5-2023

# The Effect of Yearly Labor Earnings on Commute Time to Work in South Carolina

Peter Trela
ptrela@clemson.edu

# THE EFFECT OF YEARLY LABOR EARNINGS ON COMMUTE TIME TO WORK IN SOUTH CAROLINA

---

A Thesis
Presented to
the Graduate School of
Clemson University

---

In Partial Fulfillment of
the Requirements for the Degree
Master of Science
Economic Analytics

---

by
Peter Trela
May 2023

---

Accepted by:
Dr. Scott Templeton, Committee Chair
Dr. Babur De los Santos
Dr. Devon Gorry

**ABSTRACT**

In this paper, I attempt to ascertain the effect of labor earnings on commute time to work for individuals in South Carolina by using ACS 1-year Public Use Microdata Sample Estimates. First, I use standard linear regression models with controls to determine the direction and magnitude of the association between yearly labor earnings and commute time to work. I later use standard linear regression models with limited controls to determine how the association between yearly labor earnings and commute time changes before and during the events of the COVID-19 pandemic. There exists a positive relationship between yearly labor earnings and commute time in South Carolina. In particular, a $1000 increase in labor earnings per year is associated with, on average, an 11.75-minute increase in commute time to work per year, ceteris paribus. The relationship between yearly labor earnings and commute time becomes smaller but is still positive during COVID-19. In particular, individuals during COVID-19 experienced a decrease in commute time to work of 7.25 minutes per year per $1000 of yearly labor earnings, ceteris paribus. These results, although not causal, are consistent with economic arguments that working people require compensation from their jobs for the costs of commuting. The results affected by COVID-19 suggest that the compensating differential might become smaller as video conferencing becomes more widely accepted and, in essence, reduces average commute times.

## TABLE OF CONTENTS

Page

# LIST OF TABLES

Table Page

## LIST OF FIGURES

# 1. INTRODUCTION

Millions of Americans wake up every day and trek to their jobs, but all these trips differ in some way. Some Americans live right across the street from their work and arrive at their job in less than 5 minutes. Others have long daily commutes, sometimes upwards of 3 hours, just to get to their job. Why do these commute times vary? In my thesis, I focus on a person's labor earnings to answer the question.

What is the relationship between labor earnings and commute time? Two sets of contradictory arguments can be made. On the one hand, higher labor earnings can increase the commute time for an individual. Since an individual is getting paid more, they are willing to travel further as this higher pay acts as a compensating wage differential for the longer commute time (Manning, 2003). Moreover, if earnings from work increase, then the demand for better quality of life should increase too. A person would want to live where there is less crime, better air quality, and cheaper homes. So, they would live further away from a big city and increase their commute times.

On the other hand, higher labor earnings can decrease the commute time for an individual. As labor earnings increase, those with higher earnings might choose to live closer to work because they can better afford to do so. Prices and rents for housing of a given quality tend to increase the closer the housing is located to jobs. Moreover, as one's labor earnings increase, the more leisure of a given quality an individual demands because leisure is considered a normal good. As the demand for leisure increases, people choose shorter commute times if leisure and commuting are substitutes and part of the same time constraint.

In part A of my thesis, I use both a level-level and log-log linear regression model to determine the association between labor earnings and commute time. The data for part A come from the American Community Survey Public Use Microdata Sample Estimates from 2018 (Bureau, U. S. C, 2022). In part B of my thesis, I use a standard linear regression model to determine whether COVID-19 changed the relationship between labor earnings and commute time. I combine the American Community Survey Public Use Microdata Sample Estimates from 2018 with the American Community Survey Public Use Microdata Sample Estimates from 2021 to compare individuals before and during the events of COVID-19 (Bureau, U. S. C, 2022).

Given the level-level model in part A, a $1,000 increase in yearly labor earnings is associated with, on average, an 11.75-minute increase in commute time to work per year, ceteris paribus. Given the log-log model in part A, a 1% increase in yearly labor earnings is associated with, on average, a 0.082% increase in daily commute time to work, ceteris paribus. Both of these coefficients were significant at the one percent significance level.

Given the level-level model in part B, an individual during COVID-19 experiences a decrease in yearly commute time to work of 7.5 minutes per year per $1000 of yearly labor earnings, ceteris paribus. The coefficient is significant at the one percent significance level. This negative coefficient decreases the positive association between labor earnings and commute time to work during COVID-19, but the overall association of labor earnings on commute time was still positive and significant.

## 2. LITERATURE REVIEW

Many studies have addressed the relationship between commute time and earnings. My paper is one of the first to analyze the effect of labor earnings on commute time, as almost all other papers focus on the reverse channel of causality, the effect of commute time on labor earnings. The only paper that also analyzes the effects of labor earnings on commute time to work is a paper written by Dargay et al. (2005). Dargay et al. observe the effects of labor earnings on commute time for individuals in Great Britain. Dargay et al. use the British Household Panel Survey to gather information about individuals across time.

By using the British Household Panel Survey, Dargay et al. can choose individuals who did not change jobs and observe how commute time changed given a change in labor earnings. Dargay et al. run a fixed effects model which differences out the heterogeneity of individuals. Log commute time is the variable of interest in the fixed effects model and the primary regressor is log labor earnings. Using log labor earnings and log commute time in the fixed effects model, an income elasticity is constructed to analyze the relationship between labor earnings and commute time. Dargay et al. conclude that a 1% increase in labor earnings is associated with, on average, a 0.08% increase in daily commute time, ceteris paribus.

