Research on Economic Indicators and Commute Time

Shiwei Wang

University of California, San Diego, CA 92093-0021, USA

Keywords: Commute time; Economic indicators; Linear regressions

Abstract: According to the theory of urban economics, transportation is closely related to the emergence and development of cities. With the acceleration of urbanization, commuting becomes an important part of city functions and residents' lives. Issues such as commuting patterns and commuting costs have become the focus of research in urban economics and other disciplines. and other disciplines such as urban economics. In this research, population density, housing rent fees, labor participation rate, and mean income level will be examined to determine which one could increase commute time the most with linear regressions.

1. Introduction

The outbreak of COVID-19 has freed people from commuting for a while and with less road rages. Rushing between work and home is the fixed and repetitive part of a day. Time people put in commuting suggests a great amount of information—living locations, transportation method, and the place's economic development. Past research have focused on how people search for jobs considering the distance to travel everyday (Berg and Gorter 1997)[1]. Through further research, more factors affecting commute times can be revealed to examine the economic situations at a place. While commuting time is a far more complex situation can be solely assessed by weighing variables, there can be indicators predicting a general level of commuting time.

Commute time is highly dependent on population, price of housing, number of people employed, and income. Intense rush hours always occur in busy metropolitans where price of housing is high, with a dense population. Number of people employed is the very essence source of rush hour congestions. Higher employment should link with longer commute time. Mean income can be viewed as a level indicator for a region's economic development. An increase in mean income can potentially lead to both increase in commute time and a decrease in commute time, due to personal tastes of neighborhoods to live or preference of a closer place to live from work.

In this research, population density, housing rent fees, labor participation rate, and mean income level will be examined to determine which one could increase commute time the most with linear regressions. The goal of this research is to examine the bizarre behavior of an increase in income would lead to an increase in commute time. The primary hypothesis would be that commute time is higher in places with higher level of income, labor participation rate, house rental prices, and population density. For housing rents and mean income, it is believed that as income level become higher, commute time increases because mean income indicates a more economically developed city, with more population density. Dr. Daniel Wescott states: "Higher incomes are generally associated with commuting outside one's area of residence" (Wescott 1). As labor participation rate increases, the obligation to commute falls upon more people, thus increasing commute time by adding traffic numbers. With a higher rental fee level, people tend to choose living places farther to avoid costly rents with only a few more minutes' commuting. Population density indicates a place's economic development most. With higher population density, commuting will only take longer.

2. Data

The dataset being analyzed is Neighborhood Characteristics by County. The dataset specifies a series of indicators of economic situations and community characteristics conducted by Raj Chetty[2], John Friedman, Nathaniel Hendren, Maggie Jones, and Sonya Porter, as a collaboration

  DOI: 10.25236/icemeet.2023.005

between Opportunity Insight research group in Harvard University and the Census Bureau. The time span of their research is wide and extensive. Most of the variables have data across different years that can be used as a timely comparison reaching aspects of job, income, and lifestyle of each county and can even be categorized by state with the provided FIPS code. Although the unit is by county, some of the variables are also based on households, which gives a more detailed data with more accurate predictions. In this research, the main dependent variable is set to be commute time. In the dataset of Neighborhood Characteristics by County, commute time is collected with calculating the mean commuting time among workers over 16 years old (Neighborhood Characteristics by County). As shown in table 1, commute time ranges from 7 minutes to 46 minutes, with a mean of 24 minutes. Examining its median and mean values that median is slightly smaller than mean, we can determine that the data is rightward skewed.

Table 1: Commute Time Information

```
. * Step 2_a: get vital information of commuteTime.
. univar commuteTime
                                          -------------- Quantiles --------------
Variable       n     Mean     S.D.      Min      .25      Mdn      .75      Max
-------------------------------------------------------------------------------
commuteTime    3142    24.36     5.64     7.23    20.34    23.92    27.81    46.06
-------------------------------------------------------------------------------
```

Table 2: Data Dictionary

Variable Name	Description
state	Two-digit 2010 state FIPS code.
county	Three-digit 2010 county FPIS code
Mean_commutetime2000	Mean commute time for workers over 16 years old in the county.
hhinc_mean2000	Mean household income.
Emp2000	Employment divided by total population over 16 years old. Equal to labor participation rate.
Rent_twobed2015	The median gross rent for renter-occupied housing units with two bedrooms that pay cash rent.
Popdensity2000	Number of residents per square mile.

