How Long is the Yellow Brick Road?: Does Proximity to a City Affect Relative Mobility?

David Sneddon

Strome College of Business, Economics Department, Old Dominion University

ECON 545: Urban Economics

Dr. Rex Sitti

December 2, 2024

How Long is the Yellow Brick Road?: Does Proximity to a City Affect Relative Mobility?

Intoduction

It's an old story. Young people full of optimism and dreams of prosperity leave home and head to the big city to make their dreams a reality. I would like to examine this trope, does proximity to a city have a statistically significant effect in intergenerational mobility? Prior literature (Chetty et al., 2014), opines rather than a land of opportunity, the U.S. is "better described as a collection of societies, some of which are 'lands of opportunity' " While inequality of intergenerational mobility across demographic lines has been well documented in the literature, I would like to examine how geographic location alone affects mobility.

My interest in this starts with a purely anecdotal personal observation. I spent a combined ten years as a field sales representative and sales manager for a tobacco company. You see all kinds in this line of work and my travels took me into every sort of neighborhood, both rural and urban, wealthy and impoverished, and everything in between. The lions share of our B2B clients were convenience stores, and nearly every neighborhood in America has a convenience store. One tends to get a sense of the mindset of the managers and clerks that work in these stores. In the urban setting, typically the owners or members of their family are running their stores, often times from open to close. When employees are hired, they typically change from one month to the next. Jobs in these stores are not coveted by the urban labor market.

Contrast this to the rural setting. Many of these stores are corporate owned and those employed there tend to be there for a relatively long time. The store managers are dedicated to what they are doing, take the job seriously and generally have a no-nonsense attitude. One gets a sense, especially in extremely rural areas, that a manager in one of these stores has a vested interest in the store's success. In simple terms, these jobs are viewed as "good jobs." This stands to reason, as rural towns have limited work opportunities as opposed to their urban counterparts. It was not uncommon to see a grandmother be store manager, while her daughter works in the store as a clerk. It was also not uncommon to see other family members of the store manager visit regularly to get advice, arrange child-care, borrow gas money, borrow the car, etc. I got the sense that to be a store manager for the local gas-station was stability for a family, I also got the sense that her children weren't leaving town anytime soon. The further away from a metro area I would get, the more I noticed this phenomenon.

This also begs the question: How is distance to a city affecting intergenerational mobility? While more people may on average be willing endure longer commute times for work, this endurance isn't infinite and the literature suggests that this willingness to endure is waning. Today, workers will either move closer to more lucrative opportunities, or they will make-do with relatively limited opportunities closer to home. It would appear that workers since the end of the nineteenth century have opted for the former. By 1920 over half the population of the United States lived in a city, and by 2010 only 19% lived in a rural area, with 14% living in a non-metro county (Slack & Jensen, 2020). In this context, access to the resources and labor opportunity that a city provides continues to be attractive to workers, even in a post COVID-19 world, albeit with a caveat that this relationship may be increasingly strained.

My hypothesis is that there will be a negative relationship between the distance from the most densely populated county to other counties within an MSA and the measure of relative mobility. The null hypothesis is that there is no relationship between distance and mobility. Another alternative is that there is, in fact, a positive relationship.

Literature Review

Rank-Rank Slope as a measure of mobility

The importance of location for intergenerational mobility is well documented in the literature. Chetty et al. (2014) offers a granular analysis of the U.S. geography, using measurement methods largely derived from Dahl and DeLeire (2008), and their treatment on the data from income tax returns and W-2 forms made available from the IRS was replicated in 2022, offering a post COVID-19 sample to test (Chetty et al., 2022). Their analysis tests residential segregation, income inequality, quality of schooling, social capital, and family stability. They reject the null hypothesis on each of these covariates.

Chetty et al. (2014) uses what they call rank-rank specification. This is derived from a simple correlation of the percentile rank of a child's income and their parents' income yielding a coefficient for a single variable OLS regression.