Since almost all other papers estimate the effect of commuting time on labor earnings, I will be analyzing two articles that have useful parallels but do not imply causation. The first article by French et al. (2020) investigates the effects of commute time to work on benefits-included labor earnings of young adults between the ages of 24-32 in the United States. French et al. use data for individuals between the ages of 24-32 because they assumed these individuals have the most competitive job market making the relationship between commute time and labor earnings more relevant.

French et al. ran a standard regression to analyze the relationship between an individual's labor earnings and their commute time to work. Using benefits-included labor earnings instead of benefits-excluded labor earnings is more accurate since benefits-included labor earnings encapsulate one's wage as well as other benefits such as dental and health insurance, leading to more consistent results. French et al. transform the benefits-included labor earnings variable into log labor earnings due to the skewness of the earnings function and use it as the variable of interest in their model. French et al. were also able to observe individuals who earned an hourly wage, which allowed them to use hourly earnings to run a different regression as a comparison to the labor earnings and log labor earnings regressions.

French et al.'s models use 3 variables that account for commute time to work. One variable is a continuous variable for the amount of time one-way an individual takes to get to work. This is their primary regressor. French et al. also include a dummy denoting whether a traveled distance is considered far (they used 20 minutes for this threshold). Lastly, French et al. include the log version of the continuous one-way travel time variable, which allows them to observe the elasticity between benefits-included labor earnings and commute time when benefits-included labor earnings are also logged.

French et al. also add many controls such as age, race, education, occupation, ability, marital status, nativity, and health into their model. This eliminates most of the endogeneity from missing covariates. Overall, French et al. conclude that 10 additional minutes of commuting time one way is associated with, on average, a 2.9% increase in annual personal earnings, ceteris paribus.

Another article written by Hazans (2004) investigates wage differentials between urban and rural workers and how they varied based on commuting. Investigating wage differentials is slightly different from investigating earnings, but many of the methods and conclusions were the same. The data that Hazans uses is Baltic employment and wage data. The 3 countries that he observes are Estonia, Latvia, and Lithuania since Hazans is a professor at the University of Latvia and has access to this information through the school.

Hazans runs a standard regression where his equations have wage differential regressed on age, age squared, education, gender, ethnicity dummies, fixed-term contract dummies, region dummies, and commute dummies. These commute dummies were delegated to denote whether an individual had to commute out of their municipality (to a larger city) to get to their job. The results varied based on the Baltic state in question.

Overall, Hazans finds in Latvia that commuters from outside the capital city make, on average, 16-17% more than those who do not commute from the same municipality, ceteris paribus. He finds in Lithuania that commuters from outside the capital city make, on average, 11% more than those who do not commute from the same municipality, ceteris paribus. Lastly, Hazans finds in Estonia that commuters from outside the capital city make, on average, 24% more than those who do not commute from the same municipality, ceteris paribus.

## 3. DATA DESCRIPTION

The data I use in my models for part A come from the United States Census Bureau, and I use the American Community Survey (ACS) Public Use Microdata Sample Estimates from 2018. The ACS data are data gathered by the United States Census Bureau and include an assortment of different data from a sample of the population (Bureau, U. S. C., 2022). The American Community Survey's main goal in obtaining all this yearly information on the population is to determine how to distribute yearly federal and state funds, so it is a reliable source when it comes to data on individuals in the population (Bureau, U. S. C., 2022). I primarily focus on the subset of the data about individuals from South Carolina.

The variables of importance I use from this dataset are labor earnings and commute time to work, but numerous controls such as general characteristic controls, occupation controls, education controls, location controls, and transportation controls were also extracted from the data. In the data, there were major outliers that occurred within the continuous variables, so an outlier-excluded group is created to conduct the analysis. The first restriction on the outlier group is that an individual was between the age of 18 and 65 years old. This was done to focus on the sample of those who are considered the working class. The second restriction is that an individual's earnings must be between $10,000 and $248,000. Figure 3.1 shows a distribution of yearly labor earnings without outliers. The reason for this restriction is that everyone whose earnings were above $248,000 was at the top-coded limit ($398,000) of the dataset, which leads to issues due to the fact that an individual can be making much more than the top-coded value. Everyone below $10,000 is most likely part-time which may change one's behaviors, so this group is removed as well. The third restriction is that an individual's commute time to work must be less than 230 minutes per day. Figure 3.2 shows a distribution of daily commute time with outliers. The reason for this restriction is everyone whose commute time was above 230 minutes per day was at the top coded limit of 306 minutes per day, which again, leads to issues due to the fact someone may commute much longer than specified leading to inaccurate estimates.