As shown in table 2-3, Independent variables are mean household income in 2000 (hhinc_mean2000), labor participation rate in 2000 (emp2000), house rental fee in 2015 (rent_twobed2015), and population density in 2000 (popdensity2000). The independent variables in this report have correlations between themselves by nature, and each of them is one indicator of economic situation in a county. Because the research goal is to search for a dynamic relation between changing independent variables and the changing commute times, min and max values should be given with enough attention.

Table 3: Independent Variable Information

```
. * Put all four independent variables into one summarize table.
. univar meanIncome labor_participation_rate rentFee population_density
                                          -------------- Quantiles --------------
Variable       n     Mean     S.D.      Min      .25      Mdn      .75      Max
-------------------------------------------------------------------------------
meanIncome    3142 65104.69 15318.74 22266.88 55242.07 61607.81 70959.12  1.5e+05
labor_participation_rate    3142     0.57     0.08     0.24     0.53     0.58     0.63     0.84
 rentFee    3143   684.90   208.21   172.16   560.18   637.72   756.77  2085.23
population_density    3219 24085.79  1.7e+05     0.00  1498.46  4086.15 10177.52  6.8e+06
-------------------------------------------------------------------------------
```

3. Methods and Results

3.1 Mean Income

In this research, each independent variable was analyzed by doing a regression analysis to examine their respective relationships to commute time. As shown in Figure 1, income and commute time forms a positive relation.

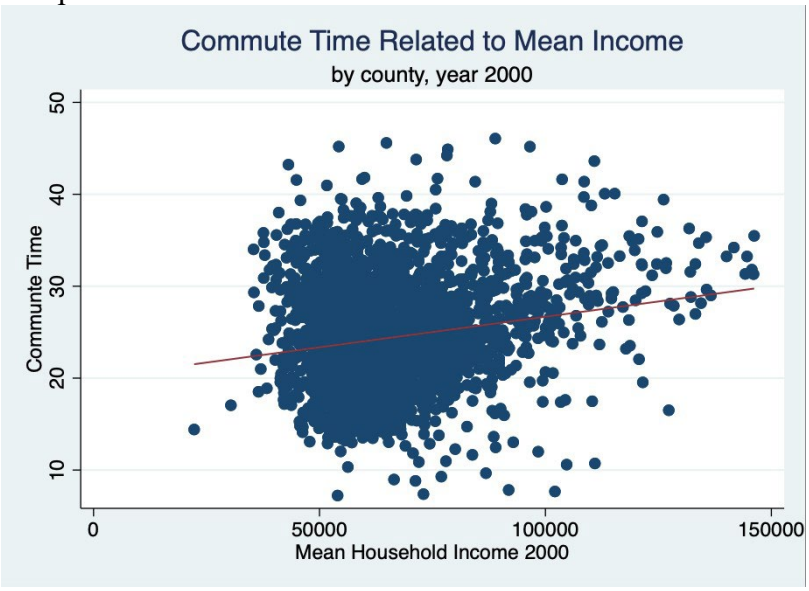

Figure 1. Commute Time Related to Mean income

Although commute time varies when income approaches 6000 from right and left side, the correlation is still visible. Further examinations with the regression result as shown in table 4 below indicates that for every increase of 1 unit of income, we can expect a 0.0042 more seconds of commuting time in this model. The plausibility of this regression is prominent. From the 95% confidence intervals, the confidence interval lies between 0.0000538 to 0.0000791, which does not include zero. Although the slope and the confidence interval are really close to zero, there are still 3.26% of the data can be explained by the income level in a county. From the other way to examine this regression result is to see if we are assuming this near-zero slope is being estimated by a random chance. To represent the random possibility, p-value indicates that the possibility is 0, smaller than the 0.001 threshold.

Table 4. Regress commuteTime meanIncome

```
. regress commuteTime meanIncome
```

Source	SS	df	MS	Number of obs	=	3,142
				F(1, 3140)	=	105.69
Model	3253.72175	1	3253.72175	Prob > F	=	0.0000
Residual	96668.493	3,140	30.7861443	R-squared	=	0.0326
				Adj R-squared	=	0.0323
Total	99922.2147	3,141	31.8122301	Root MSE	=	5.5485

commuteTime	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
meanIncome	.0000664	6.46e-06	10.28	0.000	.0000538	.0000791
_cons	20.0357	.4322454	46.35	0.000	19.18819	20.88322

3.2 Labor Participation Rate

Labor participation rate, from the description of emp2000, is itself the labor participation rate, as the formula illustrates:

$$Labor\ Participation\ Rate\ =\ \frac{Employed}{Total\ Population\ over\ 16}$$

Labor participation rate is closely related to the employment rate. When employment increase, more people will be going to work every day adding burden to the traffic. According to Figure 2, however, there is a bizarre phenomenon that when labor participation rate increases, commute time decreases.

