Effects of Population Change

Seyi Aro

2024-04-30

## Abstract

The events of the pandemic left an endearing effect on the economies of many countries. In the U, S., several new headlines reported an increase in the migration patterns with the country in the early periods of lockdown lifts. This paper explores the patterns of population change in all 50 states and District of Columbia from 2021-2022. Additionally, it uses linear regression model to analyze the relationship between population change and other economic factors.

## Introduction

In the early periods following the pandemic and lockdowns, there were several reports surrounding state-to-state migration in the U.S. According to the U.S. Census Bureau, the number of people who moved between states rose from 7.9 million in 2021 to 8.4 million in 2022 (Ismail, 2023). One of the motivating factors for this move is the rise in cost of living. A survey by a moving company found that the number of people who moved because of high cost of living within their current state doubled between 2020 and 2021 from 3% to 6.7% (United Van Lines, 2022).

This paper explores the relationship between states that saw a decrease in population and their cost-of-living indexes. The assumption is that the states that saw a decrease in population size have, on average, higher cost of living indexes than states that increased in size. It also looks at other economic factors that may contribute to the change in population among different U.S. states and District of Columbia. A linear regression model was used to explore the relationship. Between them.

## Literature

Two theories exist among economist and those that study the impact of population growth. The first is that population growth stimulates economic growth and the second posits that there is a negative relationship between the two variables.

The traditional view is that increasing rates of population has adverse effects on economic growth (Johnson, 1999). Many studies have been conducted to either support or disagree with this view. Sinding (2009) argues that there is substantial evidence to support the view that a declining demographic positively impacts economic growth in both the micro and macro levels of economic development. The author recommends implementing policies that support a reduction in birth rate among developing countries such as those in Africa.

In contrast, Thuku et. Al (2013) finds using a time series data between the period of 1963 to 2009 that among the three theories, there is a positive relationship between population growth and economic growth in Kenya, a developing country on Africa. They argue that increasing demographic reinforces technological progress and economies of scale.

While Ali et al. explains that while population growth positively impacts economic growth it negatively affects other economic factors that can in turn stifle the growth of an economy. The authors write that although growth in population leads to economic growth, it negatively impacts unemployment rates.

A great deal of research has been done on the role population growth plays on economic growth. This paper looks into how some significant economic variables affect population.

## Data

The sample data for this project was collected from three sources: the U.S. Bureau of Economic Analysis (BEA), Kaggle, and GitHub. It includes population change, employment, per capita income, and personal income data in 2021 from BEA. Cost of living data in 2023 from Kaggle and lastly, information about all 50 states and District of Columbia from GitHub.

The data retrieved from BEA contained the GeoFIPS, a number used to uniquely identify geographic areas in the U.S. and GeoName (the state names) of all 50 state in the U.S. including District of Columbia.

The cost of living data downloaded from kaggle contained indicies of the total cost of living (COL), COL rankings, grocery, transportation, housing, utilities, and miscellanaus expenses in all 50 states in the U.S., District of Columbia and Puetorico.

Lastly, state data served as bridge to join all the data.

## Methodology

This section details the process of reading and cleaning all the tables, and concatenating them into one dataframe. After downloading the files, they were loaded into R for cleaning. A quick review of the files in Excel showed that they were mostly cleaned.

Most of the data preparation process involved selecting only the rows needed for analysis. For example, all the files retrieved from the BEA had rows describing the data and its source, this were handled with one line of code that selected only the rows with the needed data. Also, using the stringr package, the strings containing unwanted characters and leading space were removed.

After cleaning all the data files individually, they were combined into one table. The data downloaded from BEA: population, employment, income per capita, and personal income were first joined by their ‘GeoName’ and ‘GeoFIPS’ columns using the join function in the dplyr package to form the ‘economy’ table.

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

To pair cost of living with the economy dataset, the table containing information about all 50 states and the district of columbia was first merged with cost of living dataset by their abbreviations and then joined to the economy data to form a ‘whole\_tb’ data.

## GeoFips GeoName stusps col\_index col\_rank pop\_21 emp\_21 perc\_21  
## 1 01 Alabama AK 124.4 47 0.5 2769464 50059  
## 2 02 Alaska AL 88.8 5 -0.2 443047 65662  
## 3 04 Arizona AR 90.3 9 1.3 4086802 56420  
## 4 05 Arkansas AZ 107.2 37 0.6 1686444 51636  
## 5 06 California CA 134.5 49 -0.3 23934549 76991  
## 6 08 Colorado CO 105.5 35 0.5 3988907 71923  
## perinc\_21  
## 1 252791.9  
## 2 48207.9  
## 3 409885.7  
## 4 156361.5  
## 5 3013676.9  
## 6 417967.7

## Results

This section discusses the results from doing analysis, and using regression models to explore the relationship between population and other variables in the dataset, specifically cost of living. It first uses a simple regression model to look at the relationship between population change and any one variable before using a multiole linear regression model for further analysis.

The table below shows a summary of all the continuous variables present in the data.