Let:

R_{ic} : National income percentile rank of a child i, who grew up in location c

c : U.S. County [1]

$\rho_{PR} = Corr(P_i, R_i)$: Slope of rank-rank relationship

P_{ic} : Parents' percentile rank of child i

$$R_{ic} = \alpha_c + \rho_{PR} P_{ic} + \varepsilon_{ic} \tag{1}$$

Both Chetty et al. (2014) and Dahl and DeLeire (2008) reject using intergenerational elasticity (IGE) as the measure of mobility. Chetty et al. (2014) indicates the "most common method of estimating IGE" as:

$$IGE = \rho_{XY} \frac{SD(log(Y_i))}{SD(log(X_i))} \tag{2}$$

[1] In Chetty et al. (2014), commuting zones were used, however I will be using U.S. counties in place of this which is available in the Chetty et al. (2022) data set.

"where $\rho_{XY} = Corr(log(X_i), log(Y_i))$ is the correlation between log child income and parent income and $SD(\cdot)$ denotes the standard deviation."

The inherent limitations of Equation 2 are expounded upon by Dahl and DeLeire (2008). In their econometric analysis they find that use of percentile ranks as opposed to natural logarithms is robust to the means by which parent and child earnings are measured and samples collected. They addressed the deficiencies in previous literature on attenuation bias and the sensitivity in previous models to periods in parents' work history where no income was earned.

Geographical Variation in Mobility

The work done by Connor and Storper (2020) examines geographical variation in social mobility. They split the country into six regions: Northeast, Midwest, West, South, Southern Plains & Mountain, and Northern Plains & Mountain. They contrast the social mobility among these regions. They find that lower income areas show some of the highest degrees of social mobility with a caveat. The high degrees of social mobility from parent to child is overstated considering the gains made by children that *leave* the poorer region. The largest gains were found to be in the Northern Plains region among children from rural communities. Indeed the rural segments of *all* regions outperformed their urban counterparts, which is explained by gains made by those that leave, not those that stayed. This lends promise to rejecting the null hypothesis.

Work from Home Trends

According to the U.S. Census Bureau (2023c), commuting times are steadily on the rise again as the effects of the COVID-19 pandemic wane. Lund et al. (2021) predicted that that work from home (WFH) trends would be persistent after the pandemic ended with up to 25% of the workforce being deployed in an entirely work from home or hybrid environment. The data are promising for WFH and there doesn't appear to be evidence for lost productivity for WFH accompanying a drop in turnover where at least hybrid models are implemented (Bloom et al., 2024). Immediately before the pandemic 9.8% of workers had a commute time of over 60 minutes. While this dropped to a low of 7.7% as lock downs were enforced and work-from-home became more en-vogue [2], the rate rose again to 8.9% by the end of 2023 (U.S. Census Bureau, 2023b).

Recent literature that suggests that return to office (RTO) policies are being implemented regardless of the data supporting a WFH model. Senior managers may blame poor performance on WFH policies in spite of evidence to the contrary, however not without a measure of regret after implementing (Ding & Ma, 2024). Therefore, this trend may not hold long-term as employees continue to demand WFH and hybrid models and the costs of implementing RTO, including the damage to employee morale, continue to manifest without an accompanying improvement in productivity.

The Great Gatsby Curve

In his work, Miles Corak (2013) suggests that areas of high income inequality negatively effect intergenerational mobility, developing the so-called Great Gatsby Curve. In Corak et al. (2010) the authors tie in mental and physical health, specifically preventative healthcare into educational outcomes. They also hone in on educational attainment by mothers having a pronounced impact on the education outcomes of their children. They do, however, stop short of drawing a causal inference for the purpose of policy recommendation and resource distribution (Corak, 2013).

In identifying causes of income inequality, Corak (2013) opines: “the American education system does not promote mobility to the extent that it could because its educational spending is more likely to benefit the relatively well-to-do.” He finds based on OECD (2011) data that while the United States spends the most per pupil of any country in the world, this is primarily spent on tertiary education. OECD (2012) succinctly notes as identified by Corak (2013): “Currently the United States is one of only three OECD countries that on average spend less on students from disadvantaged backgrounds than on other students. . . . Moreover, the most able teachers rarely work in disadvantaged schools in the United States, the opposite of what occurs in countries with high-performing education systems.”

