Research Paper: Factors of State-to-State Migration

9 December 2024

Evan Wall, James Kull

(i) Introduction

What factors influence state-to-state migration? According to our results, for every 1% difference between two states' average GDP growth over the last five years, 19.2% more people will migrate from state A to state B in a given year. The number of migrants into a state has multiple determinants, so politicians are often able to disguise insignificant factors as significant, and vice versa. This research paper is a deep dive into several factors believed to be significant in people's decision to move their residence. The results of this paper are important to both politicians and citizens. For politicians, they will gain insight as to whether the policies they support would cause an inflow or outflow of migration. In states seeking an influx of migrants, for example, politicians would be interested to learn how many migrants they may receive after a tax cut. In 2023, West Virginia passed an income and property tax reduction that was, in large part, meant to attract interstate migrants. Empirical analysis is needed in order to determine the efficacy of such programs. As for citizens, this study gives them a better idea as to whether a migration phenomenon is a result of what the people on TV are saying, or something else. The decision to migrate is certainly driven by a plethora of factors, but honing in on the significance of a few will make the overall picture less blurry.

Our research in this area revolves around two central hypotheses. First, states that have higher GDP growth are more likely to attract a greater number of migrants. GDP growth should be indicative of economic opportunity, and individuals will flock toward states that will give them a better chance to be successful in this regard. Second, people are likely to migrate to states that have lower rates of taxation than the one in which they currently reside. All else being equal, people want to live somewhere where they bring home more of their income.

(ii) Theory / Conceptual Framework

Our theory underlying our research is that migration, hereafter referred to as M, is a function of the difference between the net GDP growth over the past five years, hereafter N5Y, and the net tax burden per capita, hereafter NT, both with net defined as destination state level less origin state level. Since these data are organized into destination and origin state pairs in a given year, the subscripts S and T denote a state pair and year respectively. Such, the function defining our theoretical framework, excluding controls, can be succinctly given as:

$$M_{ST} = f(N5Y_{ST}, NT_{ST})$$

The underlying theory of the relationship between M and N5Y is that higher N5Y is indicative of State B (destination) having a level of economic prosperity that will entice more residents of State A (origin) to pack their bags and move to State B. There are two main ways in which a higher N5Y would be seen as enticing for a potential migrant, thus yielding a positive relationship between N5Y and M. The first reason is that an increase in N5Y means that there is at least one industry in the economy that is expanding. In the modern specialized economy, such an industrial expansion could entice workers in the expanding field who are in other states to make a move. Second, a relatively high N5Y could also yield better living conditions for the residents of State B. As a state is increasing its GDP and living conditions are improving, people are happy. Thus, they are likely to tell their friends and family how good it is to be living in State B. Furthermore, Molloy et al (2011) and Sasser (2010) found that the basic economic theory of individuals being utility maximizers holds when observing migration patterns; more simply put, people are going to live where they believe they will be most happy.

The theorized relationship between NT and M is negative, meaning that an increase in NT will yield a decrease in M. All else being equal, people would rather take home more of their paycheck. As the government tax revenue per capita is increasing, money is usually being taxed out of the paychecks of the state's citizens directly or indirectly. A state with a government that is bringing in more tax revenue per capita, and thus taxing its constituents more, is a less appealing place for people to live. Financial factors are theorized to be a significant driver of the migratory decision, and as such, higher taxes are something people give significant thought toward.

As a part of our analysis, we also controlled for a number of variables to minimize the effect of omitted variable bias on our result. The first two controls that are necessary in this analysis are controlling for our cross-sectional units, which in this case are states. A state can have any given number of traits that make it more or less desirable to potential migrants. Many of these traits are more or less constant over our observation period of ten years, and many of these traits, such as natural beauty or perception of residents, would be almost impossible to quantify. Both of these qualities make the state fixed effects control critical to our analysis. There are also many things that would cause more or less migrants in any given year, and controlling for that is necessary for our regression. A country-wide cold front, or other inclement weather, may disincentivize migration across the country in a given year.

Another control that we chose to include is Border, which is a binary variable equal to one if State A and State B touch and equal to zero otherwise. The theory behind this binary variable is that we want to control for the individuals who have a relatively frictionless moving process and are moving very short distances, since their moving processes probably require less consideration. For example, people moving from one suburb of Kansas City to another is technically interstate migration. In a similar vein, we also control for the distance between the states using a variable called CapitalsDist, which is the distance, in miles, between the capitals of State A and State B. We chose to use the distance between the capitals because capital cities in states tend to be, on average, centrally located as well as population centers that a good portion of migrants would be moving to. We also controlled for the net density of the two states, so that migration that is attributable to urbanization does not hinder our results. States that are more dense are also likely to be involved in growing white-collar industries, and we want to make sure that the effect of people moving to live in cities does not inflate our result of the number of people moving because of GDP growth. Lastly, we controlled for the net percentage of the population that is white, attempting to soak up the effect of any race-based migration that may be occurring during our observation window.