Figure 3.1: Distribution of Reported Labor Earnings with Outliers

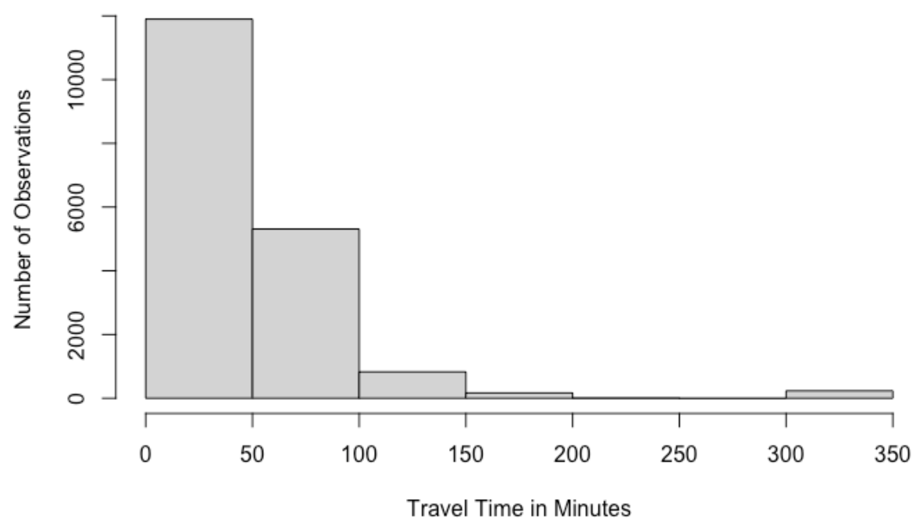
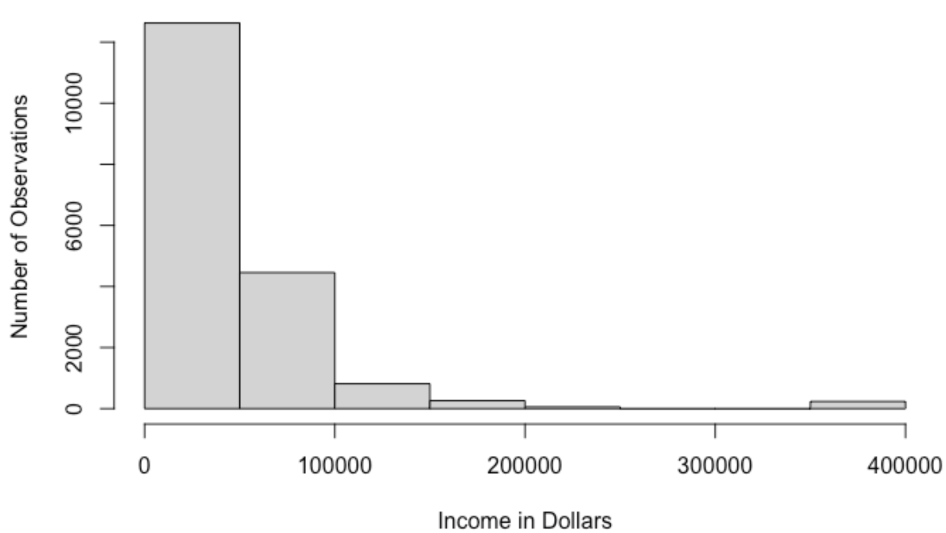



Figure 3.2: Distribution of Daily Commute Time with Outliers



The two most important variables in my data and models are labor earnings and commute time. Labor earnings is a continuous variable measured in dollars per year that focuses on the total benefits-excluded earnings an individual makes from working their job per year. Commute time is a continuous variable that focuses on an individual's commute time to work in minutes per day. Both of these variables are important to the data because they are used in the models to show what association higher labor earnings have on an individual's commute time to work.

The first set of controls used were personal characteristic controls. These variables were added to the model to control for the general characteristics of the individuals in the data (Holzer, 1991). Age is a continuous variable that denotes the age of the individual in years. Sex controls were added to control for the differences in males and females. Nativity controls were added to control for differences in individuals born in and out of the United States. Marriage controls were added to control for the differences between those married and not married. Child controls were added to control for the differences between those who have children and who do not have children. Lastly, race controls were added to control for individuals of different races. Table 3.1 shows the descriptive statistics for labor earnings, commute time, and personal characteristic controls for individual South Carolinians in 2018.

Table 3.1: Descriptive Statistics for Labor Earnings, Commute Time, and Personal Characteristic Controls in 2018 (N=14,921)

| **Variable** | **Mean** | **Standard Dev** | **Minimum** | **Maximum** | **Mean Daily Commute Time** |
|---|---|---|---|---|---|
| Commute Time (min/day) | 48.79 | 31.78 | 2 | 230 | |
| Labor Earnings (dollars/yr) | $49,212 | $34,049 | $10,000 | $248,000 | |
| Age (years) | 42.9 | 12.75 | 18 | 65 | |
| Male | 0.514 | 0.5 | 0 | 1 | 50.19 |
| Female | 0.486 | 0.5 | 0 | 1 | 47.31 |
| Native | 0.94 | 0.238 | 0 | 1 | 48.96 |
| Foreign | 0.06 | 0.238 | 0 | 1 | 46.18 |
| Married | 0.573 | 0.495 | 0 | 1 | 49.39 |
| Not Married | 0.427 | 0.495 | 0 | 1 | 47.99 |
| Child | 0.331 | 0.471 | 0 | 1 | 48.96 |
| No Child | 0.669 | 0.471 | 0 | 1 | 48.70 |
| White | 0.763 | 0.425 | 0 | 1 | 48.73 |
| Black | 0.20 | 0.400 | 0 | 1 | 49.36 |
| Native American | 0.004 | 0.061 | 0 | 1 | 48.57 |
| Asian | 0.017 | 0.13 | 0 | 1 | 45.87 |
| Pacific Islander | 0.001 | 0.028 | 0 | 1 | 36.67 |
| Other Race | 0.015 | 0.123 | 0 | 1 | 48.09 |

[a] Table was made using Outliers-Excluded Data

The second set of controls used were occupation controls. These binary variables were added to the model to control for the different occupation's individuals had. Since occupations are potentially related both to how much an individual makes and how much time it takes for an individual to get to work, these controls were crucial. I use the same occupation groups specified by the ACS. Table 3.2 shows the descriptive statistics for occupation controls for individual South Carolinians in 2018. The third set of controls used were education controls. These variables were added to the model to control for the different education levels individuals had. These education levels ranged from high school dropouts to professional degrees. Since education level is potentially related both to how much one makes and how much time it takes for one to get to work, these controls were added. Table 3.3 shows the descriptive statistics for the education controls for individual South Carolinians in 2018.

