U.S. Internal Migration Over Time

Teagan Zuniga*

December 13, 2023

*Graduate Student, Department of Political Science, University of California, Merced, Email: tzuniga3@ucmerced.edu

1 Introduction

There is an abundance of work that examines migration *to* the United States (Massey, 1987; Durand & Massey, 1992) from a foreign country or migration amongst (Fearon & Shaver, 2020) or within foreign countries (Weiss & Korn, 2006; Shultz et al., 2014). However, there has been much less work done on migration within the United States.[1] Therefore, I have three interrelated research questions for this project: 1) how has internal migration[2] within the United States changed over time? 2) when people do migrate internally, where do they migrate from and 3) where do they migrate to? Typically, migration research focuses on 1) cases outside of the United States and/or 2) cases of out-of-state (i.e., international) migration. Answering these questions would not only add to the existing research on migration, but would uniquely contribute to this growing body of research. Understanding migration is not just an international or comparative issue, but a domestic one as well.

This project was largely motivated by the recent news articles proclaiming that there was an exodus of Californians moving to Texas. One such article is titled "Droves of Californians are moving to Texas. Here's the life they are finding" (Castleman, 2023). As previously mentioned, there is little work done on migration within the United States. Therefore, prior to this project, I did not know much about U.S. internal migration. The only expectations

[1]There are some exceptions to this, of course (see Greenwood, 1975; Molloy, Smith, & Wozniak, 2011). However, there is a skew in the migration literature to studying migration outside of the United States rather than internally. Therefore, there is still much to be understood about migration within the United States.

[2]Internal migration is defined as the movement of people from one geographic unit to another, where both units are both located within the same country.

I had came from the news articles mentioned above[3] and an inclination that the bigger states, such as California, may receive more migrants than smaller states, such as Rhode Island. There were no expectations about what states would have the most people leaving.

Understanding these questions have implications for various aspects of American life. As an example, consider a situation in which migration has increased over time and this increase is concentrated within a single state, such as California. A clear consequence of this population increase is a potential strain on the state's infrastructure. Additionally, if we know that such a phenomenon exists, additional important questions arise, such as the following: why are Americans migrating to California, opposed to other states? Conversely, where are these migrants coming from and why are they leaving their home county/state? Understanding this, and other questions, can help us understand current American life as well as foresee future changes/implications.

2 Methods

The data that will be used for this project comes from Hauer & Byars (2019). Matthew Hauer is a sociologist at Florida State University whose work relates to migration and James Byars works in scientific computing in the Institute of Government at the University of Georgia. They created the dataset with the intention of it being used to study migration.

[3]This report will be unable to look at the recent phenomenon of migration between California and Texas because, as noted later in this report, the data used for this project only spans from 1990 to 2010.

As the authors note, the IRS has released migration data since 1990.[4] The IRS estimates migration using tax filings. A person's tax filing in $year_i$ is compared to their tax filing in $year_{i+1}$. A person is considered to have migrated if their address in the $year_{i+1}$'s tax filing does not match the address $year_i$'s tax filing. What is released by the IRS is aggregate counts of how many people are considered to have migrated from $county_1$ to $county_2$ in a given year.[5] This allows scholars to explore county-to-county migration in the United States; however, there are three notable problems with this data.

First, this data is likely not fully representative of the U.S. population. Since the IRS is relying on tax filings, those who do not file taxes are underrepresented in the data (Gross, 2005; DeWaard, Curtis, & Fussell, 2016), particularly undocumented individuals, the poor, the elderly, and college students (Gross, 2005). However, given that the majority of U.S. householders file tax returns (Molloy, Smith, & Wozniak, 2011), this should not be a substantial problem, especially since I do not aim to answer questions related to *who* migrates.[6]

Second, while counties tend to be fairly stable geographic units, some do change over time. This data does not account for these changes. This is because the authors aimed to preserve the original data as much as possible – illustrating the importance of mutability/immutability of the data to the authors. Therefore, it may be more plausible to focus on state-to-state migration

[4]The historic time series of the data was broken in 2011, as the IRS was improving their data processing methods, so this data only goes up to 2010.

[5]Cases where 10 or less people moved between two counties are removed from the dataset. This is to protect the anonymity of individuals.

