Master’s Capstone Draft #2

Tyler Frankenberg

DATA 698

Dr. Arthur O’Connor

April 5, 2023

City University of New York

Introduction

A mobile workforce, able and prone to migrate between regions to seek employment opportunities, is not just a characteristic of the American economy – it is an enduring element of the American cultural experience. From the mythology of the 19th century frontier, to *The Grapes of Wrath*, *A Raisin in the Sun, The Beverly Hilbillies,* and HBO’s *Girls,* our shared identity has been shaped by those who’ve left one American place to pursue their American Dream in another.

And yet, byThe patterns that have long informed our collective understanding both of ‘who lives where’ and the freedom with which we might move about the country have been changing dramatically over the course of several generations.

Central to the discourse on this change has been the changing place of cities. For most of our nation’s history, cities have served as magnets for the upwardly mobile, attracting migrants from rural areas to leave behind pastoral agrarian life and join the booming industrial economy on which the prosperity of the post-WWII period was built. Yet the emergence of new technologies for industrial automation and the connection of our urban centers to a wider global economy changed the dynamics of who moved to cities, who left them, and which cities and regions prospered. As Florida (2019), Bishop (2009), and Moretti (2012) wrote in the first wave of reckoning with the new dynamics of 21st century human geography, the transition of our economy from industrial to service oriented in the last two decades of the 20th century had made cities a different kind of magnet – one that attracted the highly educated based on the concentrated force of their technological innovation. This dynamic supercharged the economic competitiveness of highly educated, innovative urban centers, but also their cost of living, which may have further suppressed the overall rate of internal migration.

With this background in mind, a reasonable question we may consider for our own time is – does it still matter where we live in order to do our work? On the ‘supply side’, employers from established multinational firms to scrappy startups may consider – where is the best location for us to base our operations? As our economy and society continue their long, slow evolution from agricultural and industrial to technological and service-oriented, mayors and councils debate – how can we stay competitive in the 21st century economy? And as the pace of innovation in artificial intelligence technology accelerates – how will the jobs it creates or replaces affect the rate of migration between cities?

Literature Review

Surveying the economic literature on long-distance internal migration in the United States, Molloy et al. put forth three mechanisms by which the aggregate rate of migration may be demonstrated to change over time based on the characteristics of movers. First, individuals may naturally (such as by aging) or through life choices (such as through buying a home or taking a job in a different sector) become part of a demographic grouping more or less statistically likely to move. Second, particular groups’ tendency to relocate to new regions may change over time. Third, cyclical and/or structural economic factors (such as a recession, the emergence and decline of industries, or major changes to the tax code) may broadly influence internal migration rates (2011).

While inconclusive on the causes of the secular decline in internal migration since 1980, Molloy et al., (2011) profile the demographic categories among which the disparities in likelihood to relocate are greatest: in short, individuals with at least some college education are more likely to relocate across large distances than workers with less education; individuals younger than 45 are more likely than individuals over 45; and renters are more likely than homeowners.

The tendency of highly-educated, highly-skilled workers in particular to relocate to centers of technological innovation like Silicon Valley, Boston, and Seattle and become more concentrated there over this time period (Berry & Glaeser, 2005) is a particular focus of the literature. This population dynamic and its impact on America’s cities, economy, and civic life has been popularized especially in the work of and economist Enrico Moretti in *The New Geography of Jobs* (2012), whose work centers on the character of this structural shift in the labor market; of urbanist Richard Florida in *The Rise of the Creative Class* (2002, 2014, 2019), who focuses on strategies for successful revival of post-industrial cities; sociologist Bill Bishop in *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*  (2008), who examines the impact of geographic segregation on political polarization.

While each differs somewhat in the specific context and conclusions through which they examine this topic, they align on the centrality of technological innovation measured by patent activity as a key determinant of regional economic growth since 1980, and as a primary predictor of cities’ and regions’ prospects for growth and prosperity in the 21st century. The primary mechanism by which technological innovation and migration of the highly-skilled are linked in driving this growth are not merely the monetization of individual patented inventions, but so-called ‘spillovers’ in which the concentration of technological and industrial knowledge and research activity in a given firm, industry, or geographic location tends to increase productivity, and thus to have a self-accelerating effect ( Jaffe, 1986; Adams & Jaffe, 1996; Orlando, 2003).

