**Title**
Essays On U.S. Public Transportation Development



**Author**
McQuilkin, Alec Joseph

**Publication Date**
2021

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

# Essays On U.S. Public Transportation Development

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Alec J. McQuilkin

Committee in charge:

Professor Heather Royer, Chair
Professor Clément De Chaisemartin
Professor Andrew Plantinga

June 2021

The Dissertation of Alec J. McQuilkin is approved.

Professor Clément De Chaisemartin

Professor Andrew Plantinga

Professor Heather Royer, Committee Chair

June 2021

Essays On U.S. Public Transportation Development



by

Alec J. McQuilkin

To my parents Sue and Steve.

## Acknowledgements

The completion of this dissertation is a result of an extraordinary amount of hard work and dedication on my part. However, it would not have been possible without the guidance and support from the many individuals who helped me along the way. First, I would like to thank my doctoral committee Heather Royer, Clément De Chaisemartin, and Andrew Plantinga. Their input was invaluable to the development of each research chapter. I am also grateful for the help of the many UC Santa Barbara and CU Denver faculty members including Doug Steigerwald, Peter Kuhn, Hani Mansour, and Peter Rupert that helped me to focus and refine my research skills throughout my graduate studies.

I am thankful for the support of my graduate peers including Lucas Reddinger, Jacob Gellman, Guangli Zhang, Charlie Nusbaum, Hakan Özyılmaz, Emmy Robertson, and Jeff Cross along with many others. They played a vital role in helping me to overcome the many challenges that I faced while finding my way as a researcher, all while providing thoughtful feedback during the development of my dissertation.

Finally, I owe an enormous amount of gratitude to my family; my parents Sue Brozowski and Steve McQuilkin, my grandmother Shirley Brozowski, Kathy McQuilkin and my sister Molly McQuilkin. I would be lost without their support.

# Curriculum Vitæ

Alec J. McQuilkin

## Education

| | |
|---|---|
| 2021 | Ph.D. in Economics, University of California, Santa Barbara |
| 2016 | M.A. in Economics, University of California, Santa Barbara |
| 2015 | M.A. in Economics, University of Colorado, Denver |
| 2015 | M.S. in Applied Mathematics, University of Colorado, Denver |
| 2009 | B.S. in Applied Mathematics, University of Colorado, Boulder |

## Research Interests

Urban Economics, Labor Economics, Econometrics, Health Economics

## Research Papers

"Public Rail Transit and Commuting: Evidence From 30 U.S. Metropolitan Areas"
R&R *Journal of Urban Economics*

"Rail Development, Job Access, and Employment"
Job Market Paper

"Real Estate Responses To Rail Development"
With Jacob Gellman

## Research Positions

Research Assistant to Javier Birchenall — May 2020–Aug. 2020
Department of Economics
University of California, Santa Barbara

Research Assistant to Heather Royer — Sep. 2016–May 2019
Department of Economics
University of California, Santa Barbara

Research Assistant to Genevieve Pham-Kanter — Apr. 2014–Jan. 2015
School of Public Health
University of Colorado, Denver

## Teaching Positions

Head Teaching Assistant, UC Santa Barbara — 2018–2020
Introductory Econometrics, Intermediate Microeconomics

Teaching Assistant, UC Santa Barbara — 2016–2021
Intermediate Econometrics, Intermediate Microeconomics

| | |
|---|---|
| Instructor, UC Santa Barbara Extension Program<br>Personal Finance and Entrepreneurship | 2017 |
| Teaching Assistant, CU Denver<br>Introductory Econometrics, Intermediate Microeconomics/Macroeconomics | 2012–2015 |

### Academic Presentations

| | |
|---|---|
| U.S. Government Accountability Office, Center for Economics, Remote | Feb. 2021 |
| Aalto University, Department of Economics, Remote | Jan. 2021 |
| U.S. Census, Center for Economic Studies, Remote | Jan. 2021 |
| UCSB Applied Microeconomics Seminar, Santa Barbara, CA | Oct. 2020 |
| Race and Inequality: A Collaborative UCPop Event, Remote | Oct. 2020 |
| UCSB Applied Microeconomics Seminar, Santa Barbara, CA | Nov. 2019 |
| American Statistical Association CO-WY Chapter Meeting, Boulder, CO | Apr. 2015 |

### Grants and Fellowships

| | |
|---|---|
| Research Fellowship, UCSB Economics Department | May 2020–Aug. 2020 |
| Data Grant, UCSB Economics Department | Nov. 2019 |
| Fellowship Stipend, UCSB Economics Department | Sep. 2015–Aug. 2016 |