Table 3.2: Descriptive Statistics for Occupation Controls in 2018 (N = 14,921)

| Variable | Proportion | Minimum | Maximum | Mean Daily Commute Time |
|---|---|---|---|---|
| Office[d] | 0.111 | 0 | 1 | 47.14 |
| Manager | 0.110 | 0 | 1 | 48.83 |
| Sales | 0.088 | 0 | 1 | 47.19 |
| Production | 0.088 | 0 | 1 | 51.47 |
| Doctors and Nurses | 0.079 | 0 | 1 | 50.11 |
| Education | 0.075 | 0 | 1 | 42.85 |
| Transport | 0.067 | 0 | 1 | 50.79 |
| Repair | 0.046 | 0 | 1 | 53.73 |
| Cooks | 0.044 | 0 | 1 | 38.23 |
| Construction | 0.039 | 0 | 1 | 58.62 |
| Cleaners | 0.03 | 0 | 1 | 44.92 |
| Business[a] | 0.028 | 0 | 1 | 51.37 |
| Engineer | 0.026 | 0 | 1 | 52.69 |
| Computer Science | 0.023 | 0 | 1 | 53.86 |
| Healthcare Assistants | 0.023 | 0 | 1 | 49.54 |
| Protection[b] | 0.023 | 0 | 1 | 52.32 |
| Social Workers | 0.022 | 0 | 1 | 45.46 |
| Finance | 0.018 | 0 | 1 | 48.94 |
| Entertainment | 0.017 | 0 | 1 | 47.89 |
| Public Services[c] | 0.014 | 0 | 1 | 43.12 |
| Law | 0.011 | 0 | 1 | 47.25 |
| Other Science | 0.009 | 0 | 1 | 53.16 |
| Military | 0.004 | 0 | 1 | 46.2 |
| Farming | 0.003 | 0 | 1 | 51.46 |
| Mining | 0.0005 | 0 | 1 | 74.12 |

[a] Includes Resale and Wholesale Buyers, HR Employees, Event Planners, Business Analysts, and more.
[b] Includes Firefighters, Police Officers, Security Guards, Detectives, and more
[c] Includes Childcare workers, Recreational workers, Barbers, Tour and Travel Guides, and more.
[d] Includes Clerks, Office Assistants, Telephone Operators, and more.
[e] Table was made using Outliers-Excluded Data

Table 3.3: Descriptive Statistics for Education Controls in 2018 (N = 14,921)

| Variable | Proportion | Minimum | Maximum | Mean Daily Commute Time |
|---|---|---|---|---|
| High School Dropout | 0.061 | 0 | 1 | 49.94 |
| High School Graduate | 0.251 | 0 | 1 | 49.07 |
| Some College | 0.21 | 0 | 1 | 48.93 |
| Associate's Degree | 0.117 | 0 | 1 | 50.60 |
| Bachelor's Degree | 0.222 | 0 | 1 | 48.34 |
| Master's Degree | 0.106 | 0 | 1 | 47.49 |
| Professional Degree | 0.033 | 0 | 1 | 44.4 |

[a] Professional Degree combines individuals with law degrees and doctorate degrees.

[b] Table was made using Outliers-Excluded Data

The fourth set of controls used were location controls. These variables were added to control for different effects caused by living in specific regions of South Carolina. There were 18 different areas specifying where an individual lives. Since location is potentially related to how much one commutes and how much one makes per year, these local fixed effects were added. Table 3.4 shows the descriptive statistics for the location controls for individual South Carolinians in 2018. The last set of controls used were transportation controls. These variables were added to control for different commuting methods to work for individuals. There were 7 different methods for commuting to work according to the ACS, and these methods were added due to their potential relationship to both commute time and labor earnings. Table 3.5 shows the descriptive statistics for the transportation controls for individual South Carolinians in 2018.

Table 3.4: Descriptive Statistics for Location Controls in 2018 (N = 14,921)

| **Variable** | **Proportion** | **Minimum** | **Maximum** | **Mean Daily Commute Time** |
|---|---|---|---|---|
| Berkely, Charleston, and Dorchester County | 0.168 | 0 | 1 | 54.49 |
| Greenville and Laurens County | 0.152 | 0 | 1 | 50.27 |
| Calhoun, Fairfield, Kershaw, and Richland County | 0.12 | 0 | 1 | 50.90 |
| Spartanburg County | 0.074 | 0 | 1 | 47.53 |
| Lexington and Saluda County | 0.067 | 0 | 1 | 47.17 |
| Horry County | 0.06 | 0 | 1 | 43.74 |
| Darlington and Florence County | 0.043 | 0 | 1 | 48.89 |
| York County | 0.04 | 0 | 1 | 43.15 |
| Oconee and Pickens County | 0.035 | 0 | 1 | 43.20 |
| Beaufort and Jasper County | 0.033 | 0 | 1 | 44.63 |
| Anderson County | 0.032 | 0 | 1 | 45.55 |
| Clarendon, Lee, Sumter, and Williamsburg County | 0.032 | 0 | 1 | 44.65 |
| Cherokee, Chester, Newberry, and Union County | 0.029 | 0 | 1 | 45.95 |
| Allendale, Bamberg, Barnwell, Colleton, Hampton, and Orangeburg County | 0.029 | 0 | 1 | 51.36 |
| Aiken and Edgefield County | 0.027 | 0 | 1 | 52.09 |
| Chesterfield, Lancaster, and Marlboro County | 0.021 | 0 | 1 | 44.97 |
| Abbeville, Greenwood, and McCormick County | 0.02 | 0 | 1 | 46.56 |
| Dillon, Georgetown, and Marlon County | 0.019 | 0 | 1 | 44.23 |