[6]This IRS data does not include demographic data so it would be impossible to answer questions about who, even if the data was representative of the U.S. population.

Figure 1: Data Journey

than county-to-county migration since state borders have not changed between 1990 and 2010.

Lastly, while the IRS migration data has great potential for the migration research community, the IRS posts the data in a very hard to manage format! The data is posted across over 2000 data files. Luckily, the goal of Hauer & Byars (2019) is to process this data and provide it to researchers as a single, flat file. While this is very helpful, it is important to note that this is an extra step in the data's journey and opens up the possibility for more error in the data. Luckily, Hauer & Byars (2019) post their R code, allowing users to check for any such errors. A simplified version of the data journey is presented in **Figure 1**.

To restate, I have three interrelated research questions: 1) how has internal migration within the United States changed over time? 2) when people do migrate internally, where do they migrate from and 3) where do they migrate to? Therefore, the data provided by Hauer & Byars (2019) on U.S. county-to-county migration is suitable to answer my research questions.[7]

[7]There are other datasets that are frequently used to study subnational migration in the

To test these questions, I conduct an exploratory data analysis (EDA). Specifically, I rely on visualizations of the data to make preliminary, non-causal conclusions. The first portion of the following section will present descriptive statistics of the data. The second portion will present the preliminary findings related to my research question.

3 Results

3.1 Descriptive Statistics

The data from Hauer & Byars (2019) began as a district-year dataset. Districts change over time, as previously mentioned, and the dataset does not capture these changes. As a result, this report focuses on migration between states, not within states, and the dataset was reformatted accordingly. Specifically, the dataset was changed from district-year to state-year.

This was a tedious task, however, as the locations were included as FIPS codes[8] rather than district and state names. For simplicity, I will outline the four main steps of this process.[9] First, the district FIPS code was split into the state identifier (i.e., the first two numbers) and district identifier within

United States (the Decennial Census and the American Community Survey); however, those data sources rely on self-reported migration, are not in as readily usable formats, and rely on less concise temporal ranges. For example, the Decennial Census asks, "Where did you live five years ago?" This would not allow me to measure year-to-year migration. Therefore, I have chosen to use the processed IRS migration data provided by Hauer & Byars (2019).

[8]A FIPS code, as described by the Federal Communications Commission, are "...numbers which uniquely identify geographic areas. The number of digits in FIPS codes vary depending on the level of geography. State-level FIPS codes have two digits, county-level FIPS codes have five digits of which the first two are the FIPS code of the state to which the county belongs."

[9]Further details of this process are provided in the code.

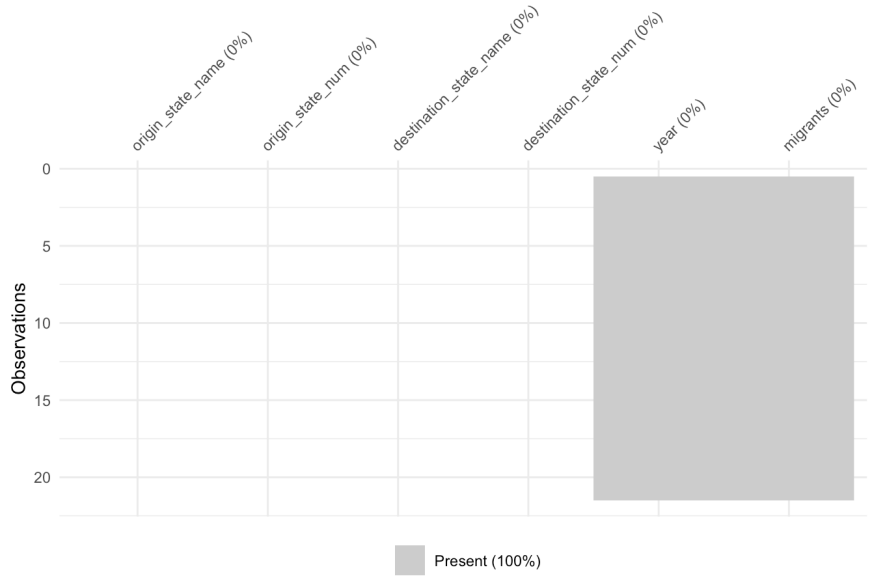

Figure 2: Missingness in the Data

the code (i.e., the remaining numbers). Second, I created a state FIPS code key that matches the state code with the state name. Third, I merged the dataset onto this key. Lastly, the dataset was aggregated up to the state-year level. This resulted in 51 unique states[10] and 21 unique years.