### Affiliations

| | |
|---|---|
| Broom Center for Demography, UC Santa Barbara | Jun. 2017–Jun. 2021 |

### References

Heather Royer
Associate Professor
Department of Economics
UC Santa Barbara
North Hall 2020
Santa Barbara, CA 93106
royer@econ.ucsb.edu

Clément de Chaisemartin
Assistant Professor
Department of Economics
UC Santa Barbara
North Hall 2038
Santa Barbara, CA 93106
clementdechaisemartin@ucsb.edu

Andrew Plantinga
Professor
Bren School of Environmental Science and Management
UC Santa Barbara
Bren Hall 3424
Santa Barbara, CA 93106
plantinga@bren.ucsb.edu

## Permissions and Attributions

1. The content of Chapter 3 is the result of a collaboration with Jacob Gellman. It is reproduced here with his consent.

**Abstract**

Essays On U.S. Public Transportation Development

by

Alec J. McQuilkin

In Chapter 1, I study the impacts of rail transit expansions in 30 U.S. metropolitan areas by comparing changes in commuting patterns for neighborhoods located near newly opened stations to similar neighborhoods that remained more distant. The effects depend heavily on socioeconomic status with below-median poverty neighborhoods experiencing a meaningful increase (1.97 percentage points) in the proportion of public transit commuters. Higher poverty neighborhoods see no change in public transit commuting as increases in rail usage are completely offset by reductions in bus usage. Regardless of poverty rate, these changes in commute mode are not accompanied by commute time savings.

In Chapter 2 I investigate the extent to which public, rail transit can improve employment outcomes of targeted populations by enhancing job access to urban residents. Using a novel, person level panel data set, I exploit changes in rail access resulting from the large number of stations opened between 1968 and 2017 to compare changes in outcomes for workers located near new stations relative to those that remained more distant. I find that the introduction of a rail station within 2 miles of a household leads to a 2.73 percentage point increase in the probability of employment among household heads. The effects appear immediately following a station's opening, but continue to grow in magnitude until stabilizing at an elevated level three years afterward. The results provide evidence that rail infrastructure development can be an effective policy tool in improving the economic outcomes of urban residents.

In a joint work with Jacob Gellman, Chapter 3 explores the impacts of rail infrastructure on residential property values using a comprehensive data set that incorporates all urban rail development throughout the U.S. from 1996 to 2017 to investigate the differential effects of rail development on home values. Prior research documents a positive causal relationship of public rail transportation infrastructure on real estate values. However, the extent of the effect appears to differ considerably both within and across cities based on location, system, and property characteristics. We find similar positive effects of rail development on home values while highlighting this inter and intracity heterogeneity.

# Contents

# Chapter 1

# Public Rail Transit and Commuting: Evidence From 30 U.S. Metropolitan Areas

## 1.1 Introduction

Population growth in urban settings places a strain on local transportation infrastructure. In the United States, where residential and employment location steadily decentralized throughout the latter half of the 20th century (Baum-Snow (2010)), this strain became particularly acute. Metropolitan areas across the U.S. have increasingly turned to public rail transit to alleviate such burden. However, the high cost and permanence of these capital investments have made rail development a contentious policy issue.[1]

[1]Light rail construction costs (excluding rolling stock expenses) range from approximately $20 million (2005 dollars) per mile in cities such as Baltimore, Denver, Sacramento, and St. Louis (Baum-Snow, Kahn and Voith (2005)) to as much as $179 million per mile in Seattle (Central Puget Sound Regional Transit Authority (2006)). Heavy rail and subway construction can be far more expensive with Phase 1 of the New York City Second Avenue Subway costing approximately $2.2 billion per mile (Rosenthal (2017)).

The fundamental objective of rail investment is to reduce the cost of public transit travel. The high speed and right-of-way advantages make rail an efficient option for movement between connected locations. An increase in the number of stations shortens the journey to and from transit access points. Additionally, rail transit may offer several other advantages such as improved comfort, lower stress, the elimination of parking and driving costs, and the mitigation of personal environmental damage. Accordingly, proponents argue that the introduction or expansion of rail transit systems can entice drivers out of their vehicles reducing the negative externalities associated with automobile travel (e.g., congestion, pollution).

Proponents often suggest that expanded rail infrastructure can be particularly valuable to the urban poor who struggle to afford private vehicles (Garrett (2004)). As employment opportunities decentralized across the U.S., low-income and often minority individuals remained concentrated in inner-city locations unable to access an increasing proportion of jobs (Kain (1968)). By enhancing the mobility of these transit dependent individuals, rail development may act to improve the geographic scope for employment reducing this problem of spatial mismatch (Ihlanfeldt and Sjoquist (1998); Pang (2017); Turner (2019)). However, while bus public transit users are indeed disproportionately poor, the same cannot be said of rail users (U.S. Census).[2] Moreover, expenditure for rail capital investment and operation may act to crowd out spending on bus transit, leading to reductions in service. Empirical evidence can help to shed light on the these proposed impacts.