(iii) Literature Review

United States citizens migrate to maximize their utility is the theory that underlies the literature on interstate migration. The works of Molloy et al and Sasser are foundations to this theory. Molloy's work built a basic framework on the determinants of migration, measured at the county, state, and regional level. These determinants included demographic data, cost of living, wages, along with market conditions; Sasser's followed a similar set of parameters. However, both papers found two separate key findings that point in a new direction of research. Molloy's work found a steady decline in the overall proportion of citizens who are migrating in the United States (proportionally the same at the county, state, and regional levels). Their results point to an increasing frictional cost of moving as the primary cause. Sasser found a more seminal discovery that state market conditions have a high and strong correlation to the proportion of individuals moving out of a state.

While the previously stated findings help build the framework on our model, current work by economists this calendar year have driven more directly at research surrounding our research question. Cassidy et al published findings that concluded the introduction of income tax has a negative correlation with state populations (specifically within the context of migration. Cassidy found that an introduction of an income tax found a small but statistically significant decrease in state population and an increase in migration out of the state. They also found that this small correlation has been growing larger over time. Possibly the most important implication when it comes to our regression model is a posed endogeneity problem. Cassidy found that many states in the mid-1900's introduced an income tax to compensate for a smaller number of taxed citizens.

Our regression model aims to get at a couple items that are novel to this area of research. First, our regression aims to decipher marginal differences in overall tax rate, not just the introduction of a certain tax. Secondly, we are diving deeper by using state to state data points (ie. specifically measuring migration to Alabama to Alaska and Alabama, Alabama to Arizona, etc.). We believe our model should more strongly predict how marginal changes in effective tax rate impact state to state migration compared to previously published papers. This will add to the existing body of work and perhaps drive granular research into other marginal migration factors.

(iv) Data and Descriptive Analysis

Our data are sourced from various US government agencies, which we believe to be high-quality. These agencies include the Bureau of the Census, Bureau of Economic Analysis, Federal Housing Finance Agency, and the Bureau of Labor Statistics. All data are from the years 2010-2019 and are specific to the destination and origin state of migrants. We believe that our panel data will improve the quality of our estimation strategy over a purely time series or cross-sectional data set.

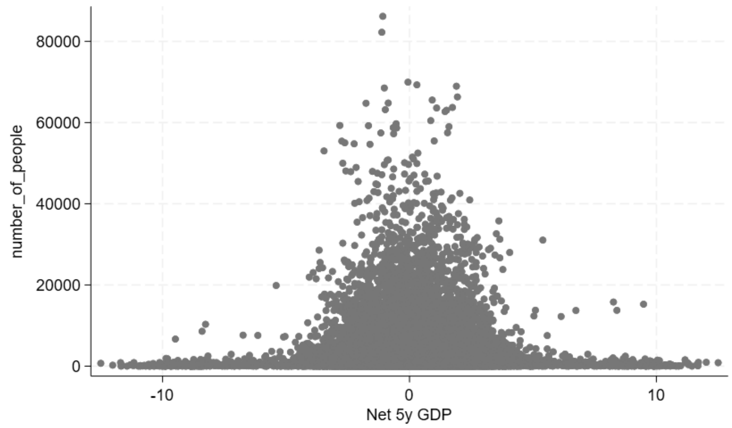

Fig. 1: Distribution of Migrants Dependent on GDP Growth

Our endogenous variable is the number of people who migrated from their origin state to their destination state in a given year. Number of people has a mean of 2,928, standard deviation of 5,715, a minimum of 0, and a maximum of 86,164. The three largest observations of interstate migrations are all from California to Texas in the years 2016, 2018, and 2019. This state pairing has seen a massive trend in migration, with only one year, 2011, falling below 60,000 migrants from California to Texas. The states that sent the least amount of total migrants from 2010-2019 were Vermont, North Dakota, and Wyoming. These data were sourced from the US Bureau of the Census in the Migration Flows data set. Something of note for the purpose of our regression is that we had to replace the cases of 0 migration with 1 to run our left log regression.

