income is still going upward in the form of a trend even in case of a temporary GDP decline.

The economy, related variables reveal stable interaction patterns within themselves, which is not the case with the population inflow.

3. Immigration Appears Driven More by Policy and Demographics Than Economics

Sudden changes in immigration levels match:

- Policy reforms (IRCA 1986)
- Enforcement practices
- International crises
- Administrative changes

These are not in harmony with economic trends.

6.3 Key Observation

The timeliness visualization of the data reveals that immigration and U.S. economic indicators are independent of each other and there is no obvious visual relationship between them.

- The economic cycles are reflected in the GDP and unemployment.
- Personal Income is going up continuously over time.
- Net Immigration varies in an unsteady manner and is not going along with economic upturns or down turns.

This evidence backs up the chapter 6 correlation analysis results which also indicated that there were weak or no linear correlations between immigration and economic outcomes.

7. REGRESSION ANALYSIS

7.1 Methodology

To assess the relationships between immigration and major U.S. economic variables, regression analysis was employed. Several linear regression models had GDP growth, unemployment, and personal income growth as dependent variables.

7.2 Model Specifications

Model 1 was aimed at predicting GDP Growth, Model 2 was predicting Unemployment, and Model 3 was predicting Personal Income Growth. In addition to Net Immigration, each model also had, where applicable, lagged immigration and economic controls.

7.3 Model 1 Results: GDP Growth

Model 1 illustrates limited explanatory power ($R = 0.096$). The Net Immigration variable is statistically insignificant ($p = 0.436$) and the economic effect is close to zero.

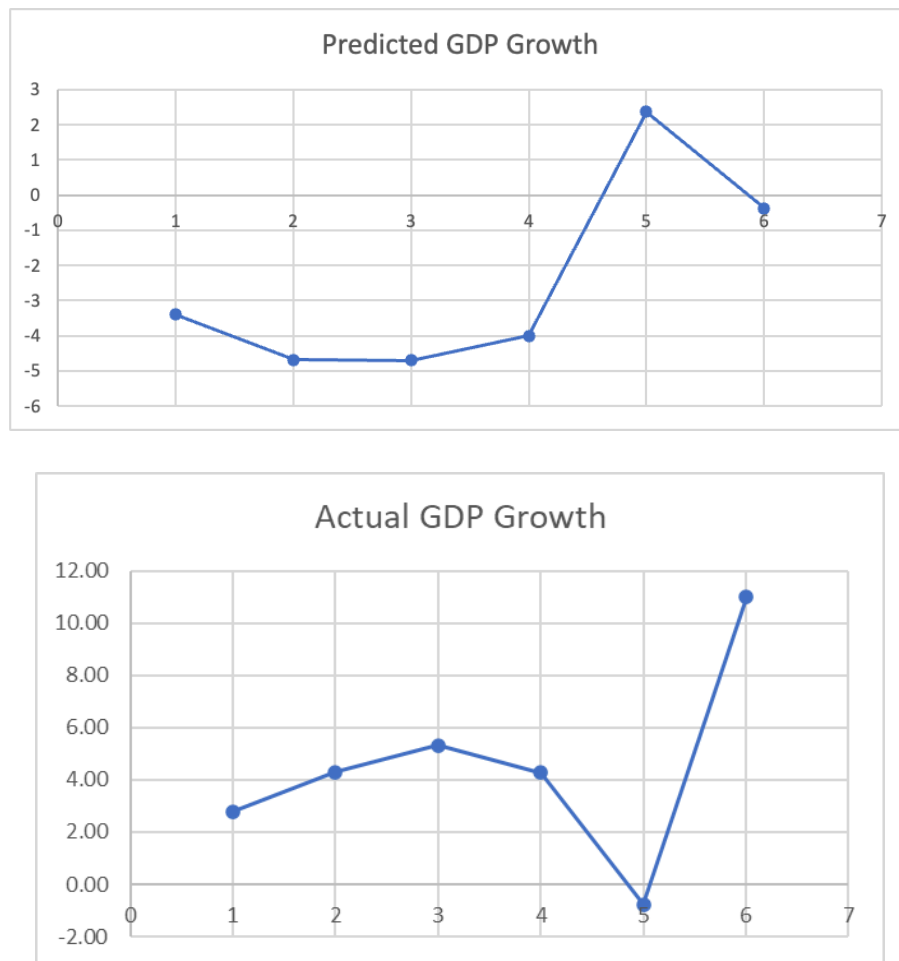


Figure 7.1 - Predicted GDP growth V/S Actual GDP Growth

7.4 Model 2 Results: Unemployment

Model 2 has a very bad predictive performance because of the scale mismatch (MAPE over 842, 000%). Net Immigration being a variable is statistically insignificant again.