[2] Rising from 5.7% in 2019 to 17.9% in 2021, falling back to 13.8% by the end of 2020 (U.S. Census Bureau, 2023b)

Rural-Urban Divide

On the other hand, in their working paper Ahsan et al. (2020) posit: “The returns to education are usually higher in the urban areas because the manufacturing and services activities are more skill intensive and may give rise to agglomeration externalities.” [3]

Ahsan et al. (2020) conclude that the quality of investment in schooling, which tends to fall behind investment in urban schooling, can serve as a counter-balance to a home life and family background less conducive to educational attainment. They go on to cite peer interactions in a higher performing school as the primary counterweight to family influence. They also strongly suggest that this should influence policy recommendations for investment in rural schools.

[3] Ahsan et al. (2020) cite World Bank Group (2016) as evidence.

Method

I will be testing whether the distance between the most populous county in a Metropolitan Statistical Area (MSA) and other counties within will have a statistically significant impact on relative mobility. Data analysis is conducted using R (R Core Team, 2024) within the RStudio environment (RStudio Team, 2020). All "packages" mentioned should be understood to be R packages.

Let:

M_c : Rank-rank slope measuring relative mobility in a county

d : Distance from each county to the most population dense county (centroids)

γ_s : Fixed effects for densest counties in a CSA

$$M_c = \beta_0 + arsinh(\beta_1(d)) + \gamma_c + \varepsilon_c \tag{3}$$

Data

Geographic Data

Raw county-level geographic data is provided by the `tigris` package using `tigris::counties()` (Walker, 2024). County centroids are then calculated using the `sf` package's `sf::st_centroid()`. Latitude and longitudes were extracted using `sf::st_coordinates()` (Pebesma & Bivand, 2023). County listings and FIPS codes were taken from the core based statistical areas (CBSAs), metropolitan divisions, and combined statistical areas (CSAs) data set provided by the U.S. Census Bureau (2023a). Underlying code and data has been made available on GitHub (Sneddon, 2024) and the appendix.

Population Densities

It was necessary to obtain population densities in each county in order to find the most populous counties within a CSA. These data were obtained by calculating the densities using population data obtained from the 2020 decennial census by the U.S. Census Bureau (2020) using the `censusapi` package, courtesy of Recht (2022) and divided by the county land area extrapolated from the spatial data provided by `tigris::counties()` (Walker, 2024) and calculated using `sf::st_area()` (Pebesma & Bivand, 2023).

Rank-Rank Slope

The Rank-Rank slope from the replication data provided by Chetty et al. (2022) is specifically from "Online Data Table 3: Intergenerational Mobility Statistics by County". This is then merged with the data already collected to form a complete data set used in the test.

Distances

Distances are calculated using great circles and the haversine formula. In simple terms, the distance between centroids are "as the crow flies" and does not take into account driving distance. This is calculated using `geosphere` package and the `geosphere::distHaversine()` function (Hijmans, 2022).

Results

Regression

The `fixest` package and specifically the `fixest::feols()` function are used for the regression. Standard errors are robust to heteroscedasticity. Fixed effects for the densest counties within a CSA are chosen primarily to account for the variance in intergenerational mobility by region indicated in Connor and Storper (2020).

Table 1

Fixed-effects regression using inverse hyperbolic sine of distance from most population dense county in a CSA predicting the Rank-Rank Slope and thus the intergenerational mobility of a county.

Dependent Variable:	M_c
$arsinh$(Distance)	-0.172
	(0.0334)
Fixed-effects	
Densest Counties	917 Counties
Fit statistics	
Observations	1,764
R^2	0.7539
Within R^2	0.0386

Heteroscedasticity-robust standard-errors
Significance: $\alpha = 0.001$

As shown in Table 1, we can reject the null hypothesis, as there is strong evidence that a 1% increase in distance from the most populous county will result in a decrease of 0.172 of the rank-rank slope with a high level of significance with $\alpha = 0.001$. This indicates that proximity to the most populous county and thus proximity to the most urban area of a CSA has a positive effect on intergenerational mobility.