Our primary exogenous variable is net 5 year GDP, which is calculated by taking the 5 year GDP growth average of the destination state and subtracting the 5 year GDP growth average of the origin state, measured as a percentage. Since this variable is calculated in this way, the mean value is 0 for both 5 Year GDP growth average and all of our other net variables, helping us interpret spread. The maximum and minimum values are 12.5% and -12.5%, which was the difference between Nevada, -1.88%, and North Dakota, 10.62%, in 2012. As seen in Fig. 1 above, the majority of the distribution is found within the range of (-5,5), although there are enough data points outside of this range, 562, that we are confident that our data have enough spread to not hinder our regression results. For one of our secondary regressions, we use Net GDP growth which is only the difference in GDP growth from the previous year, rather than the 5 year average. This variable has a max of 25.4%, meaning that the maximum difference in GDP growth over a single year between two states was 25.4%. Wyoming, -2.2%, and North Dakota, 23.2%, had this massive differential in 2012. All GDP growth data were gathered from the State Annual Summary Statistics by the Bureau of Economic Analysis.

Net Tax per Capita is another exogenous variable which is calculated by taking the difference between the per capita tax burden of the destination state and subtracting it by the per capita tax burden of the origin state. The maximum value of this statistic is 8.09, meaning that the highest tax burden per capita difference was $8,090. This took place between Alaska, $9,770, and New Hampshire, $1,680, in 2012. The standard deviation of Net tax per capita was $1,460, meaning 95% of the differences between states were $2,862 or less per capita. These data were collected from the Annual Survey of State Government Tax Collections by the US Bureau of the Census.

Net Housing Pricing Index (HPI) is an exogenous variable which is calculated by taking the difference between the housing price index of the destination state and subtracting it by the housing price index of the origin state. The HPI of all states was normalized to 100 in 2000. The maximum value for the net HPI was 130, between Michigan, 125, and Hawaii, 255, in 2019. The standard deviation Net HPI was 31.8, meaning 95% of our observations obtained a difference in HPI of 62.3 or less. These data were collected from the Demographic and Housing Estimates data set from the US Bureau of the Census.

Net Unemployment is an exogenous variable which is calculated by taking the difference between the unemployment rate of the destination state and the origin state. The maximum value is 10.1 meaning that the highest difference in unemployment rate between two states was 10.1%. All three of the highest differentials included Nevada as it disproportionately struggled to recover from the Great Financial Crisis of 2008-2009. The standard deviation net unemployment was just under 2% (1.97%) meaning the vast majority of our observations observed a difference of 4% or less between the destination and origin. These data were collected from the Unemployment Rates for States by the US Bureau of Labor Statistics.

Net Density is an exogenous variable which is calculated by taking the difference between the population density of the destination and the origin state. We observed an average difference in density of 292 people per square mile, while also observing a maximum of 1017 people per square mile (and an equal and opposite minimum). These data were collected from the Demographic and Housing Estimates by the US Bureau of the Census.

The border exogenous variable is a qualitative variable that is determined by whether or not the destination and origin state share a border. Our observations shared a border 8.7% of the time.

Capital distance is a variable that measures the distance between the capital of the destination and origin state in miles, calculated using the Google Maps coordinates of the capitals. Our observations had a mean value of 890 miles and a maximum value of 5,118 miles. The standard deviation of these observations was 40.6 miles.

Net White is an exogenous variable that measures the difference between the destination state and origin state's percentage of the population that is white. These data were obtained from the US Census Bureau's ACS Demographic and Housing Estimates data set.

(v) Estimation Strategy

We use an ordinary least squares regression to estimate our theorized relationships. We ran this regression on all 24,500 combinations of destination & origin states across the period 2010-2019. We chose to not include data from 2020 onwards, in order to prevent potential distortions due to the Covid-19 pandemic. In order to see a variety of results, we run 6 regressions. Our baseline regression includes ln(M) (as a function of N5Y and NT, with the control variables of Border, CapDist, NetDensity, NetHPI, PctWhite, NetUnemploymentRate and state and year fixed effects. We chose to make our baseline a semi-log because we believe that people are more likely to respond to changes in economic conditions in proportional terms, where absolute changes in statistics result in percent increases in the number of migrants . Additionally, percentage changes in migration are more easily interpretable by our intended audience, policy makers and the general public, and they generalize better to states with differing populations. We also ran several regressions in which we included the population of the sending state as a control variable, but ultimately omitted it from our baseline regression due to a very high VIF. We also believe that most of the effect of population is picked up in fixed effects. We also ran a regression without controls, but the results had a weaker R^2 value. Additionally, we ran a regression with the net GDP growth of the previous year instead of the five year trailing growth rate and found the results to be less clinically significant, although they were still statistically significant. Another finding of interest while running regressions was that the sign on NetUnemploymentRate in the regressions in which it was included was the opposite of expected. This is likely due to it having a very high correlation with GDP growth. By holding N5Y constant, a NetUnemploymentRate is essentially a measure of job availability in an economy that is performing as well as any other economy. Thus, the expected sign would be positive.

