Brendan Bebb

STATS 485

Professor Ben Hansen

3 February 2023

# Estimating the Probability of Economic Mobility and Opportunity

## I. Introduction

For many individuals in their late teen years, the biggest imposing question that surrounds their futures is how they are going to monetarily succeed in life. In this work, data collected from a random sample of citizens from various cities across the United States is used to estimate the probability of a 16-18 year old individual who moves from the lowest quintile of economic status to the highest quintile by their thirties. The true probabilities of moving up in economic status for each city are captured in 95% confidence intervals known as the Jeffreys Interval, which is similar to the standard Wald interval, but works to optimize overall coverage as an equal-tailed interval. Findings include the optimal region and nation to live in to have the best economic mobility, as well as the role of city population size in providing this opportunity.

## II. Methodology

The data used in this project samples citizens of 40 different commuting zones across the United States. Each commuting zone is also labeled with which region of the United States it resides in: northeast, west, south, or midwest. The sampling size was determined by taking a fixed proportion of the overall population of each commuting zone, meaning fewer from smaller-populated zones and more from larger-populated zones. All of the sampling was

collected uniformly at random to account for various confounding variables and to eliminate any weighting that might be necessary to make note of for any sample across various regions.

Commuting zones were chosen as the level of measurement for this because of the availability of the data. Each commuting zone tracks the economic mobility of its citizens. Commuting zones are not reflective of state population mobility, as opportunities are different in various commuting zones in each state. The same applies for the nation as a whole: commuting zones merely serve as an *estimate* of the probability of economic mobility, which is why one must factor in population size of each commuting zone when analyzing the data gathered.

Individuals from each commuting zone were assessed based on their economic mobility. Individuals were labeled as "upmovers" if they met the following criteria: (1) at the time of sampling they were in their early thirties, (2) they are currently in the highest quintile of economic status, and (3) they were in the lowest quintile of economic status when aged 16-18. All individuals who were in the lowest quintile, regardless of economic mobility, were also tracked to find the proportion of those individuals who "upmoved." Given that the sampling of individuals was performed uniformly at random, some assumptions made are that individuals from the same commuting zone all have the same probability of economic mobility, regardless of what opportunities are available to each citizen. In a similar fashion, since this study involves binomial proportions, it is assumed that each individual moves up independently of each other, meaning that Individual A moving up does not affect Individual B's chances of moving up. This proportion is then used to calculate the aforementioned Jeffreys' Intervals for each of the 40 commuting zones in an effort to capture the true population proportion of upmovers.

The Jeffreys' Interval was chosen in this work due to the greater coverage probability for small proportion values. The Wald interval struggles to capture the true population proportion at

a confidence level of 95% when given sample proportions very close to 0. As found by Brown, Cai, and Disgupta (2001), the standard interval has "major problems arise [in terms of] coverage probability when *p* is near the boundaries." The Jeffreys' Interval serves as a strong alternative to the Wald interval because of its equal-tailed distribution, which has the upper and lower confidence limits "modified to avoid the undesirable result" that the coverage probability approaches 0 around the proportion boundaries of 0 or 1 (Brown et al., 2001). In addition, Brown, Cai, and Disgupta (2001) outline the shorter length that accompanies the Jeffreys' Interval, serving as a strong estimator to capture the true population proportion of upmovers in a more precise and accurate interval.

For each commuting zone, a 95% Jeffreys' Interval was calculated and added into the dataset. The results were then added together for each respective region to find the true population proportion of upmovers for each respective region. Upon finding variability amongst regions, the proportions were then calculated for nations described in Colin Woodard's *New York Times* article (2018). Woodard, too, believed that regions were too large and generic to categorize the United States, so he provides in his article "nations" that the commuting zones could belong to (Woodard). As seen in the appendix, each commuting zone also has their correlating nation assigned to it, helping one better understand locationally where each nation resides in the United States. Based on the variability amongst these nations, another proposed variable to capture the best area to live in for economic mobility was to calculate the probability of being an upmover based on the population size of a city.

## III. Results

After creating the 95% Jeffreys' Intervals for each of the 40 commuting zones, it was discovered that Phillipsburg, Kansas was the best commuting zone to improve one's economic

probability with an interval of (0.218, 0.270). On the contrary, Orlando, Florida provided the worst opportunity to become an upmover with an interval of (0, 0.0089). A complete set of each commuting zone and its interval can be located in the appendix that accompanies this paper. The commuting zones' population sizes, number of lower-quintile-starting individuals, and upmover proportions were then added together for each region to provide an estimate of the population proportion for each respective region. The results obtained were as follows:

| **Region** | **Estimated Proportion of Upmovers** |
|---|---|
| northeast | 0.081 |
| west | 0.086 |
| midwest | 0.053 |
| south | 0.037 |

The west region has the highest probability of being an upmover, but the northeast is comparable in probability as well. The variability amongst these regions is enough evidence to investigate the proportions of these commuting zones grouped under their respective nations as well. When the zones are added together for their nations, the results are as illustrated in the table below.