In other words, the presence of successful tech- and research-driven firms attracts not just highly-skilled employees, but additional firms and startup activity that attract still more highly-skilled employees. Moreover, the benefits of concentrated tech-driven economic growth is not limited to the highly-skilled and highly-educated; for example, Moretti shows that the wage growth impact of a high concentration of educated workers on workers *without* college degrees is four- to five- times higher than the impact on workers *with* college degrees; with the implication being that the research spillover-driven economic growth in tech centers spurs non-tech economic activity in these centers as well (Moretti, 2012).

Mechanisms aside from research spillovers and the acceleration of tech-based economic growth are less unanimously accepted as causes driving the migration of the highly-skilled. The premise of Florida’s work is that twenty-first century cities thrive largely on their success attracting what he calls the ‘Creative Class’, i.e. white collar workers in knowledge-based fields like tech and design, to settle there (Florida 2002, 2014, 2019). He rates cities on their potential to attract these so-called ‘Creatives’ using an index in which technological innovation measured by patent activity is combined with cultural elements like the presence of a large homosexual population (as a proxy for overall cultural diversity and tolerance) as well as measures of workforce education levels (pp. 228 – 264). His methodology is representative of a school of thought that sees the *non-economic reasons* for migration and settlement as a co-equal determinant of cities economic outcomes alongside economic causes; this is often characterized by the shorthand phrase “jobs follow people”, in contrast with a competing family of theories labeled “people follow jobs” (Russell, 2017).

A representative corpus of the “people follow jobs” approach is the work of economist Enrico Moretti, whose *The New Geography of Jobs* examines the migration of university-educated workers to centers of technological innovation (Moretti, 2012). While agreeing with Florida that the migration of highly-skilled individuals shapes cities’ economic outcomes (pp.154 – 177), he disagrees that thriving arts and entertainment scenes or the presence of vibrant artisanal niche economies are root causes attracting this migration (and he’s largely silent on the question of measures of diversity or cultural tolerance). He argues that the distinguishing cultural elements of successful twenty-first century cities are instead *effected* by the presence of highly-skilled workers; in particular their disposable income and educated tastes create the demand for the cultural and artisanal amenities that follow (pp. 188-190).

Why do I focus specifically on the impact of innovation in Artificial Intelligence separate of technological advances in other domains? First, patented Artificial Intelligence technologies have been demonstrated to increase firm’s hiring by 25% and revenue growth by 40% when compared with similar firms not using AI, while simultaneously increasing worker productivity and accelerating wage inequality between high- and low-skilled positions within the firm (Alderucci et al., 2019). Second, recent research on the broader category of software patents showing that such technological advances accounted for the bulk of differential spillover effects within regions since 1980 when compared with patents in other domains (in fact masking an overall stability in non-software patents across cities)(Chattergoon & Kerr, 2022). Both of these findings suggest that the impact of emerging AI technologies may be significant enough in its spillover impact to have its own measurable effects on the regions where its development is concentrated.

There is considerable value in understanding whether this translates to an effect on human settlement different from that of other technologies, not least because the effects of economic growth and migration driven by technological advance on the inhabitants of regions that attract and lose migrants are profound. In a landmark study, Chetty et al. demonstrated that the future potential opportunities of children born in innovative regions is demonstrably higher than those born in less technologically innovative regions in the United States (2014).

Moreover, the external effects of population change driven by technological innovation are neither uniformly distributed nor are they uniformly positive in their impact, as scholars have demonstrated in tying innovation and associated population growth to ?[lower housing affordability] (source), ?[greater inequality] (source), political polarization (Bishop, 2009) and even ?[the breakdown of traditional civic and social norms like club membership, volunteerism, and charitable giving] (?Putnam, 1999 as summarized in Bishop, 2009).

Thus, there are many reasons to assume emergent AI technologies’ impact on regional economic activity and migration might continue or accelerate well-established trends in play over the last 40 years; and yet, the fact remains this is largely still unexplored. Moreover, as the 2020 COVID-19 pandemic’s lockdowns and transitions to remote work spurred both a disruption in patterns of residential settlement away from urban centers as well as the development in new AI-enhanced technologies to address a changed workplace and society. In this paper, I do not attempt to address the pandemic-driven impacts, as the data for many of my sources becomes less available or less reliable since 2020. This does not change the fact that the changed context we now operate in demands a fresh examination of the situation over the last 10-20 years.