Limitations and Future Directions

This essay serves as a starting point for the effects of proximity to urban areas and intergenerational mobility. There is still more analysis to undertake. One fairly obvious limitation is how distance was calculated, the distances were calculated using great circles as opposed to actual drive time. There is likely some omitted variable bias for condition of road infrastructure. If road networks are more robust between counties this may have some effect on the results. Analysis can also be conducted so see this effect over time.

There is also a question on how significant of an effect this covariate would have in the overall model proposed by Chetty et al. (2014). I also make no assertion of a causal relationship, I only submit that there is evidence of a corollary.

Policy Recommendations

It would appear that there is significant inequality of opportunity in rural areas as opposed to urban centers. At the same time there is significant inequality of *result* in urban areas, whereas in rural areas income inequality is far less pronounced. This is well supported by the literature and the empirical analysis conducted for this paper adds to that support.

P

d

Conclusion

Overall, it would appear that proximity to an urban area does correlate to relative mobility. This is supported by the high degree of significance in the regression. There are undoubtedly more opportunities for income advancement in urban centers as opposed to rural areas. In addition to being highly intuitive, the literature strongly support this assertion.

References

Ahsan, M. N., Emran, M. S., & Shilpi, F. J. (2020). *The rural-urban divide and intergenerational educational mobility in a developing country: Theory and evidence from indonesia* (Policy Research Working Paper WPS 9464). World Bank Group. http://documents.worldbank.org/curated/en/414131604430277237/The-Rural-Urban-Divide-and-Intergenerational-Educational-Mobility-in-a-Developing-Country-Theory-ar

Bloom, N., Han, R., & Liang, J. (2024). Hybrid working from home improves retention without damaging performance [Journal Article]. *Nature*, *630*(8018), 920–925. https://doi.org/10.1038/s41586-024-07500-2

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity: The geography of intergenerational mobility in the united states. *Quarterly Journal of Economics*, *129*(4), 1553–1623. http://qje.oxfordjournals.org/content/129/4/1553.full.pdf

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2022). *Replication Data for: Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States* (Version V1) [dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/NALG3E

Connor, D. S., & Storper, M. (2020). The changing geography of social mobility in the united states. *Proceedings of the National Academy of Sciences*, *117*(48), 30309–30317. https://doi.org/10.1073/pnas.2010222117

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *The Journal of Economic Perspectives*, *27*(3), 79–102. https://www.jstor.org/stable/41955546

Corak, M., Curtis, L., & Phipps, S. (2010). *Economic mobility, family background, and the well-being of children in the united states and canada* (Discussion Paper 4814). Institute of Labor Economics (IZA). https://doi.org/10.2139/ssrn.1570432

Dahl, M., & DeLeire, T. (2008). *The association between children's earnings and fathers' lifetime earnings: Estimates using administrative data* (DP 1342-08). Institute for Research on Poverty, University of Wisconsin–Madison. https://irp.wisc.edu/publications/dps/pdfs/dp134208.pdf

Ding, Y., & Ma, M. (Shuai). (2024). Return-to-office mandates. *SSRN Electronic Journal*.

https://ssrn.com/abstract=4675401

Hijmans, R. J. (2022). *Geosphere: Spherical trigonometry*. https://CRAN.R-project.org/package=geosphere

Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., & Robinson, O. (2021). *The future of work after COVID-19*. McKinsey Global Institute. https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19#/

OECD. (2011). *Education at a glance 2011: OECD indicators*. OECD Publishing. https://doi.org/10.1787/eag-2011-en

OECD. (2012). *OECD economic surveys: United states 2012*. OECD Publishing. https://doi.org/10.1787/eco_surveys-usa-2012-en

Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Recht, H. (2022). *Censusapi: Retrieve data from the census APIs*. https://CRAN.R-project.org/package=censusapi

RStudio Team. (2020). *RStudio: Integrated development environment for r*. RStudio, PBC. http://www.rstudio.com/

Slack, T., & Jensen, L. (2020). The changing demography of rural and small-town america. *Population Research and Policy Review*, *39*(5), 775–783. https://doi.org/10.1007/s11113-020-09608-5

Sneddon, D. (2024). *Dasneddon/IGM*. https://github.com/dasneddon/IGM.