| **Nation** | **Estimated Proportion of Upmovers** |
|---|---|
| El Norte | 0.101 |
| Left Coast | 0.088 |
| New Netherlands | 0.076 |
| Yankeedom | 0.062 |
| Far West | 0.061 |
| New France | 0.056 |

| | |
|---|---|
| Midland | 0.039 |
| Deep South | 0.029 |
| Greater Appalachia | 0.025 |

When broken up by nation, El Norte has the highest estimated probability of drastically improving one's economic status. This is not overly surprising, as two of the four commuting zones within the El Norte nation are also within the west region. This solidifies the idea that more western commuting zones provide a higher probability of social upmoving.

Within the El Norte region, it is worth noting that three of the four commuting zones have populations greater than 200,000. Additionally, as mentioned above, Phillipsburg, Kansas has the highest probability of becoming an upmover, but this zone only has a population size of 974. This raises the question if there is any sort of advantage to economic mobility prevalent by grouping commuting zones based on population. The commuting zones were then divided amongst five different levels of population size, which are outlined in the appendix.

The results find that the highest probability of social upmovement lie within cities with a population of 150,001 to 200,000 citizens, with an estimated population probability of 0.092, whereas cities with a population of 100,001 to 150,000 have the lowest population probability of 0.025. The remaining three brackets of population size are all comparable at approximately 0.075.

## IV. Discussion

Based on the results found above, in simplest terms, it appears that the most ideal commuting zone to live in would be one in the west region, El Norte nation with a population size of 100,001 to 150,000 citizens for the best improvement of economic status. However, a

commuting zone with these criteria did not appear in the 40 zones used in compiling this data. Of the three grouping tactics used, it appears that the region is the most beneficial way to estimate the population proportion, as there are an equal number of 10 commuting zones for each region. Dividing these 40 commuting zones based on population size or nation provide large variability, but that could also be due to an uneven distribution of observations amongst those groupings.

It is not sufficient to look simply at the distribution based on region, however, because of the vast amount of space that each region covers. As stated above, to account for the different amount of commuting zones in each region, a weighting technique must be implemented. This is accomplished by adding together the population size and number of low-starters for each zone in the regions to obtain a proportion for the region as a whole. While each region has a vast distribution of population among commuting zones, one cannot account for regional differences in opportunities, such as average income and overall cost of living. Between the other two measurements used above, there appears to be more of a correlation between the commuting zone's nation than population size, but more data points are necessary to make this assertion.

Overall, it does not appear that there is any benefit to the population size of a commuting zone. As previously mentioned, Phillipsburg, Kansas has the highest interval of upmover probability, but there are several factors that have influenced that. Phillipsburg only has a population of 934 citizens and it sampled 200 people randomly; Detroit, Michigan also had 200 people sampled randomly, despite having a population size of 392,121.

If one chooses to overlook that, another reason that there is a higher social mobility in Phillipsburg could be the growing number of ethanol plants that inhibit the small zone, creating more job opportunities for its citizens (Selfa et al, 2015). Phoenix, Arizona was another commuting zone in the El Norte nation as well as the west region, which is feasible, as Phoenix

is one of the most populous cities out west and with the large size of the commuting zone comes more job opportunities. There are more factors to consider when selecting the optimal commuting zone for social mobility, but these divisions studied provide a strong baseline to estimate ideal commuting zones and the opportunities that accompany them.

In its entirety, the Jeffreys' Interval provided strong estimates for each of the commuting zones. One area that the Jeffreys' Interval struggles in, however, is that the numerical endpoints of very small $n$ cannot be computed without mathematical software (Brown et al., 2001). This is not something that needs to be taken into account in this study, as the smallest $n$ was 200, making the Jeffrey's Interval a very viable option to carry out calculations involving economic mobility of individuals in various commuting zones. As illustrated in the graphics made by Brown, Cai, and Disgupta (2001), the Jeffreys' Interval performs much better around the $p$ boundaries of 0 and 1 than the standard interval, but it still struggles a bit in its accuracy in capturing the true population proportion at the 95% confidence level. The Jeffreys' Interval was chosen for this report over the Wilson Interval due to its equal-tailed nature, which increases the coverage probability around the boundaries.

Works Cited

Brown, Lawrence D., et al. “Interval Estimation for a Binomial Proportion.” *Statistical Science*, vol. 16, no. 2, 2001, https://doi.org/10.1214/ss/1009213286.

Selfa, Theresa, et al. “Promoting Ethanol in Rural Kansas: Local Framings and Cultural Politics.” *Journal of Rural Studies*, vol. 39, June 2015, pp. 63–73, https://doi.org/10.1016/j.jrurstud.2015.03.008.

Woodard, C. (2018, July 30). The Maps That Show That City vs. Country Is Not Our Political Fault Line. *The New York Times*.