**Research Question & Hypothesis**

By measuring the effect of US counties’ artificial intelligence patent applications over the years 2011-2019 on their population growth through domestic in-migration over the same period, I endeavor to shed light on the extent to which the emerging Artificial Intelligence economy affected the dynamics of settlement in the United States in the period between the Great Recession and the global COVID-19 pandemic. I begin with the null hypothesis that the effect of AI patent activity was nonexistent or negligible (as compared with the measured effect of other economic and geographic influences over the same period) and test this hypothesis using several proxy variables in both ordinary least squares and multiple regression models. To control for confounding factors, I examine the relationship in the context of destination and origin counties’ relative personal tax burdens on citizens, labor market health, housing and rental market affordability, percentage of population with bachelor’s degrees, and presence of natural amenities.

Data and Variables

*Dependent Variable: Percent net population change through domestic migration, 2011-2019*

The analyses use as their dependent variable the county-by-county percent net population change from domestic migration, calculated from the Census Bureau’s annual population estimates for 2011-2019 (as published in the 2019 and 2022 vintages). These data are published annually by the Census Bureau to aid federal agency planning and allocations in the years in between the decennial census, and are updated with each ‘vintage’ to reflect the latest improvements in data collection and statistical methodology. I’m primarily concerned in this analysis with the county-level total population estimate and count of net domestic migrants, but the dataset also includes births, deaths, and count of net international migrants.

The percent net population change through domestic in-migration for a given county is calculated by taking the sum of **domestic migrants** over the years 2011-2019, and dividing by the starting (i.e. 2011) **population estimate:**

Text

Description automatically generated

For the 2416 US counties examined over this time period, this dependent variable ranges from a minimum of -0.1837 to a max at 0.6501. The histogram approximates a normal distribution centered at net-zero-percent in migration, with a noticeable rightward skew that appears most prominent beyond the interquartile range. The mean is very close to the median, suggesting it is not overly influenced by the rightward skew.

Chart, histogram

Description automatically generated Table

Description automatically generated with medium confidence

*Independent Variables: Measures of AI Patent Activity at the county level, 2011-2019*

Using US Patent and Trademark Office data on utility patent applications by county of residence of the first listed inventor, classified as pertaining to a variety of AI technologies using the algorithm developed by Giczy, Pairolero, & Toole (2021), I measure each county’s AI patent applications per thousand residents in 2011 as the county’s starting point for contribution to technological innovation in artificial intelligence. I also measure each county’s total number of Artificial Intelligence patent applications contributed over the study period as a measure of the local aggregate contribution to US innovation in AI over the period.

Map

Description automatically generated

Map

Description automatically generated

The total patents dataset is maintained by the United States Patent and Trademark Office’s PatentsView portal (and is also made conveniently accessible in full through Google BigQuery). For AI patents, I’ve joined the PatentsView dataset with the results of a binary classification for “AI” applied to the same universe of patents and patent applications by Giczy, Pairolero, & Toole. Noting that the established patent classification schemes maintained by USPTO and international bodies have not kept up with the rapid emergence of new AI technologies, this team used a natural language processing model to classify patents as pertaining to one of several technologies like image processing, speech processing, and machine learning, as well as hardware and AI modeling tools. For the purposes of this analysis, I’m using only their top-level classification **any\_ai**, which equals 1 when the patent is classified as *any* of their subdomains.

To illustrate the importance of using both measures, consider the AI patent applications data for Los Angeles County, CA vs. Los Alamos County, NM in 2011:

Graphical user interface, text, website

Description automatically generated

While Los Angeles County, CA, hub of the nation’s second-largest urban area, dwarfs tiny Los Alamos County, NM, in terms of population, we can see that the latter produced over 4 times as many AI patent applications per thousand residents that year. This is not entirely surprising, given the fact that Los Alamos is home to a campus of US government laboratories conducting some of the world’s most advanced and top-secret military and energy research.