(vi) Conclusion

Ultimately, our estimation strategy allowed us to generate clinically and statistically significant evidence of the substantial effects of tax rate and GDP growth on interstate migration patterns. In our regression of choice, as well as every other one in which we ran state and year fixed effects, both Net5yGDP and TaxRate were found to have coefficients statistically significant at the 0.01 level. For more detailed statistical results, please see appendix one. Our large number of observations led to smaller standard errors for our coefficients, thus contributing to the statistical significance of our results.

Additionally, for both key variables, we determined clinical significance. Specifically, we predict that a one standard deviation change in the net five year average GDP growth rate (destination-origin) will yield a 0.24 standard deviation change in the number of migrants into a given destination state in a given year. For tax rates, we predict that a one standard deviation change in the average tax rate per capita will yield a 0.13 standard deviation change in the number of migrants in a given period. Migration is such a multifaceted decision that these results are extraordinarily clinically significant, and they demonstrate that both GDP growth and tax rate play a significant role in the decision-making of potential migrants.

The results from this analysis should be applicable for both policymakers and potential migrants. For migrants, it is helpful to know the perhaps subconscious factors that could be influencing one's decision to move to a certain state. Perhaps the more compelling application for this research is in how it could help policymakers. For example, if a state has started to experience higher GDP growth in the last few years, this research tells the state's officials that they should begin preparing for an influx of migrants from other states. Additionally, the research on the migratory results of the tax burden are another factor for officials to consider when mulling over a potential change in the tax code.

While we are satisfied with the results of our research within the given time frame, there is more exploration that could be done. Some areas of particular interest would be adding more exogenous variables that vary across states across time, as we attempt to gain a clearer picture of the true factors of migration. One that could be particularly interesting to examine is crime rate, since that is purported to be a major factor in the migratory decision, and finding data for that analysis would be relatively straightforward.

(vii) References

Cassidy, Traviss, Mark Dincecco, and Ugo Antonio Troiano. 2024. "The Introduction of the Income Tax, Fiscal Capacity, and Migration: Evidence from US States." *American Economic Journal: Economic Policy*, 16 (1): 359–93.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives*, 25 (3): 173–96.

Sasser, Alicia C. "Voting with Their Feet: Relative Economic Conditions and State Migration

Patterns." *Regional Science and Urban Economics*, vol. 40, no. 2, 2010, pp. 122–35,

https://doi.org/10.1016/j.regsciurbeco.2010.02.001.

Appendix i: Regression Results

Net Migration Regression

VARIABLES	(1) Number of People	(2) Number of People	(3) ln(Number of People)
Net5yGDP	348.09089***	454.65087***	0.19216***
	(60.371)	(77.772)	(0.018)
NetTaxPerCapita	-676.40292***	-464.18144***	-0.19888***
	(117.680)	(75.435)	(0.010)
NetHPI		12.52165***	-0.00155
		(4.101)	(0.001)
CapitalsDist		-1.55236***	-0.00061***
		(0.213)	(0.000)
Border		7,341.52518***	1.65591***
		(713.709)	(0.088)
NetUnemploymentRate		369.12921***	0.13068***
		(59.159)	(0.011)
NetDensity		0.62198	0.00006
		(0.658)	(0.000)
NetWhite		-9,167.56801***	-3.48982***
		(1,157.938)	(0.127)
Constant	163.18267	11,459.17669***	9.76259***
	(270.984)	(1,412.457)	(0.154)
Observations	24,500	24,500	22,767
R-squared	0.212	0.475	0.576
State Effects	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: We also regressed GDP growth rate (instead of 5-year trailing average). However, we believe that a multiyear trend is the better way to observe the actual effect of GDP growth on migratory behavior. We also ran regressions with a population of origin variable. However, we decided that most of the effect of this control was already being controlled for in the state effects. If you would like the full regression results, please contact one of the authors of this paper.

Appendix ii: Descriptions of Variables Used in Regressions

Net Migration	Endogenous variable; migration from origin to destination in year x, measured in people
Net5yGDP	Exogenous variable; GDP growth, averaged over the previous five years, measured as percent, (destination-origin)
NetTaxPerCapita	Exogenous variable; Nominal thousands of dollars, (destination-origin)
NetHPI	Control variable; Housing Price Index, base year 2000
CapitalsDist	Control variable; Distance from origin to destination state capital
Border	Control variable; does the origin state share a border with the destination state, dummy variable, (yes = 1)
NetUnemploymentRate	Control variable; unemployment rate, measured as a percent, (destination-origin)
NetDensity	Control variable; population (people)/state area (square-miles), (destination-origin)
NetWhite	Control variable; percent of state population that identifies as white, (destination-origin)