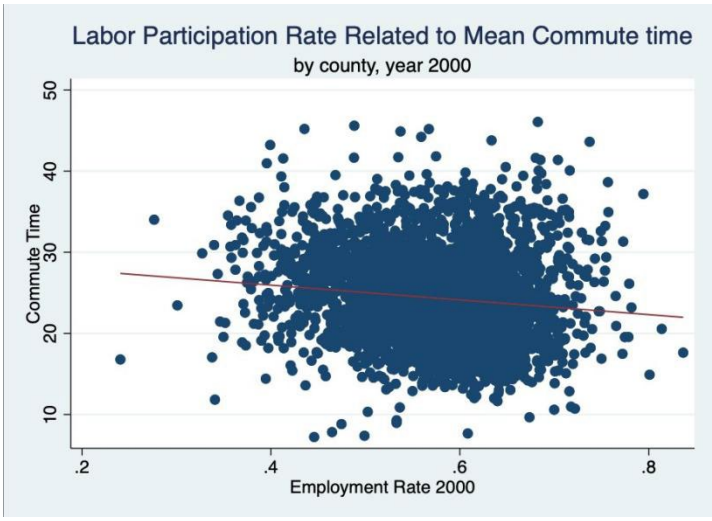

Figure 2. Labor Participation Rate Related to Mean Commute time

To determine if the result can be accepted, regression results are being investigated, as figure 3 shown. The slope is -0.09, meaning every increase in 1 percent point of labor participation rate, there will be 0.09 minute decrease in commuting time. From the column specifying 95% confidence interval, we know the confidence interval is from -0.117 to -0.647 which does not include zero, indicating the result is significant enough. The p-value is also 0, meaning there is no possibility of estimating the slope at a random chance.

```
. regress commuteTime labor_participation_rate
```

Source	SS	df	MS		
Model	1466.31259	1	1466.31259	Number of obs	= 3,142
Residual	98455.9021	3,140	31.3553828	F(1, 3140)	= 46.76
				Prob > F	= 0.0000
				R-squared	= 0.0147
				Adj R-squared	= 0.0144
Total	99922.2147	3,141	31.8122301	Root MSE	= 5.5996

commuteTime	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
labor_participation_rate	-.090775	.0132742	-6.84	0.000	-.116802	-.064748
_cons	29.57675	.7691814	38.45	0.000	28.0686	31.0849

Figure 3. The regression results

3.3 Housing Rent

Housing rent is also an economic as well as commuting time indicator. When the mean housing rents are relatively lower between counties, there are more people willing to rent houses near their workplaces to reduce commuting time. According to figure 4, an increase in housing rent will increase commuting time, because the increase in housing price forces people to buy farther places where the locations are not ideal, but prices are reasonable. Knowing how housing rents and commuting time relates, significance of this graph should be examined.