On the other hand, if we take a look at the same two counties’ gross patent production over the time period, we see a much different picture:

Graphical user interface, text

Description automatically generated

Here, we see that over the period 2011-2019, over 20 times as many AI patents. Does Los Alamos punch above its weight in terms of innovation? Absolutely. Is it as likely to attract domestic migration as Los Angeles County? Almost certainly not in aggregate given its size disadvantage. But on a per-capita basis, its relative *specialization* as a research center should be visible in its ‘pull’, especially when considering the educational and innovation levels of the counties it pulls migrants *from.* This illustrates the need to consider control variables, which I discuss in the next section.

*Control variables*

*Education level*

Another important independent variable to the analysis is the percentage of county population with at least a bachelor’s degree. The Census Bureau publishes periodic survey results of population by education level, the most recent of which were in 2000, 2011, and 2020. For convenience, I’m using the data as published by the Department of Agriculture’s Economic Research Service, which also provides potentially relevant independent variables including each county’s ‘Rural-Urban Continuum Code’ which classifies the character of its built environment (<https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>).

*Natural Amenity Scale*

To control for counties relative geographic appeal, I turn to David McGranahan’s Natural Amenity Scale (1999) to characterize the geographic appeal of counties in the lower 48 United States. The scale is a weighted index constructed from variables including July and January temparature, rainfall, and sunshine, topographic variation, proximity to water, and level of urban development. Speaking generally, the highest scoring counties are drier, warmer, sunnier, more mountainous, less developed, and/or closer to water, while the lowest scoring counties are colder, wetter, cloudier, flatter, more developed, and/or further from a body of water. The use of the data set helps control for migration motivated in whole or in part by climate and geographic concerns.

Map

Description automatically generated

*Tax burden on individuals*

Sociologists have examined the so-called ‘Tax Flight’ phenomenon in which high-income individuals migrate in search of a lower tax liability (for example, see Young, C., Varner, C., Lurie, I. Z., & Prisinzano, R, 2016). Each origin and destination county’s state and local personal tax burden can be calculated as the proportion of the county’s per capita Adjusted Gross Income collected in the form of State or Local income taxes, personal property taxes, real estate taxes, or general sales taxes, minus refunds. This is calculated from another part of the Internal Revenue Service’s SOI publication – the Tax Stats County datasets (<https://www.irs.gov/statistics/soi-tax-stats-county-data>).

*Housing and Rental Market affordability*

Jeanty, Partridge, and Iriwin (2010) develop a spatial simultaneous equation for the effect of housing market factors on migration, while Jia et al. discuss housing affordability and access as factors in the dynamics of migration since 1980 (2022). Housing and rental market affordability may be determined, as measured by the Federal Housing Finance Administration annual residential real estate price index (https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx) and the Fair Market Rent price as established by the Department of Housing and Urban Development (https://www.huduser.gov/portal/datasets/fmr.html).

*Labor Market Health*

Origin and destination labor market health may be measured by unemployment rate, with a low unemployment rate signaling a healthy labor market and a high unemployment rate signaling a distressed labor market. Partridge et al (2012) analyze an apparent diminishing influence of labor market changes on county-to-county migration in the early 2000s when compared with the historic trend. The data source will be the US Department of Agriculture Economic Research Service’s summary of Department of Labor unemployment data (https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/).

*Time Period Considerations*

Finally, a note on the years specifically included in the study. 2011 – 2019 in particular were chosen for a number of unrelated reasons. Because the education data is published only periodically, I essentially started from the choice of either 2000 or 2011 as a starting year. While data is available going back to 2000 for each of the other variables, the focus on AI patents (of which only 20 were applied for *nationwide* in that year) led me to settle on 2011, when the technologies were more firmly established (see fig 1 below, which shows that by 2011 the number of AI utility patent applications had grown to over 32,711, which accounted for over 14% of US utility patent applications).

This 9-year time period is likely long enough to smooth out periodic local labor and housing market shocks that may affect any given year’s flows for specific counties or regions. Though of course we cannot interpret results without the context of the 2008 housing market crash and 2008 – 2010 Great Recession in mind, starting our analysis in 2011 should avoid skewing the data based on the most *temporary* impacts of these shocks. Similarly, ending in 2019 allows me to focus on the movement of workers without the effects of the wide structural economic, cultural, and political re-alignments following the 2020 pandemic and recession; in some cases, moreover, the data is less available, less reliable, or both starting in 2020.