##   
## Please cite as:

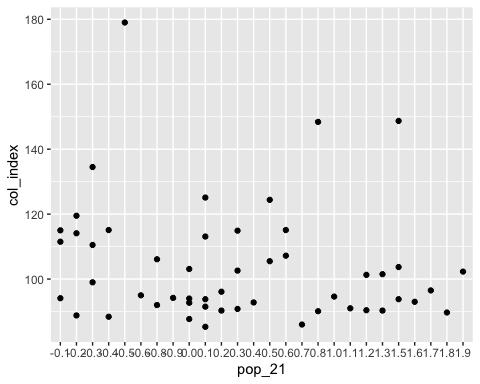
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

##   
## <table style="text-align:center"><tr><td colspan="6" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Statistic</td><td>N</td><td>Mean</td><td>St. Dev.</td><td>Min</td><td>Max</td></tr>  
## <tr><td colspan="6" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">col\_index</td><td>51</td><td>104.002</td><td>18.241</td><td>85.300</td><td>179.000</td></tr>  
## <tr><td style="text-align:left">col\_rank</td><td>51</td><td>26.490</td><td>15.305</td><td>1</td><td>52</td></tr>  
## <tr><td style="text-align:left">emp\_21</td><td>51</td><td>3,975,531.000</td><td>4,543,920.000</td><td>420,429</td><td>23,934,549</td></tr>  
## <tr><td style="text-align:left">perc\_21</td><td>51</td><td>62,459.980</td><td>9,726.161</td><td>46,577</td><td>97,498</td></tr>  
## <tr><td style="text-align:left">perinc\_21</td><td>51</td><td>419,466.900</td><td>526,660.100</td><td>39,603.500</td><td>3,013,677.000</td></tr>  
## <tr><td colspan="6" style="border-bottom: 1px solid black"></td></tr></table>

The results shows that 16 states in the U.S. saw a reduction in population size. Among them are, perhaps not surprisingly, New England states like New York, New Jersey, Pennsylvania, Masschussetts, Maryland, etc. Other notable states are Carlifornia, Hawaii, Illinois and Alaska. On the other hand, 35 states saw an increase in population after the effects of the pandemic. Furthermore, the average cost of living index among states with negative change in population is 109.8, compared to 101.4 among states that saw a positive change.

Before building a regression model, a scatter plot of population change plotted against cost of living shows a positive relationship. This is different from the assumption that cost of living negatively impacts population.



The summary table for the regression model that analyzes the relationship between population and cost of living index shows that one point increase in the cost of living index in a particular state results in a 0.005 percent decrease in the population. However, the relationship between both variables is not statistically significant (p-value: 0.418 > 0.005).

What happens when other variables are added to the model?

##   
## <table style="text-align:center"><tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"></td><td><em>Dependent variable:</em></td></tr>  
## <tr><td></td><td colspan="1" style="border-bottom: 1px solid black"></td></tr>  
## <tr><td style="text-align:left"></td><td>pop\_21</td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">log(col\_index)</td><td>0.033</td></tr>  
## <tr><td style="text-align:left"></td><td>(0.706)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">log(emp\_21)</td><td>2.654<sup>\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(1.211)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">log(perc\_21)</td><td>0.047</td></tr>  
## <tr><td style="text-align:left"></td><td>(0.815)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">log(perinc\_21)</td><td>-2.564<sup>\*\*</sup></td></tr>  
## <tr><td style="text-align:left"></td><td>(1.161)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td style="text-align:left">Constant</td><td>-7.527</td></tr>  
## <tr><td style="text-align:left"></td><td>(11.544)</td></tr>  
## <tr><td style="text-align:left"></td><td></td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">Observations</td><td>51</td></tr>  
## <tr><td style="text-align:left">R<sup>2</sup></td><td>0.131</td></tr>  
## <tr><td style="text-align:left">Adjusted R<sup>2</sup></td><td>0.056</td></tr>  
## <tr><td style="text-align:left">Residual Std. Error</td><td>0.719 (df = 46)</td></tr>  
## <tr><td style="text-align:left">F Statistic</td><td>1.735 (df = 4; 46)</td></tr>  
## <tr><td colspan="2" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</em></td><td style="text-align:right"><sup>\*</sup>p<0.1; <sup>\*\*</sup>p<0.05; <sup>\*\*\*</sup>p<0.01</td></tr>  
## </table>

Using a multiple regression model, one unit increase in cost of living index leads to 0.00159 increase in pouplation change and the relationship is not statistically significant since p-value, 0.784, is greater than 0.005. Also, one percent increase in employment leads to 2.633 percent change in population. A similar result is observed in income per capita which increase population by 0.05 percent. While one percent increase in personal income leads to a 2.545 decrease in population.

The overall model is not statistically significant since p-value (0.1586) is greater than 0.05 and F-stat is 1.735. Lastly, only 5.6 percent of the variance in the change in population is explained by the independent variables.

## Implications

This study is largely exploratory and therefore, makes no conclusion about the relationship between variables. An indepth study of the effect of population and other economic variables will require larger data perhaps a time series data that analyzes the relationships among variables during the periods before the pandemic and after.

## Conclusion

This project explores the process of using R to collect and clean data, perform analysis and interpret statistical models using simple and linear regression models.

## References