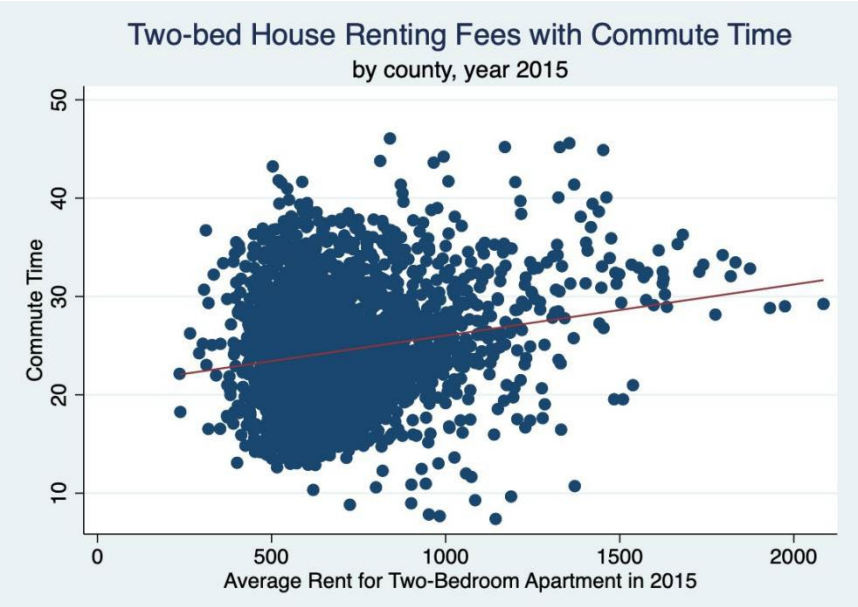

Figure 4. Two-bed House Renting Fees with Commute Time

As displayed in Figure 5, the slope is 0.0052, meaning an increase in one dollar of rental fee makes the commute time increase by 0.312 seconds on this model statistically. The 95% confidence level ranges from 0.0042 to 0.0061, which does not include zero. It is confident to say that our result is plausible. There are 3.6% of the commute time is being explained by housing rent fees. Last, the p-values are also zero.

```
. regress commuteTime rentFee
```

Source	SS	df	MS	Number of obs	=	3,066
				F(1, 3064)	=	114.63
Model	3477.622	1	3477.622	Prob > F	=	0.0000
Residual	92955.9286	3,064	30.3380968	R-squared	=	0.0361
				Adj R-squared	=	0.0357
Total	96433.5506	3,065	31.4628224	Root MSE	=	5.508

commuteTime	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
rentFee	.0051953	.0004852	10.71	0.000	.0042438	.0061467
_cons	20.83614	.3503472	59.47	0.000	20.1492	21.52307

Figure 5. Regress commuteTime rentFee

3.4 Population Density

It should be obvious that as a county gets more populated, commute time would increase. This is illustrated by the regression between population density and commute time in Figure 6.

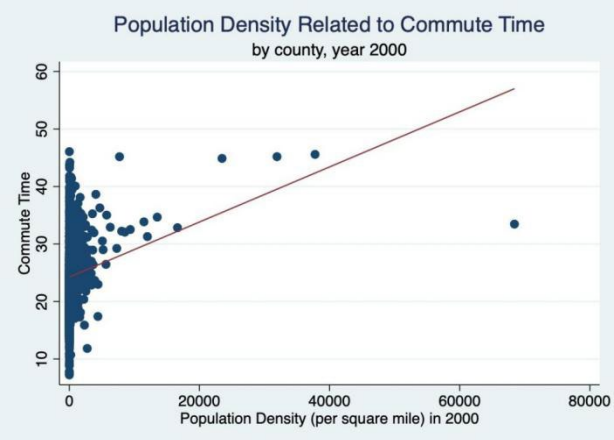

Figure 6. Population Density Related to Commute Time

Although the scatter plot, by observation, concentrates where population density is equal to zero, further regression result should be analyzed as shown in Figure 7. As shown, the coefficient is 0.00048, which implies that every increase in 1 unit of population density, it will cause commute time increase by 0.029 seconds in this model. The 95% confidence level ranges from 0.0003657 to 0.0005924 without zero included and the p-value is also 0. We can see there are 2.14% of the commute time can be explained by population density.

```
. regress commuteTime population_density
```

Source	SS	df	MS		
Model	2138.49001	1	2138.49001	Number of obs	= 3,141
Residual	97760.6658	3,139	31.1438885	F(1, 3139)	= 68.66
				Prob > F	= 0.0000
				R-squared	= 0.0214
				Adj R-squared	= 0.0211
Total	99899.1559	3,140	31.8150178	Root MSE	= 5.5807

commuteTime	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
population_density	.0004791	.0000578	8.29	0.000	.0003657	.0005924
_cons	24.24457	.1005929	241.02	0.000	24.04734	24.44181

Figure 7. Regress commuteTime population_density

Lastly, correlations of each independent variables are calculated through Stata to see if some of the variables can be used to explain commute time together (As shown in figure 8).

```
. correlate meanIncome labor_participation_rate rentFee population_density
(obs=3,065)
```

	meanIn~e	labor_~e	rentFee	popula~y
meanIncome	1.0000			
labor_part~e	0.6543	1.0000		
rentFee	0.8060	0.3975	1.0000	
population~y	0.1687	0.0274	0.3255	1.0000