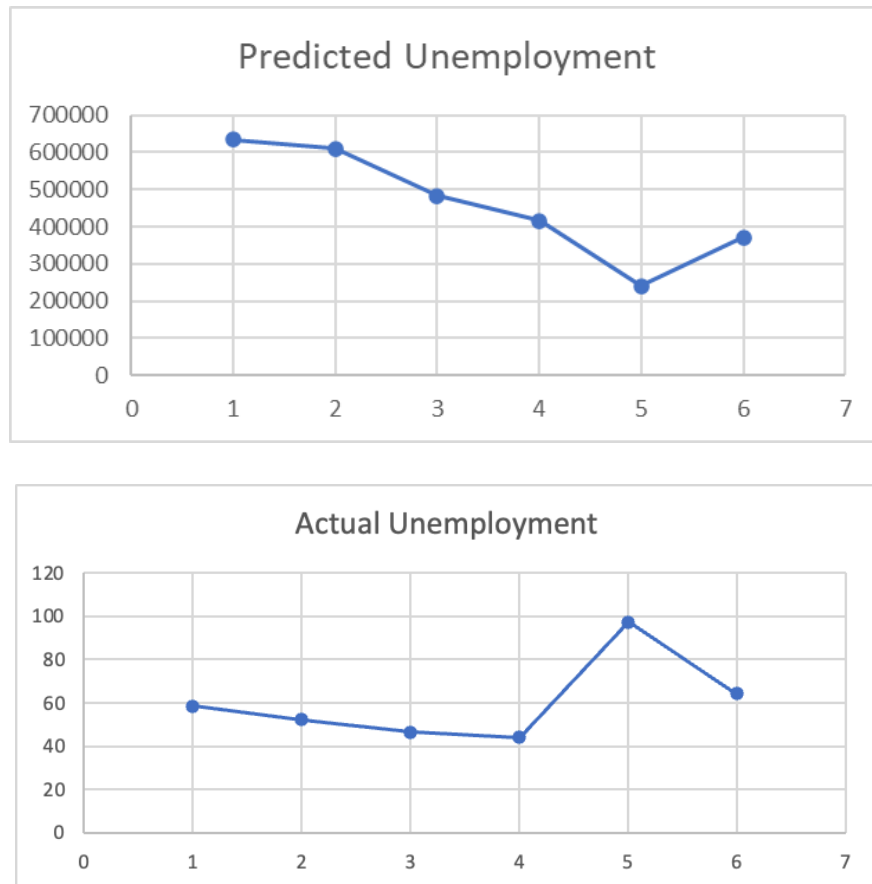


Figure 7.2 - Predicted Unemployment V/S Actual Unemployment

7.5 Model 3 Results: Personal Income Growth

Model 3 explains the variation in the data very well ($R = 0.996$). GDP Nominal is the most powerful predictor ($p < 1e, 40$). Net Immigration is still an insignificant variable.