U.S. Census Bureau. (2020). *2020 decennial census data*. https://api.census.gov/data/2020/dec/dp

U.S. Census Bureau. (2023a). *Core based statistical areas (CBSAs), metropolitan divisions, and combined statistical areas (CSAs)* [dataset]. https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html

U.S. Census Bureau. (2023b). *Means of transportation to work: Public transportation and worked from home*

[American Community Survey Report]. U.S. Census Bureau. https://www2.census.gov/programs-surveys/commuting/guidance/acs-1yr/Mean-public-worked-from-home.pdf

U.S. Census Bureau. (2023c). *Percentage of workers traveling 60 minutes or more* [American Community Survey Report]. U.S. Census Bureau. https://www2.census.gov/programs-surveys/commuting/guidance/acs-1yr/Percent-60-or-more-minutes.pdf

Walker, K. (2024). *Tigris: Load census TIGER/line shapefiles*. https://CRAN.R-project.org/package=tigris

World Bank Group. (2016). *Indonesia's rising divide* [Report]. World Bank Group. http://documents.worldbank.org/curated/en/267671467991932516/Indonesias-rising-divide

Appendix

Code

```
library(tigris)

options(tigris_use_cache = TRUE)

library(sf)

library(dplyr)

library(censusapi)

library(units)

library(geosphere)

library(fixest)

options(scipen=999)

library(readr)




###IMPORT GEOGRAPHIC DATA


counties <- counties()

counties$centroid <- st_centroid(counties$geometry)

counties$lon <- sf::st_coordinates(counties$centroid)[,1]

counties$lat <- sf::st_coordinates(counties$centroid)[,2]

counties <- counties[,c("GEOID", "lon", "lat")]

cbsa <- read.csv("csa-counties.csv")

cbsa$GEOID <- paste0(formatC(cbsa$FIPS.State.Code, width=2, flag=0),
                     formatC(cbsa$FIPS.County.Code, width=3, flag=0))
```

```
###IMPORT COUNTY POPUATIONS

county_pop <- getCensus(
  "dec/dhc",
  vintage = 2020,
  vars = c("NAME", "P1_001N"),
  region = "county:*")

county_pop$GEOID <- paste0(county_pop$state, county_pop$county)
county_pop <- merge(counties,county_pop, by="GEOID")
county_pop$area <- set_units(st_area(county_pop$geometry), "mi^2")
county_pop$density <- county_pop$P1_001N/county_pop$area

##MERGE POPULATIONS AND GEOGRAPHY
cbsa <- merge(cbsa,county_pop, by="GEOID")
cbsa$densest <- 0
cbsa$distance <- set_units(0, "m")

##FIND THE DENSEST COUNTY

densecounty <- function(combstatarea){
  cbsaframe <- cbsa[cbsa$CBSA.Code==combstatarea,]
  highest <- max(cbsaframe$density)
  return(cbsaframe$GEOID[cbsaframe$density==highest])
}

for (i in unique(cbsa$CBSA.Code)){
```

```
  flag <- densecounty(i)
  cbsa$densest[cbsa$CBSA.Code==i] <- flag
}

for (i in 1:nrow(cbsa)){
  flag <- cbsa$densest[i]
  cbsa$distance[i] <- distHaversine(c(cbsa[i,"lon"],
                                      cbsa[i,"lat"]),
                                    c(cbsa$lon[cbsa$GEOID==flag],
                                      cbsa$lat[cbsa$GEOID==flag]))
}

#IMPORT RANK-RANK SLOPE

county_rank_slope <- read_csv("county_rank_slope.csv")
cbsa <- merge(cbsa, county_rank_slope, by="GEOID")
cbsa$RS <- cbsa$RS*100

##REGRESSION

reg1 <- feols(RS ~ asinh(distance) | densest,
              data = cbsa, vcov = "hetero")

summary(reg1)



write_csv(cbsa, "cbsa.csv")
```