[a] Table was made using Outliers-Excluded Data

Table 3.5: Descriptive Statistics for Transportation Controls in 2018 (N = 14,921)

| Variable | Proportion | Minimum | Maximum | Mean Daily Commute Time |
|---|---|---|---|---|
| Car | 0.974 | 0 | 1 | 49.02 |
| Walk | 0.009 | 0 | 1 | 25.28 |
| Other Transport | 0.008 | 0 | 1 | 42.83 |
| Bus | 0.004 | 0 | 1 | 71.75 |
| Bike | 0.003 | 0 | 1 | 40.41 |
| Taxi | 0.001 | 0 | 1 | 31.89 |
| Motorcycle | 0.001 | 0 | 1 | 50.95 |

[a] Table was made using Outliers-Excluded Data

The data I used for my models in part B is a combination of the ACS Public Use Microdata Sample Estimates from 2018 and ACS Public Use Microdata Sample Estimates from 2021. This combined dataset aims to have observations before and during the events of COVID-19. A dummy variable named COVID is created to represent whether or not the individual was from the 2018 or 2021 dataset. On top of that, a variable called COVIDxEARN was created to represent the difference in the effect of labor earnings on commute time before and during COVID-19.

The ACS Public Use Microdata Sample Estimates from 2021 has yet to add some of the controls used in part A, so limited controls were accessible for use. All of the personal characteristic controls excluding race and child controls were still available in the data. The controls for education were also available to be used. The controls for transportation, location, and occupation were unavailable, so they were excluded from this part of the research. Earnings for both datasets are adjusted to be reported in 2018 dollars (Bureau, U. S. C., 2022). Table 3.6 shows the descriptive statistics of all variables used for the COVID-19 model in part B.

Table 3.6: Descriptive Statistics of Variables Used in the COVID-19 Model (N=30,379)

| | **2018 Data (N=16,224)** | | | **2021 Data (N= 14,155)** | | |
|---|---|---|---|---|---|---|
| **Variable** | **Mean** | **Std Dev** | **Mean Commute Time** | **Mean** | **Std Dev** | **Mean Daily Commute Time** |
| Commute Time (min/day) | 49.96 | 33.26 | | 48.98 | 33.32 | |
| Labor Earnings[b] ($/yr) | $49,623 | $34,557 | | $47,487 | $32,021 | |
| Age (years) | 42.67 | 12.84 | | 42.644 | 12.96 | |
| Male | 0.522 | 0.499 | 51.49 | 0.527 | 0.499 | 51.29 |
| Female | 0.478 | 0.499 | 48.29 | 0.473 | 0.499 | 46.58 |
| Native | 0.937 | 0.243 | 50.13 | 0.939 | 0.24 | 49.10 |
| Foreign | 0.063 | 0.243 | 47.34 | 0.061 | 0.24 | 48.57 |
| Married | 0.57 | 0.495 | 51.17 | 0.565 | 0.496 | 50.01 |
| Not Married | 0.43 | 0.495 | 48.35 | 0.435 | 0.496 | 47.84 |
| High School Dropout | 0.06 | 0.238 | 51.00 | 0.058 | 0.234 | 52.20 |
| High School Graduate | 0.251 | 0.434 | 49.56 | 0.265 | 0.441 | 49.46 |
| Some College | 0.212 | 0.409 | 50.37 | 0.205 | 0.404 | 49.04 |
| Associate's degree | 0.117 | 0.322 | 51.36 | 0.117 | 0.321 | 52.34 |
| Bachelor's Degree | 0.221 | 0.415 | 50.08 | 0.220 | 0.414 | 48.16 |
| Master's Degree | 0.105 | 0.307 | 48.85 | 0.098 | 0.298 | 46.18 |
| Professional Degree[a] | 0.033 | 0.178 | 46.17 | 0.037 | 0.188 | 44.09 |

[a] Professional Degree combines individuals with law degrees and doctorate degrees.

[b] Earnings for both years is measured in 2018 dollars.

[c] Table was made using Outliers-Excluded Data

## 4. MODELS

For part A, I use two different models to test the effects of earnings on commute time to work. The first model is a level-level standard regression model with controls. This model is represented as:

$$COMTIME_i = \beta_0 + \beta_1 EARN_i + \beta_2 \gamma + \beta_3 \lambda + \beta_4 \rho + \beta_5 \tau + \beta_6 \pi + \varepsilon \quad \textbf{(4.1)}$$

where $COMTIME_i$ is an individual's daily commute to work in minutes, EARN is an individual's labor earnings in dollars per year, $\gamma$ are variables specifying an individual's personal characteristics, $\lambda$ are variables specifying an individual's occupation, $\rho$ are variables specifying an individual's education level, $\tau$ are variables specifying an individual's location, $\pi$ are variables specifying an individual's transportation method to work, and $\varepsilon$ is the error term in the regression. The coefficient of interest in this model is $\beta_1$ as it shows the potential effect of higher labor earnings on commute time to work, ceteris paribus.