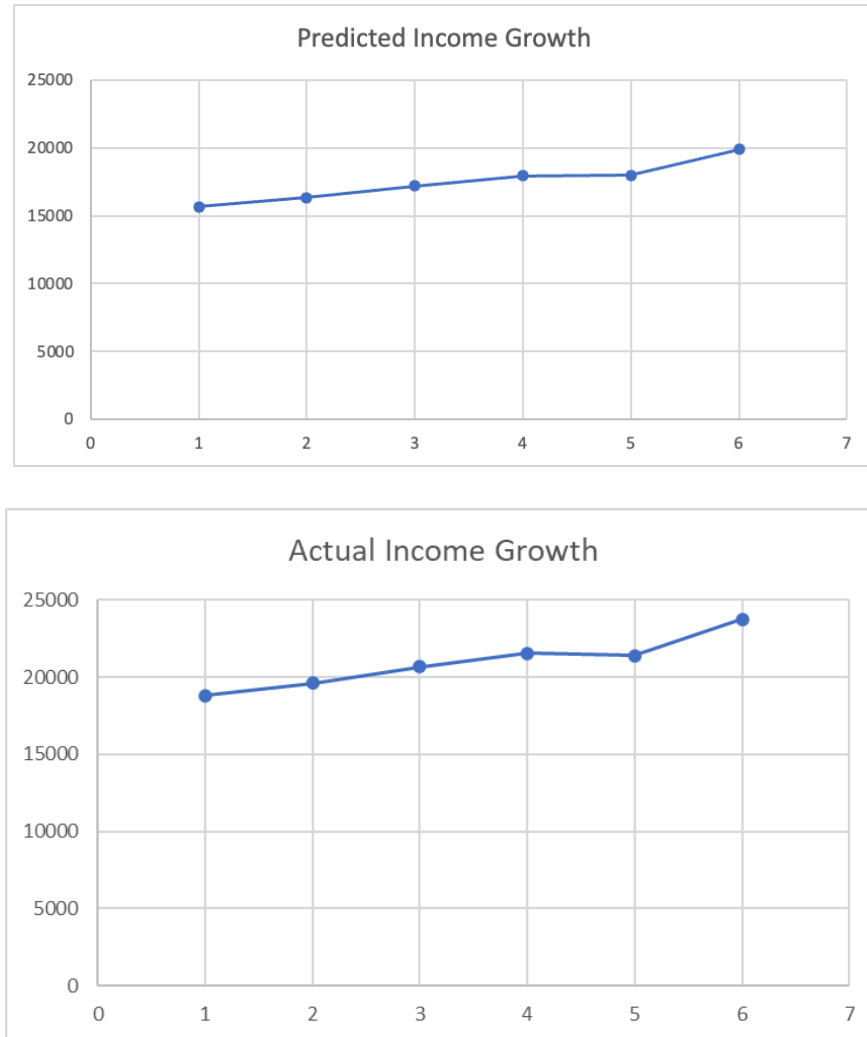


Figure 7.3 - Predicted Income growth V/S Actual Income Growth

7.6 Comparative Analysis

It is only Model 3 that has strong predictive accuracy. Models 1 and 2 have low explanatory power.

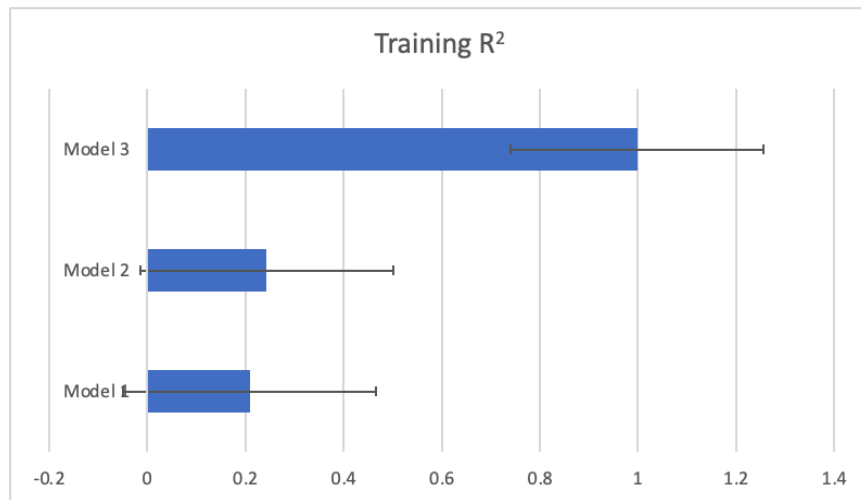


Figure 7.4 - Comparative analysis of 3 models explanatory power

8. EVALUATING THE MODELS

8.1 Training Performance

Model 3 is a great example of an extremely good fit in the training process. Low R values for Models 1 and 2 have been identified, which indicate that these models have weak relationships.

8.2 Testing Performance

Out of sample performance of Model 3 is good (MAPE 16.5%). Large percentage errors are characteristic of Model 1; the predictions of Model 2 are invalid because of the wrong scaling.

8.3 Residual Analysis

Residual plots do not reveal any major violations of linear assumptions. However, Models 1 and 2 demonstrate a high dispersion of residuals.