Chart, line chart

Description automatically generated

*Total Artificial Intelligence patent applications and defensive publications produced in the United States by year, 2000 – 2019.*

REFERENCES

Adams, J. D., & Jaffe, A. B. (1996). Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data. *The RAND Journal of Economics*, *27*(4), 700–721. https://doi.org/10.2307/2555878

Alderucci, D., Branstetter, L. G., Hovy, E., Runge, A., Ryskina, M., & Zolas, N. (2019). *Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata1*. https://www.semanticscholar.org/paper/Quantifying-the-Impact-of-AI-on-Productivity-and-Alderucci-Branstetter/49d415cf593be38c6cd97a183dadc7d7b48bab72

Berry, C., & Glaeser, E. (2005). *The Divergence of Human Capital Levels Across Cities* (No. w11617; p. w11617). National Bureau of Economic Research. https://doi.org/10.3386/w11617

Bishop, B. (2009). *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Houghton Mifflin Harcourt.

Chattergoon, B., & Kerr, W. R. (2022). Winner takes all? Tech clusters, population centers, and the spatial transformation of U.S. invention. *Research Policy*, *51*(2), 104418. https://doi.org/10.1016/j.respol.2021.104418

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). *Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States* (Working Paper No. 19843). National Bureau of Economic Research. https://doi.org/10.3386/w19843

Florida, R. (2014). *The Rise of the Creative Class--Revisited: Revised and Expanded*. Basic Books.

Jaffe, A. B. (1986). *Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits and Market Value* (Working Paper No. 1815). National Bureau of Economic Research. https://doi.org/10.3386/w1815

Jeanty, P. W., Partridge, M., & Irwin, E. (2010). Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. *Regional Science and Urban Economics*, *40*(5), 343–352. https://doi.org/10.1016/j.regsciurbeco.2010.01.002

Jia, N., Molloy, R., Smith, C. L., & Wozniak, A. (2022). *The Economics of Internal Migration: Advances and Policy Questions* (SSRN Scholarly Paper No. 4028260). https://doi.org/10.17016/FEDS.2022.003

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, *25*(3), 173–196. https://doi.org/10.1257/jep.25.3.173

Moretti, E. (2012). *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Orlando, M. J. (2003). Measuring R&D Spillovers: On the Importance of Geographic and Technological Proximity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.387840

Partridge, M. D., Rickman, D. S., Olfert, M. R., & Ali, K. (2012). Dwindling U.S. internal migration: Evidence of spatial equilibrium or structural shifts in local labor markets? *Regional Science and Urban Economics*, *42*(1), 375–388. https://doi.org/10.1016/j.regsciurbeco.2011.10.006

Russell, J. (2017, June 14). *Do Jobs Follow People or Do People Follow Jobs?* Pacific Standard. https://psmag.com/economics/jobs-follow-people-people-follow-jobs-69354

DATASETS

Federal Housing Finance Administration (2022). Annual House Price Indexes.

Accessed February 26, 2023 from [https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index Datasets.aspx#qat](https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index%20Datasets.aspx#qat).

Internal Revenue Service. SOI Tax Stats – County Data. Updated December 19, 2022.

Accessed February 26, 2023 from <https://www.irs.gov/statistics/soi-tax-stats-county-data>.

United States Census Bureau, (2022). United States Population Estimates, Vintage 2022. Accessed from <https://www2.census.gov/programs-surveys/popest/datasets/>.

United States Census Bureau, (2019). United States Population Estimates, Vintage 2019. Accessed from <https://www2.census.gov/programs-surveys/popest/datasets/>.

United States Census Bureau, (2009). United States Population Estimates, Vintage 2009. Accessed from <https://www2.census.gov/programs-surveys/popest/datasets/>.

United States Department of Agriculture (2022). Unemployment and median household income for the U.S., States, and counties, 2000-2021. Economic Research Service. Last Updated June 3, 2022. Accessed February 26, 2023 from <https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>

United States Department of Housing and Urban Development. Fair Market Rents (40th Percentile Rents). Office of Policy Development and Research. Accessed February 26, 2023 from <https://www.huduser.gov/portal/datasets/fmr.html>.