The second model is a log-log standard regression model with controls. This model is represented as:

$$\log(COMTIME_i) = \beta_0 + \beta_1 \log(EARN_i) + \beta_2 \gamma + \beta_3 \lambda + \beta_4 \rho + \beta_5 \tau + \beta_6 \pi + \varepsilon \quad \textbf{(4.2)}$$

where $\log(COMTIME_i)$ is the log of an individual's daily commute time to work in minutes, $\log(EARN_i)$ is the log of an individual's labor earnings in dollars per year, $\gamma$ are variables specifying an individual's personal characteristics, $\lambda$ are variables specifying an individual's occupation, $\rho$ are variables specifying an individual's education level, $\tau$ are variables specifying an individual's location, $\pi$ are variables specifying an individual's transportation method to work, and $\varepsilon$ is the error term in the regression. The coefficient of interest in this model is $\beta_1$ as it shows the potential effect of higher labor earnings on commute time as an elasticity, ceteris paribus.

In part B, I use a linear regression model to test how the effects of labor earnings on commute time vary before and during the events of the COVID-19 pandemic. The model is represented as:

$$COMTIME_i = \beta_0 + \beta_1 \, EARN_i + \beta_2 \, COVID + \beta_3 \, COVIDxEARN + \beta_4 \, AGE + \beta_5 \, AGESQ + \beta_6 \, MALE + \beta_7 \, NATIVE + \beta_8 \, MARRIED + \beta_9 \, \rho + \varepsilon \quad \textbf{(4.3)}$$

where $COMTIME_i$ is an individual's daily commute time to work in minutes, $EARN_i$ is an individual's labor earnings in dollars per year, COVID is a dummy variable denoting whether the individual is in the dataset before or during COVID-19, COVIDxEARN is the interaction between COVID and EARN, $\rho$ are variables specifying an individual's education level, and $\varepsilon$ is the error term in the regression. The coefficient of interest is $\beta_3$ as it shows how much the potential effect of labor earnings on commute time differs when an individual lives during the COVID-19 time period.

## 5. RESULTS AND DISCUSSION

In part A, the first model that I estimate is the level-level standard regression model with controls. The interpretations of the coefficients were based on the average workdays per year being 250. Given this model, a $1000 increase in labor earnings is associated with, on average, an 11.75-minute increase in commute time to work per year, ceteris paribus. The coefficient on labor earnings is significant at the one percent level (Table 5.1).

When comparing my results to those found by French et al. (2020), my results are much smaller in magnitude. French et al. find, when switching the direction of the relationship, that a $1000 increase in labor earnings per year is associated with, on average, a 105-minute increase in commute time per year, ceteris paribus. This estimate is almost ten times larger than what I found in my data. One reason for this is that French et al.'s observations are young individuals from the ages of 24-32 years old while I observe the whole working class (ages 18-65). Since younger individuals may experience less of the negative effects of commuting, they are willing to accept a higher compensating differential for longer commute times. French et al. also focus on individuals across the United States while I only focus on individuals from South Carolina. This may play a role in the comparison of results.

Table 5.1: Regression Output for Level-Level Standard Regression Model with Controls and Commute Time as the Dependent Variable

| Variable | Coeff | Std. Err[a] | Pr(>\|t\|) |
|---|---|---|---|
| EARN[b] | 0.047*** | 0.000001 | 0.000002 |
| AGE | 0.439*** | 0.16 | 0.006 |
| AGESQ | -0.0052*** | 0.0019 | 0.005 |
| MALE | 0.23 | 0.61 | 0.7 |
| NATIVE | 3.47*** | 1.14 | 0.002 |
| MARRIED | 0.64 | 0.59 | 0.28 |
| CHILD | -0.49 | 0.62 | 0.43 |
| Intercept | 15.88** | 7.35 | 0.03 |
| Controls | YES | | |
| | | | |
| Number of Observations | 14921 | | |
| Pr(>\|f\|) | 2.2e-16 | | |
| R-Squared | 0.0419 | | |
| *=significant at α=0.10; **=significant at α=0.05; ***=significant at α=0.01 | | | |

[a] Standard Errors Reported are Robust Standard Errors
[b] Labor Earnings in Thousands
[c] Outliers-Excluded Data Used in Model

The second model that I estimate is the log-log standard regression model with controls. Given this model, a 1% increase in labor earnings is associated with, on average, a 0.082% increase in daily commute time to work, ceteris paribus (Table 5.2). The coefficient is significant at the one percent significance level using both regular and robust standard errors.

Comparing these results to those found by Dargay et al., they are very similar. Dargay et al. find that a 1% increase in labor earnings is associated with, on average, a 0.08% increase in daily commute time to work. This result is very interesting due to the fact that Dargay et al. use panel data and minimize much of the endogeneity caused by reverse causality, yet they still find similar results. The population differences and lack of controls that Dargay et al. use may still lead to some omitted variable bias in their models, but the similarities show promising signs when observing the relationship between labor earnings and commute time.

Why would the relationship between labor earnings and daily commute time to work be positive? Higher labor earnings lead an individual to travel further to work as they are being compensated for their long commute times. Higher labor earnings are also associated with an individual demanding a higher quality-of-life. One with a higher labor earnings would want to live where there is less crime, better air quality, and cheaper housing. Since these characteristics are more common away from big cities, one would have to take a longer commute to get to work.