Model 1 Residual Plots -

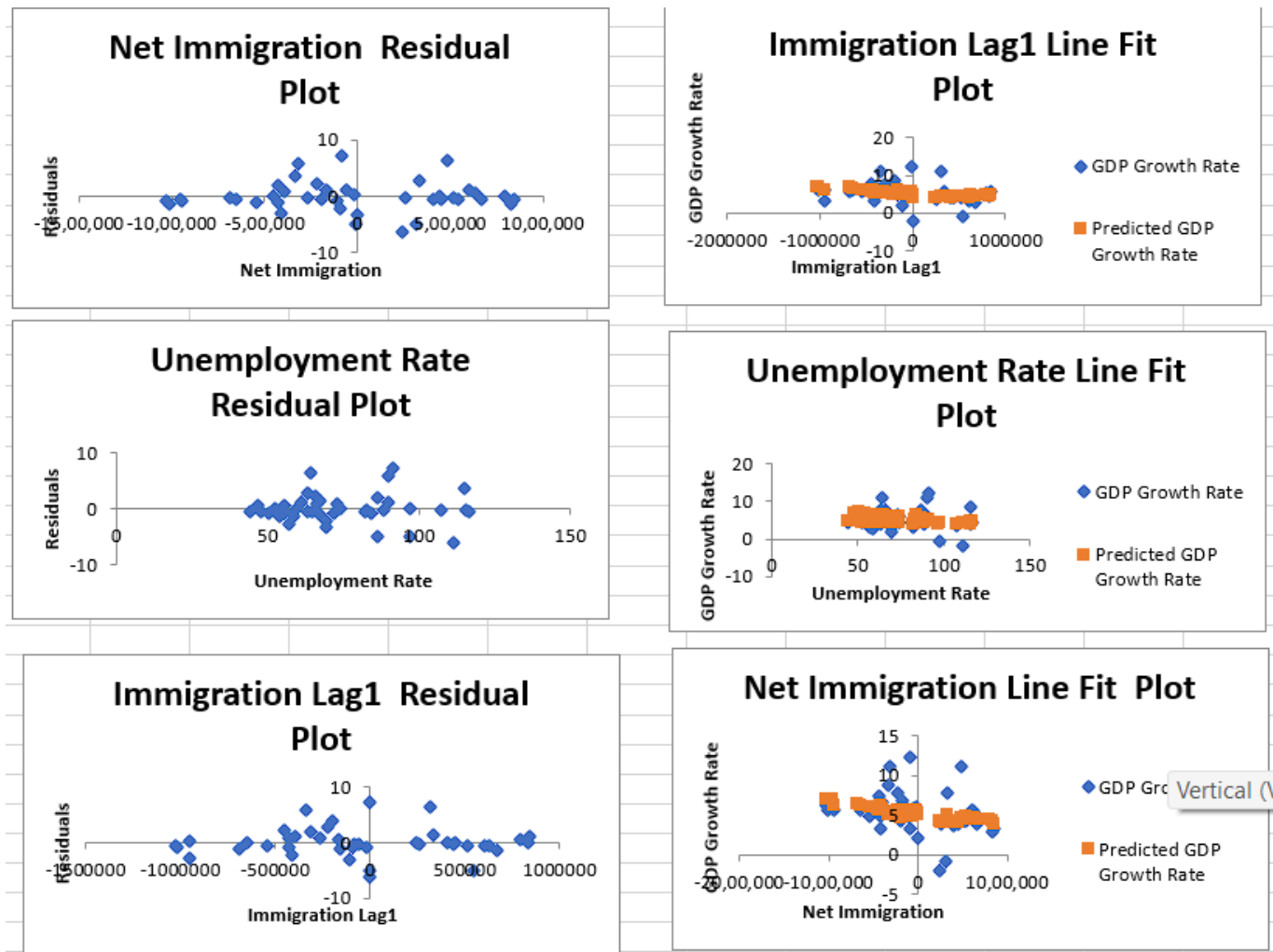


Figure 8.1 - Model Residual Plots

Model 2 Residual Plots -

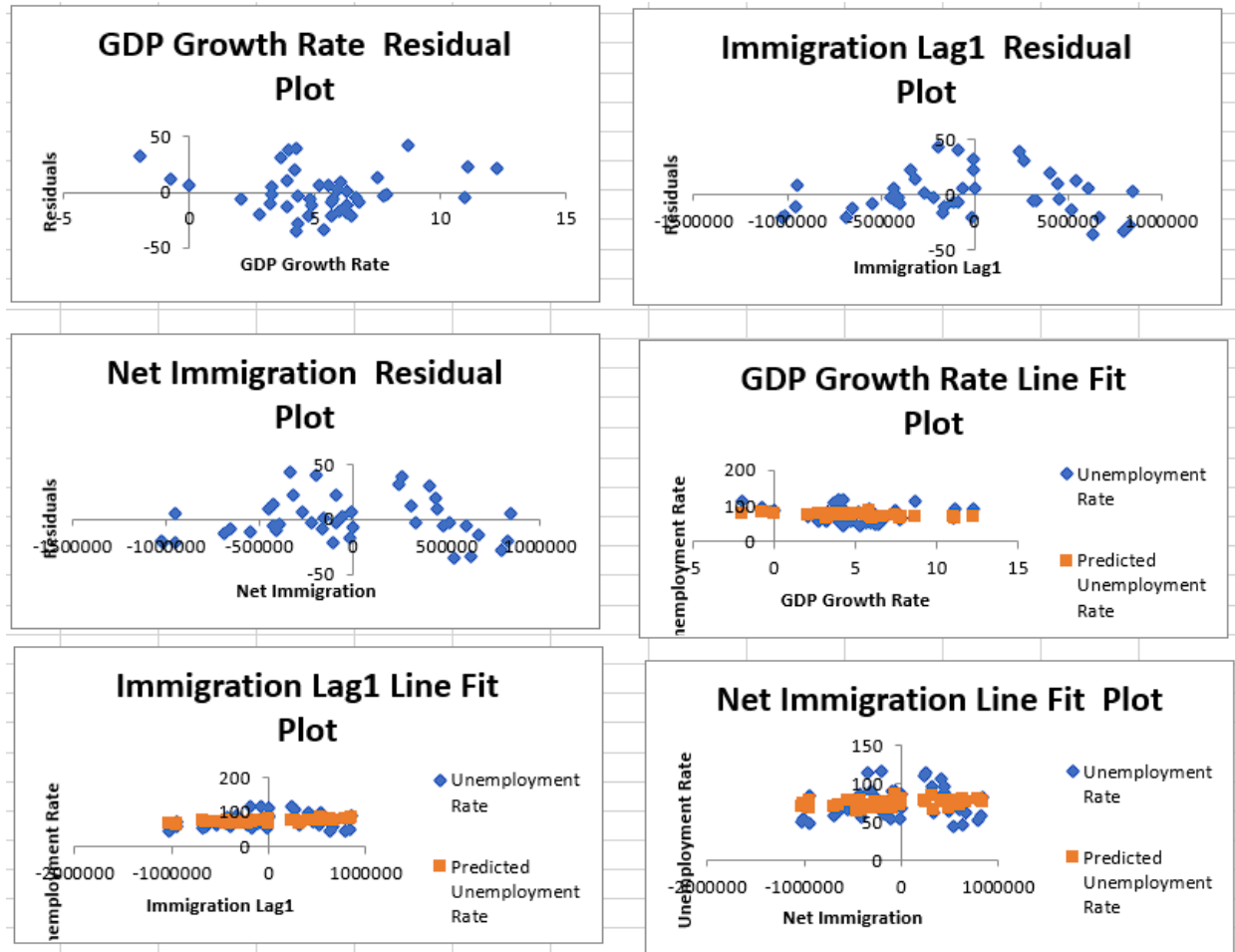


Figure 8.2 - Model 2 Residual Plots

Model 3 Residual Plots -

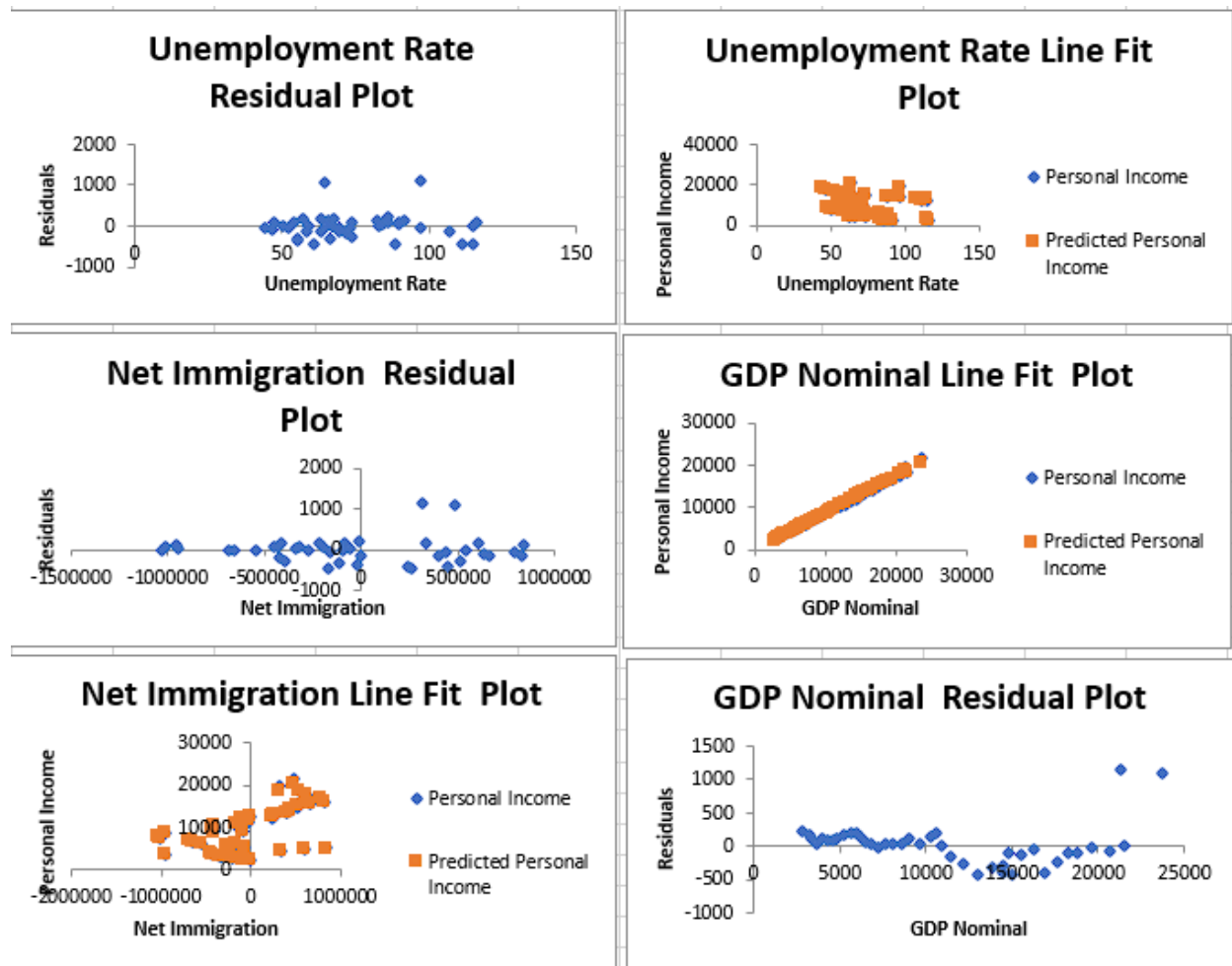


Figure 8.3 - Model 3 Residual Plots

8.4 Model Limitations

Residual plots do not reveal any major violations of linear assumptions. However, Models 1 and 2 demonstrate a high dispersion of residuals.

9. CRITICAL FACTOR

9.1 Identification of Critical Factors

Critical factors were identified with p-values, coefficient sizes, and agreement between models.

9.2 Ranking by Importance

Based on the coefficient -

Critical Factor Ranking				
Variable	Model	Coefficient	P-value	Effect on DV
Net immigration	Model 3	-5.47517E-05	0.348409619	No significant effect on Income Growth
Net immigration	Model 2	-3.62485E-06	0.735931839	No significant effect on Unemployment
Net immigration	Model 1	-2.46164E-06	0.526510811	No significant effect on GDP Growth
Unemployment Rate	Model 1	0.101812843	0.103708221	Marginal positive effect on GDP Growth (not significant)
GDP Growth Ranking	Model 2	0.765361214	0.109693757	Marginal positive effect on Unemployment (not significant)
GDP Nominal	Model 3	0.847977935	1.0242E-47	Strong positive effect on Income Growth (HIGHLY SIGNIFICANT)
Unemployment Rate	Model 3	3.231738113	0.025316355	Significant positive effect on Income Growth