The next key step was to check for whether there was an missingness in the data. **Figure 2** presents all of the variables in the data set as well as whether there are any missingness in those variables. There is no missingness in any of the variables, so this will not be a concern when conducting the preliminary analysis.

At this point in the paper, with the presentation of **Figure 2**, it is beneficial to describe each of the variables in the dataset. Each variable is described

[10]This includes the District of Columbia

below:

- **origin_state_name**: The name of the state where the migrant is from
- **origin_state_num**: The FIPS state code of the state that the migrant is from
- **destination_state_name**: The name of the state where the migrant has traveled to
- **destination_state_num**: The FIPS state code of the state that the migrant has traveled to
- **year**: The year that the migration took place
- **migrants**: The total number of migrants that traveled between the origin state and destination state within the given year

Next, I looked at whether there was a skew in the number of migrants. It would be expected for there to be a positive skew, as most state-years likely have few to no migrants in a year. **Figure 3** presents the distribution of the migration variable. In this plot, the x-axis represents the number of migrants and the y-axis is the number of cases that had that many migrants. As expected, most state-years have no migrants. The number of cases decreases as you move along the x-axis. The next section will cover the preliminary results.

3.2 Preliminary Analysis

My first item to answer is how has internal migration within the United States changed over time. **Figure 4** shows how the total number of migrants has

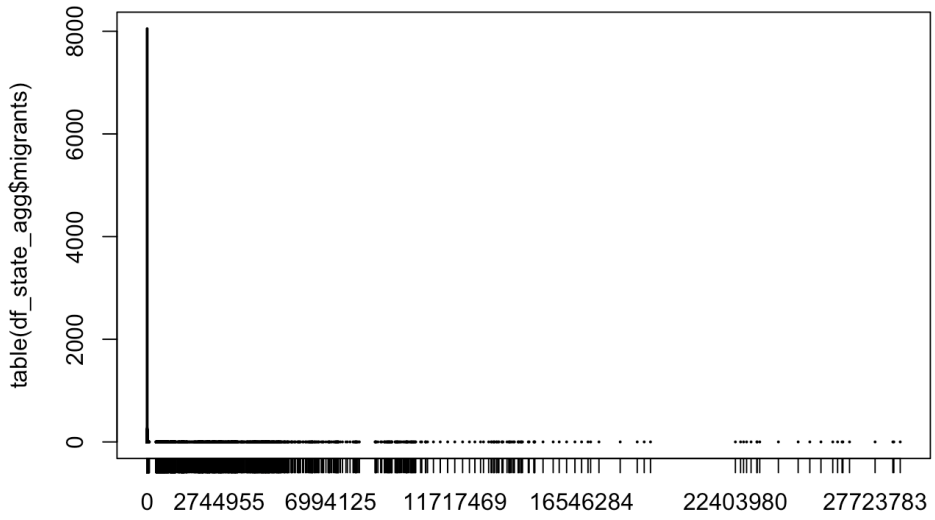

Figure 3: Migration Skew

changed over time. There has been a consistent increase in migrants over time, but why?

There may be several reasons for this. First, the total population in America has grown drastically during this time period. The U.S. Census Bureau reports a 9.7 % population increase between 2000 and 2010 alone. There was a 13.2% increase in the 1990's (Mackun, 2021). Therefore, this population increase may explain the consistent increase in migrants over time.

Second, there may be factors that are enticing individuals to migrate that are present now that were not in earlier years. Although this data does not allow an exploration of this, it is possible that economic, politics, or social changes in specific states may draw people to or push people out of these states. Consider, for example, if gay marriage was legalized in state were it was previously illegal. That may draw specific populations to that state.