Table 5.2: Regression Output for Log-Log Standard Regression Model with Controls and Log Commute Time as the Dependent Variable

| Variable | Coeff | Std. Err[a] | Pr(>\|t\|) |
|---|---|---|---|
| Log(EARN) | 0.082*** | 0.011 | 4.82e-13 |
| AGE | 0.0056 | 0.0035 | 0.111 |
| AGESQ | -0.000068* | 0.00004 | 0.09 |
| MALE | -0.008 | 0.013 | 0.56 |
| NATIVE | 0.05* | 0.024 | 0.08 |
| MARRIED | 0.0014 | 0.025 | 0.91 |
| CHILD | -0.004 | 0.04 | 0.76 |
| Intercept | 2.34*** | 0.21 | 2.2e-16 |
| Controls | YES | | |
| | | | |
| Number of Observations | 14853 | | |
| Pr(>\|f\|) | 2.2e-16 | | |
| R-Squared | 0.045 | | |
| *=significant at α=0.10; **=significant at α=0.05; ***=significant at α=0.01 | | | |

[a] Standard Errors Reported are Robust Standard Errors
[b] Outliers-Excluded Data Used in Model

In part B, I estimate the COVID-19 model with controls. The interpretations of the coefficients were based on the average workdays per year being 250. Given this model, individuals during COVID-19 experienced a decrease in commute time to work of 7.5 minutes per year per $1000 of yearly labor earnings, ceteris paribus (Table 5.3). This coefficient is significant at the 1 percent significance level. Though this comparative inter-period association between labor earnings and commute time is negative, the net association of labor earnings on commute time to work is still positive and significant in both periods.

Why would the positive effect between labor earnings and commute time decrease during the events of COVID-19? The main argument is that there is much less traffic as many people moved to an online work environment. During the events of COVID-19, many individuals were forced to work from home to contain the spread of the COVID-19 pandemic. This means there are many people off the road, which leads to less traffic and fewer accidents. Since there is less traffic and fewer accidents, those still commuting to work enjoy a shorter commute time. Another argument is that individuals want to locate closer to hospitals to decrease the commute time to hospitals under extreme conditions. Since hospitals are located closer to cities, and a majority of jobs are also located in the city, these individuals would be moving closer to work indirectly. Lastly, many individuals altered their preferences for work and leisure during the events of COVID-19 (Alsharawy et al., 2020). Since COVID-19 was a very stressful time period for everyone, one may value leisure much more than before, therefore, an individual will work less than they did before the events of COVID-19 (assuming that working less is directly related to commuting less).

Table 5.3: Regression Output for COVID-19 Model with Limited Controls and Commute Time as the Dependent Variable

| Variable | Coeff | Std. Err[a] | Pr(>\|t\|) |
|---|---|---|---|
| EARN[b] | 0.079*** | 8.7e-6 | 2.2e-16 |
| Covid | 0.53 | 0.66 | 0.42 |
| COVIDxEARN[b] | -0.03*** | 0.000011 | 0.007 |
| Age | 0.9*** | 0.11 | 2.2e-16 |
| Age^2 | -0.01*** | 0.0013 | 1.54e-14 |
| Male | 2.84*** | 0.39 | 4.52e-13 |
| Naive | 2.23*** | 0.77 | 0.004 |
| Married | 0.71* | 0.41 | 0.087 |
| Intercept | 23.7*** | 2.32 | 2e-16 |
| Education Controls | YES | | |
| | | | |
| Number of Observations | 30734 | | |
| Pr(>\|f\|) | 2.2e-16 | | |
| R-Squared | 0.014 | | |
| *=significant at α=0.10; **=significant at α=0.05; ***=significant at α=0.01 | | | |

[a] Standard Errors Reported are Robust Standard Errors
[b] Labor Earnings in Thousands
[c] Outliers-Excluded Data Used in Model

## 6. LIMITATIONS

The first limitation of my study is that the assumption of a zero conditional mean breaks down. One important variable left out of the models is a variable denoting whether an individual lives in the city. Living in the city can potentially lead to higher wages but also potentially leads to lower commute times. These effects could potentially lead to a downward bias in my labor earnings coefficient. A downward bias could lead to a smaller coefficient, decreasing the potential association between labor earnings and commute time.

Another interesting variable to add would be a quality-of-life variable. Higher quality of life can be associated with higher wages but can also be associated with higher commute times, which would lead to a positive bias in the labor earnings coefficient. Overall, both of these variables potentially bias the labor earnings coefficient in all the models, and it would be beneficial to include city and quality-of-life variables in the models if they were available.

Potential endogeneity from either reverse causality or simultaneity bias may also exist in my models. It is uncertain whether commute time to work is the one that affects labor earnings or if labor earnings are the one that affects commute time to work. It could also be the case that these two factors are simultaneously determined, which would lead to simultaneity bias. The models in my paper showed that some linear relationship exists between the two variables, but the true direction of the relationship is not certain.

The last limitation of my model is that benefits-excluded labor earnings are used in the model instead of benefits-included labor earnings. When capturing the total benefit from working, or the true wage from working, we want to factor in all possible benefits that an individual gets from a job. These benefits are not strictly one's wage. Benefits such as health and dental insurance are not included in the labor earnings provided in my data. Benefits may act as a compensating wage differential that pushes an individual to choose a job, so not including them in the research may be limiting the association found in the results. There may be a problem here if benefits as a proportion of labor earnings vary with labor earnings.