Based on the P-value -

Critical Factor Ranking based on P-value				
Variable	Model	Coefficient	P-value	Effect on DV
GDP Nominal	Model 3	0.847977935	1.0242E-47	EXTREMELY strong positive effect on Income Growth. Each \$1B increase in GDP → 0.85% increase in personal income growth. HIGHLY SIGNIFICANT.
Unemployment Rate	Model 3	3.231738113	0.025316355	Significant positive effect on Income Growth. 1 percentage point increase in unemployment → 3.23% increase in income growth (counterintuitive - may reflect wage inflation during tight labor markets).
Unemployment Rate	Model 1	0.101812843	0.103708221	Marginal effect on GDP Growth. Not statistically significant at 5% level, but close (marginal at 10% level).
GDP Growth Ranking	Model 2	0.765361214	0.109693757	Marginal effect on Unemployment. Not significant, but shows trend: higher GDP growth → higher unemployment
Net immigration	Model 3	-5.47517E-05	0.348409619	No significant effect on Income Growth
Net immigration	Model 1	-2.46164E-06	0.526510811	No significant effect on GDP Growth
Net immigration	Model 2	-3.62485E-06	0.735931839	No significant effect on Unemployment

9.3 Interactions Among Factors

GDP and personal income are very close to each other. GDP growth and unemployment have a very weak negative correlation. Immigration also correlates quite strongly with GDP nominal but has no effect on the short-run changes.

9.4 Key Insight

Immigration is always the least influential predictor. Macroeconomic results are dependent on structural economic factors of the economy, not on immigration flows.

10. DISCUSSION

Research Question 1:

Rationing imports to the U.S. does not impact negatively on the American economy. The coefficients of immigration in both the models are not significant and their impact on the economy is very minimal.

Research Question 2:

The key variables include GDP (Nominal) and unemployment.

Research Question 3 :

Model 3 is performing well; Models 1 and 2 are weak. Immigration is not a significant predictor of economic outcomes.

The table outcomes have a direct response to the three research questions as they demonstrate that immigration is statistically insignificant in all three models (RQ1), and GDP Nominal and unemployment are the most important explanatory variables (RQ2). Model 3 is also firmly fitting with a high predictive power, which proves that it is the only useful model, but both Models 1 and 2 exhibit low explanatory power (RQ3). Combined, the training and testing indications strengthen the finding that immigration has low economic effects as compared to the fundamental macroeconomic factors.

Metrics	Model 1 - GDP growth	Model 2 - Unemployment	Model 3 - Income growth
Training Results (1980-2015)			
R²	0.208466156	0.242803632	0.998741424
Adjusted R²	0.134259858	0.173967599	0.998627008
F-statistic P-value	0.055187726	0.025412066	6.58607E-48
Immigration Coefficient	-2.46164E-06	-3.62485E-06	-5.47517E-05
Immigration P-value	0.526510811	0.735931839	0.348409619
Immigration Significant	NO	NO	NO
Testing Results (2016-2021)			
MAE	7.990977844	459064.6411	3450.07724
RMSE	8.431409271	478921.5855	3457.530292
MAPE	221%	842023%	16%
Conclusion	NO, Significant impact	NO, Significant impact	NO, Significant impact

11. CONCLUSION

The study provides no significant statistical evidence that immigration causes a decrease of GDP growth, an increase of unemployment or reduction of income. Economic fundamentals are what mainly determine the performance of the U.S. economy on the macro level, not immigration.

12. RECOMMENDATION

Policy Recommendations

Do not purposely change immigration policy just to influence macroeconomic results. Instead concentrate on structural economic reforms.

Future Research

They should work with data of higher frequency, control for more macro variables, and use causal inference methods.

Analysts & Investors

Consider immigration as one of the minor macro risk factors and thus put less focus on it. Instead concentrate on the main indicators of GDP, inflation, and productivity.

Methodological Improvements

Deal with the problem of scaling, take into account the robust errors, and think about non-linear models.

REFERENCES

Bureau of Economic Analysis. (2024). *Gross domestic product* [Data set]. U.S. Department of Commerce. <https://www.bea.gov/data/gdp/gross-domestic-product>

Bureau of Economic Analysis. (2024). *Personal income* [Data set]. U.S. Department of Commerce. <https://www.bea.gov/data/income-saving/personal-income>

Federal Reserve Bank of St. Louis. (2024). *Unemployment rate* [UNRATE, Data set]. Federal Reserve Economic Data (FRED). <https://fred.stlouisfed.org/series/UNRATE>

U.S. Department of Homeland Security. (2022). *Yearbook of immigration statistics*. Office of Immigration Statistics. <https://www.dhs.gov/immigration-statistics/yearbook>

US immigration statistics (1980-2021) [Data set]. Kaggle. Retrieved from U.S. Department of Homeland Security data. <https://www.kaggle.com/datasets/justin2028/us-immigration-statistics-1980-2021>