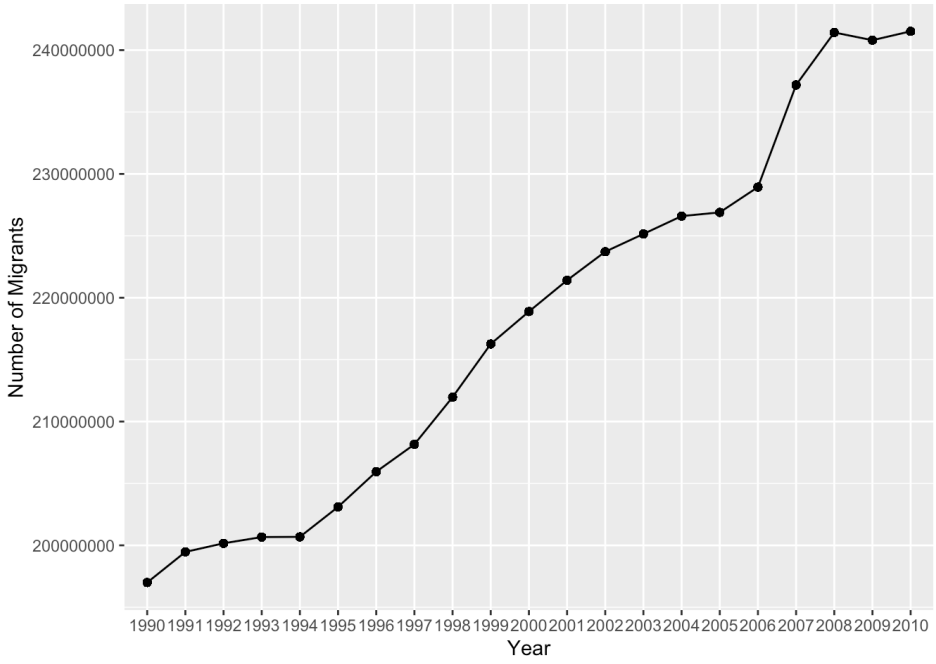

Figure 4: Total Number of Migrants Over Time

This may be true for a multitude of political issues (e.g., abortion rights or legalization of marijuana). On the other hand, economic reasons may pull people out of their origin state and to specific destination state. Returning to the motivating example of this paper, a lot of public discourse claims that Californians are moving to Texas because the land and homes are significantly cheaper there than in California. People who cannot buy a home in California can in Texas. These reasons, amongst others, may explain this increase in migrants over time.

The last two questions relate to where people migrate to and from. However, in order to do that, I had to create variables that represent how many migrants when to a given state in a specific year, as well as how many migrants

left a state in a given year. This was able to do be done by summarizing how many migrants left from or went to a state in a given year. To be more specific, let's take the total number of migrants arriving to a destination state. To create this variable, I summarized how many migrants that arrived to that destination state in a given year from all origin states.

The second question this paper aims to answer is where do people migrate to when they do migrate. In other words, what are the destination states of migrants within the United States? **Figure 5** is a bar plot of all of the destination states in descending order. This plot is extremely hard to read since there is a total of 51 states. Therefore, in order to better answer the question at hand, the bar plot was subsetted to just the top 10 destination states.[11] This is pictured in **Figure 6**

The top 10 destination states, respectively, are California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, Michigan, New Jersey, and Georgia. Since there were no prior expectations of which states people would migrate to, this list is not necessarily unsurprising. Unfortunately, the data does not allow an exploration of state characteristics to see whether these states vary on any important factors (economic, social, or political).

The last question this paper aims to answer is where do people migrate from when they do migrate. Put another way, what are the origin states of migrants within the United States? **Figure 7** presents all of the origin states in ascending order, and **Figure 8** presents just the top 10 origin states. The original origin state bar plot was subsetted for the same reason as the original

[11]Looking at the top 10 was selected arbitrarily. Future iterations could look at the top 15, 20, etc.