Figure 8. Result analysis

4. Discussion

Through a series of regressions made, there is a strong correlation between commute time and economic conditions, specifically have positive relations with income, housing rent fees, and population density, as well as negative relation with labor participation rate. The data have confirmed the hypothesis that as income, rental fee, and population density increases, commute time increases. However, regression result suggests that as labor participation rate increases, commute time decreases. Since the correlation between labor participation rate and mean income is over 50%, it is plausible that as the participation rate can be affected by income level. Another possible explanation being that labor participation rate implies that a place with lower population base, meaning the economic situation is relatively lower. In places with lower economic conditions, commute time would decrease because of lower income, lower housing rent fees, and lower population density. Another possible explanation starts with the insertion from Dr. Diane Westcott: "It is known that central cities have the highest unemployment rates" (Westcott 1979)[3]. With higher labor participation rate—higher employed population, the county's economic condition would be less of a major city but rather smaller in economic scale. With a lower economic scale, commute time would decrease because the cities itself in counties are relatively smaller. Although the four of the variables cannot entirely describe economics situations entirely, they represent a great deal of how economic impact people's life, such as income, rental price level, and commute time in this research. Some of them are also descriptive in macroeconomic sense. For example,

income is used as a method to calculate Gross Domestic Product named Income Method; labor participation rate also suggests the ratio between employed people and total population over 16 years old. Population density can also be a vital piece of information in macroeconomics. Because commuting is highly dependent on transportation method, city transportation situations, and population, which all relates to economic conditions of a county, independent variables should be or have a strong relation with economic parameters.

There are some limitations of this research, most profoundly the year collected for variable rent price level. For all other variables, their time was collected in year 2000, while rent price level variable was collected in 2015. There must be changes within this 15-year period within a county, either an economic expansion driving rent level up or an economic depression that causes a recession in local housing market. However, the scope of this research is within the boarder of the U.S by county, and the changes within the same country can be regarded on approximately the same pace, because the currency exchange rates, and economic policies are all dedicating on stabilizing markets from severe fluctuations. It is unlikely to have a county's economic situation spikes up shortly after year 2000. Therefore, it is safe to insert that the housing rent price level collected in year 2015 can still hold accountable for 2000's data.

Due to the limitations of data collections, the explanation of the unusual behavior between labor participation rate is insufficient to some level. It would be more convincing with the data of employment population in addition to the existing dataset. With running the regression between the employed population and commute time examining the positivity of the slope, one can determine that if it is the employment population or the expanding (or shrinking) population count is affecting commuting time. Within labor participation rate, the base population is the population with over 16-year-old, but population density counts the population regardless of the age, making it impossible to get only the employment population out of the existing variables.

5. Conclusion

Running the regression with commute time and the four chosen variables, labor participation rate and commute time has the strongest causal relationship. Labor participation rate may be affected by population over 16 years old, or the employed population. Overall, income level, labor participation rate, housing rental price, and population density are all factors that can affect one's commute time. The primary relationship determination can help further economic explorations on employment rate, happiness scale with respect to commute time, or income and neighborhoods. With the relationships presented above, further analysis can be made. People with higher income have a wider range of choice over commute methods, making commute become more convenient and faster, meaning with the same transportation method, people with higher income are more willing to afford a place with nicer environment and can be a little further from work. As the employed population increases among total population over 16, counties get smaller in economic scale because of the base population is relatively small with respect to the employed population. Hence, with cities in smaller scales, commute time are naturally decreased as well. The same applies to population density, another variable tested in this report. Metropolitans and major cities are always busy and filled with people. With lower population density, cities—counties—become smaller and thus lowering commute time. Rental fee is also the indicative variable for a county's economic level. With busier cities, rental fee always higher than cities with lower economic levels. It is safe to think about renting a two- bedroom-two-bathroom house in New York City compared to renting the same house in Buffalo city in Erie County.

References

[1] Van den Berg, G., & Gorter, C. (1997). Job Search and Commuting Time. *Journal of Business & Economic Statistics*, 15(2), 269-281. doi:10.2307/1392310

[2] Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). The opportunity

atlas: Mapping the childhood roots of social mobility (No. w25147). National Bureau of Economic Research.

[3] Westcott, D. (1979). Employment and commuting patterns: A residential analysis. Monthly Labor Review, 102(7), 3-9. Retrieved June 12, 2021, from http://www.jstor.org/stable/41841039