## 7. **CONCLUSION**

What kind of large-scale negative effects could a positive association between labor earnings and commute time to work have on individuals in South Carolina? Those with longer commute times to work could be subjected to more stress associated with commuting (Halonen et al., 2019). On top of that, those with longer commute times could be less likely to have meaningful relationships with their co-workers (Halonen et al., 2019). These effects would imply that individuals are not accurately weighing the potential effects of commuting time. Both of these effects lead to worse mental health for individuals in South Carolina. What kind of large-scale negative effects could this have on firms in South Carolina? Individuals who have longer commute times could be more likely to miss work (Van Ommeren and Gutiérrez-i-Puigarnau, 2011). Lower work attendance leads to lower firm productivity and higher turnover of employees, leading to a less efficient firm.

From the results in part B, the potential positive association between labor earnings and commute time to work seems to be falling. What is a potentially negative effect for an individual when they live closer to their jobs? One may feel constrained or overwhelmed by their job leading to no separation between work and leisure (Halonen et al., 2019). This may be unhealthy for an individual's mental health.

Many limitations existed in my models which causes bias in the estimates. Omitted variable bias existed in my models as important variables such as city dummies and quality-of-life were not available in the data. Many of the controls in part A were not available in part B of my studies, which adds more uncertainty. Reverse causality and simultaneity bias also plagued my models in both parts A and B due to the fact that the true direction of causality cannot be determined without the use of an instrument, which was not available in this dataset. Not having benefits-included labor earnings as the independent variables for my models in both parts A and B also causes issues as benefits-excluded labor earnings do not solely represent the true benefit of working a job.

Any future endeavors into this topic would have to address these limitations listed above. An instrumental variable can be used to extract the true effect of labor earnings on commute time. A natural or quasi-natural experiment can also be used to guarantee the direction of causality. One paper by Mulalic et al. conducts a quasi-natural experiment using firm relocation in Denmark, but firm information is difficult to come by, especially in the United States. Mulalic et al. find that increasing the commuting distance implies a moderate wage increase three years after the firm relocation (0.15%) but find no significant effect of an increase in wages directly following the firm relocation. It would be very interesting to see whether these results found in Mulalic et al. would be similar if a study was run in the United States. Overall, limitations exist within this study, but with broader data and more sophisticated techniques, there is potential to find a reliable estimate for the effects of labor earnings on commute time to work in South Carolina.

## WORKS CITED

Alsharawy, A., Ball, S. B., Smith, A., & Spoon, R. (2020). Fear of COVID-19 changes economic preferences: Evidence from a repeated cross-sectional Mturk Survey. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3750351

Bureau, U. S. C. (2022, December 15). Public use microdata sample (PUMS). Census.gov. Retrieved April 7, 2023, from https://www.census.gov/programs-surveys/acs/microdata.html

Bureau, U. S. C. (2022, October 25). Sample ACS & PRCS forms and instructions. Census.gov. Retrieved April 7, 2023, from https://www.census.gov/programs-surveys/acs/about/forms-and-instructions.2018.html#list-tab-9466845

Bureau, U. S. C. (n.d.). Explore Census Data. Retrieved April 7, 2023, from https://data.census.gov/

Carvajal, M. J., & Popovici, I. (2016). Interaction of gender and age in Pharmacists' labour outcomes. Journal of Pharmaceutical Health Services Research, 7(1), 23–29. https://doi.org/10.1111/jphs.12118

Dargay, J. (n.d.). The effect of income on commuting time using panel data. The effect of income on commuting time using panel data - researchgate. Retrieved April 13, 2023, from https://www.researchgate.net/publication/228454905_The_Effect_of_Income_on_Commuting_Time_Using_Panel_Data

French, M. T., Popovici, I., & Timming, A. R. (2020). Analysing the effect of commuting time on earnings among young adults. *Applied Economics*, *52*(48), 5282–5297. https://doi.org/10.1080/00036846.2020.1761537

Halonen, J. I., Pulakka, A., Vahtera, J., Pentti, J., Laström, H., Stenholm, S., & Hanson, L. M. (2019). Commuting time to work and behaviour-related health: A fixed-effect analysis. Occupational and Environmental Medicine, 77(2), 77–83. https://doi.org/10.1136/oemed-2019-106173

Hazans, M. (2004). Does commuting reduce wage disparities? *Growth and Change*, *35*, 360–390. https://doi.org/10.2139/ssrn.748247

Holzer, H. J. (1991). The spatial mismatch hypothesis: What has the evidence shown? Urban Studies, 28(1), 105–122. https://doi.org/10.1080/00420989120080071

Manning, A. (2003). The real thin theory: Monopsony in modern Labour Markets. Labour Economics, 10(2), 105–131. https://doi.org/10.1016/s0927-5371(03)00018-6

Mulalic, I., Van Ommeren, J. N., & Pilegaard, N. (2013). Wages and commuting: Quasi-natural experiments' evidence from firms that relocate. *The Economic Journal*, *124*(579), 1086–1105. https://doi.org/10.1111/ecoj.12074

Pums accuracy of the data 2010 - census.gov. (n.d.). Retrieved April 7, 2023, from https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/ACS_Accuracy_of_Data_2010.pdf

van Ommeren, J. N., & Gutiérrez-i-Puigarnau, E. (2011). Are workers with a long commute less productive? an empirical analysis of absenteeism. Regional Science and Urban Economics, 41(1), 1–8. https://doi.org/10.1016/j.regsciurbeco.2010.07.005