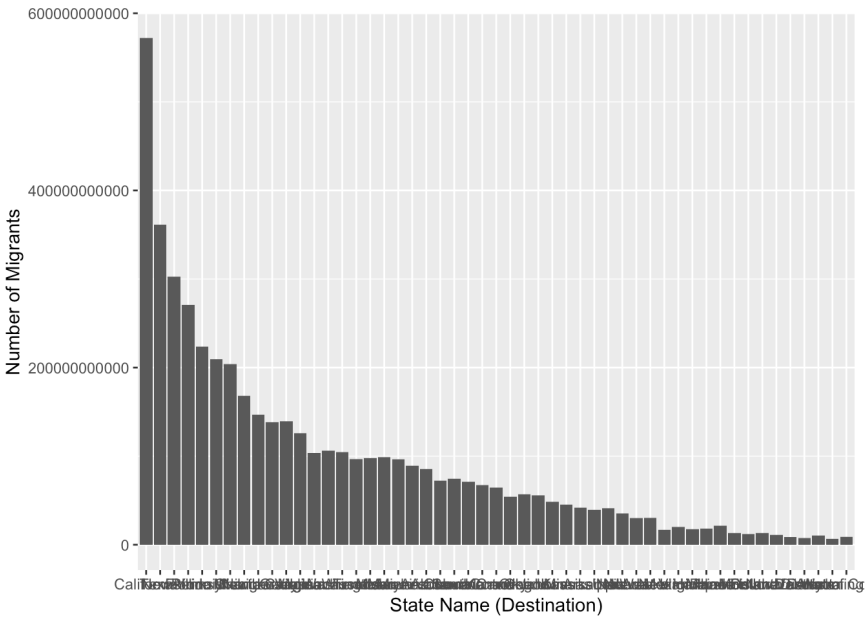

Figure 5: Number of Migrants by Destination State

destination state bar plot - including all states made the figure impossible to read.

Figure 8 shows that the top 10 origin states are California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, Michigan, New Jersey, Georgia. One may notice that is the same top 10 list as the top 10 destination states! An obvious explanation is that these are all large states and a more appropriate answer may be to look at a migration rate measure ($\frac{TotalNumberofMigrants}{TotalStatePopulation}$). This is a potential avenue of future research.

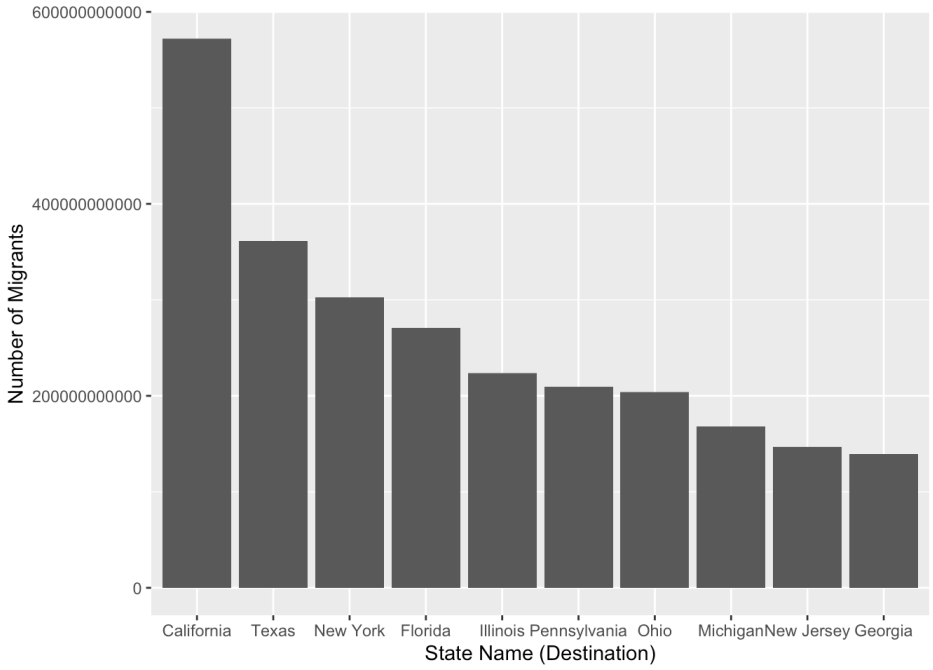

Figure 6: Number of Migrants by Destination State (Top 10)

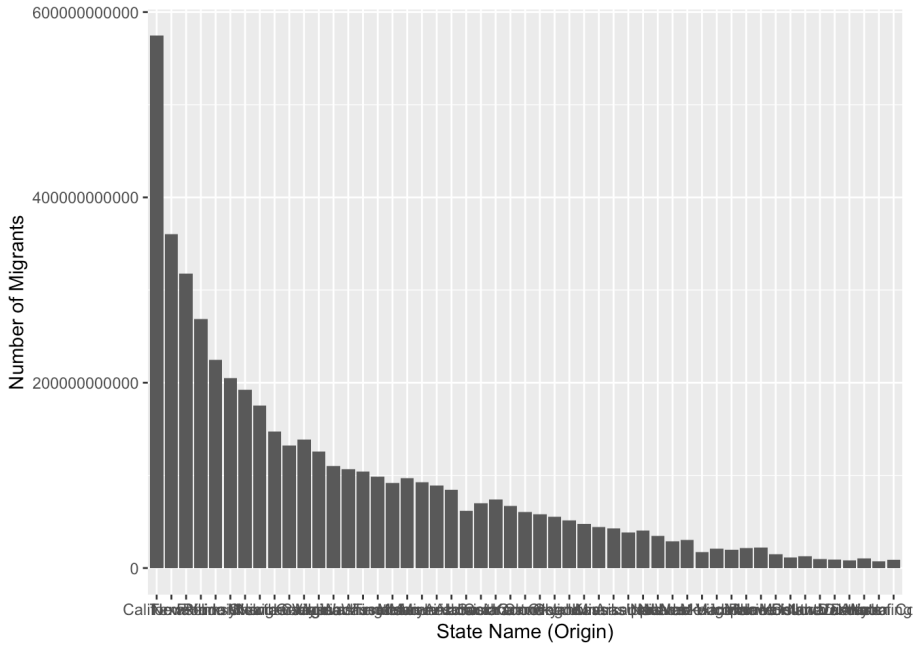

Figure 7: Number of Migrants by Destination State

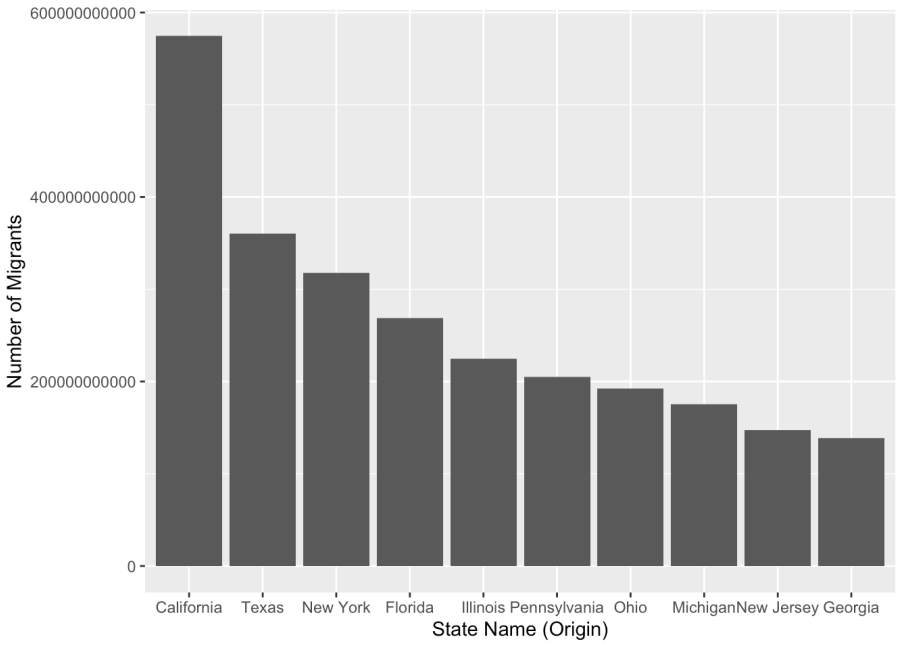

Figure 8: Number of Migrants by Destination State (Top 10)

4 Discussion

The preliminary findings presented here imply that there has been a consistent increase in internal migrants over time. Additionally, the top destination *and* origin states are California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, Michigan, New Jersey, Georgia. There are multiple directions for future research related to the questions presented in this paper.

First, the most glaring spot for future research is to tease out why the most common destination states are the same as the most common origin states. My first expectation is that the results would change when the total state population is accounted for. However, it may be possible that other factors are driving this findings. For example, it may be possible that flight prices between these states are the cheapest. Whatever the underlying cause, future work should collect more data on state characteristics (such as total population, economic growth, distance between states, etc.) to better understand why this finding occurs.

Second, future work should explore *why* internal migration has increased over time. As previously noted, there may have been economic, social, political, or other changes that have occurred that has lead to there being a gradual increase. It may be possible that internal migration has steadily increased over time simply due to population growth, but this should be further explored and empirically shown (if true!) in future work.

Third, future researchers could improve upon the presented study by engaging with more rigorous empirical designs. The two previous future steps are both areas that would lend themselves to better empirical tests. Specifically,

questions that aim to understand the determinants of U.S. internal migration should be tested with other empirical methods, such as using regressions. This would greatly improve on the study presented here as scholars could then make much stronger claims about internal U.S. migration than are presented in this paper.

Work on U.S. internal migration has a lot of room to grow. Understanding these migration patterns will allow scholars to better understand migration generally, as well as given American officials a better understanding of current and future migration in order to better prepare for future situations. Internal migration is not just an issue found outside the United States and understanding how and why it occurs in the United States will given us a better look into American life and politics.

5 References

Castleman, Terry. "Droves of Californians Are Moving to Texas. Here's the Life They Are Finding." Greater Milwaukee Today, November 13, 2023. https://www.gmtoday.com/news/national/droves-of-californians-are-moving-to-texas-here-s-the-life-they-are-finding/article_c3f9961a-8238-11ee-bba3-f7e71ae62c4b.html#:~:text=Here's%20the%20life%20they%20are%20finding,-Terry%20Castleman%20%2D%20Los&text=More%20than%20100%2C000%20Californians%20moved,who%20made%20the%20opposite%20move.&text=Jayne%20Jordan%2C%2061%2C%20was%20a,and%20moved%20to%20Azle%2C%20Texas.

DeWaard, Jack, Katherine J. Curtis, and Elizabeth Fussell. "Population recovery in New Orleans after Hurricane Katrina: Exploring the potential role of stage migration in migration systems." *Population and environment* 37 (2016): 449-463.

Durand, Jorge, and Douglas S. Massey. "Mexican migration to the United States: A critical review." *Latin American Research Review* 27, no. 2 (1992): 3-42.

Fearon, James D., and Andrew Shaver. "Civil war violence and refugee outflows." *Unpublished paper. Stanford University* (2020).

Federal Communications Commission. "Federal Information Processing System (FIPS) Codes for States and Counties." States and Counties FIPS Codes. Accessed December 12, 2023. https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt.

Greenwood, Michael J. "Research on internal migration in the United States: A survey." *Journal of economic literature* (1975): 397-433.

Gross, Emily. "Internal revenue service area-to-area migration data: strengths, limitations, and current uses." *Statistics of Income. SOI Bulletin* 25, no. 3 (2005): 159-160.

Hauer, Mathew, and James Byars. "IRS county-to-county migration data, 1990–2010." *Demographic Research* 40 (2019): 1153-1166.

Mackun, Paul. "Population Distribution and Change: 2000 to 2010." Census.gov, December 16, 2021. https://www.census.gov/newsroom/blogs/random-samplings/2011/03/population-distribution-and-change-2000-to-2010.html#:~:text=The%202010%20Census%20reported%20308.7,the%201980s%20of%209.8%20percent.

Massey, Douglas S. "Understanding mexican migration to the United States." *American Journal of Sociology* 92, no. 6 (1987): 1372-1403.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. "Internal migration in the United States." *Journal of Economic perspectives* 25, no. 3 (2011): 173-196.

Shultz, James M., Dana Rose Garfin, Zelde Espinel, Ricardo Araya, Maria A. Oquendo, Milton L. Wainberg, Roberto Chaskel et al. "Internally displaced "victims of armed conflict" in Colombia: the trajectory and trauma signature of forced migration." *Current psychiatry reports* 16 (2014): 1-16.

Weiss, Thomas G., and David A. Korn. *Internal displacement: Conceptualization and its consequences.* Routledge, 2006.