In this study, I investigate how local commuting outcomes respond to public rail transit infrastructure investment. Using a novel data set that includes the timing and location for all public, urban rail stations opened throughout the U.S. between 1970 and

[2]The median income among rail users is actually higher relative to that of the overall U.S. population.

2010, I exploit variation in rail transit access within metropolitan areas across time.[3] I compare changes in commute mode for census tracts located in close proximity of newly opened stations to tracts that remained more distant.[4] I also explore changes to the tract level median commute time to evaluate the potential for time savings resulting from improved rail access. This staggered adoption difference-in-differences framework serves to estimate the neighborhood level causal effect of increased rail transit access on usage and commute time.

The results suggest a small overall increase in the proportion commuting by public transit (0.661 percentage points (p.p.)) in neighborhoods located near new stations. There is no associated change in automobile commuting. While the proportion of rail commuters increases by 1.75 p.p., 59.4 percent of this is offset by reductions in bus usage. The increase in rail commuting is stable, regardless of location relative to the city central business district (CBD). Decreases in bus usage are confined to neighborhoods located closer to the CBD.

Special attention is paid to the differential impacts by neighborhood socioeconomic status as measured by the census tract poverty rate. I find that the impacts on commute mode are strongest in below-median poverty neighborhoods which experience a 1.97 p.p. increase in public transit usage following treatment. This change is driven by a large increase in rail commuting (2.22-3.56 p.p) with minimal impacts on bus usage. These well-off neighborhoods experience a decrease in automobile commuting that is roughly equivalent to the increase in public transit commuting. For above-median poverty neighborhoods, there is little evidence of an overall change in public transit usage. The proportion of rail commuters increases by 1.11 p.p., but reductions in bus usage completely offset this change. Regardless of poverty rate, I find no evidence of commute

[3]This includes all light rail, heavy rail, subway, and commuter rail systems.
[4]Neighborhoods and census tracts are used interchangeably throughout this paper.

time savings.

The validity of these results in assessing the causal impacts of rail access is predicated on two assumptions. First, changes in commuting outcomes across time for tracts located near newly opened rail stations must be comparable to those of tracts that remained more distant. I evaluate this assumption by examining differences in outcomes over time showing no evidence of differential trends prior to treatment. Additionally, I apply two econometric procedures which modify the sample so that comparisons are made between groups that are more similar based on pre-treatment observable characteristics. The estimated effects are robust to each approach.

Second, my main specification assumes that commuting outcomes of tracts farther than 1 mile from a new rail station are unaffected by it's introduction. As the geographic extent of the impact is ex ante unclear, I estimate an alternative specification that allows changes in commuting outcomes to vary non-linearly in distance from the station. I show that much of the effect is restricted to tracts located within 1 mile providing support for my preferred specification.

This study provides three main contributions to the literature concerning public transportation infrastructure development. First, by including all U.S. urban rail expansions between 1970 and 2010, I build on several studies that have focused on either a single expansion (Heilmann (2018); Severen (2018); Tsivanidis (2018)) or a subset of those included in this study (Baum-Snow and Kahn (2000); Baum-Snow, Kahn and Voith (2005); Glaeser, Kahn and Rappaport (2008); Kahn (2007)). Using information for each of the 1,431 stations opened during this time period allows me to construct a comparatively comprehensive data set of U.S. rail infrastructure. The estimates measure the average national effects which are less influenced by city-specific factors. Additionally, I explore how the impacts differ across cities, including many recent expansions that were not included in the similar Baum-Snow, Kahn and Voith (2005).[5] For this reason, the estimates may be more pertinent to policy discussions regarding future expansions.

The extent of the data also provides an excellent opportunity to investigate the heterogeneous effects of rail infrastructure by neighborhood socioeconomic status with greater precision than earlier work. Previous studies have documented that relatively well-off neighborhoods experience the largest gains to the introduction of large transit infrastructure projects (Baum-Snow and Kahn (2000); Heilmann (2018); Tsivanidis (2018)). Similarly, I show that commuting outcomes are far more responsive to the introduction of rail infrastructure in lower poverty neighborhoods. I also demonstrate that these differential impacts are not an artifact of the initial spatial distribution of income within a given metropolitan area.

Second, similar studies often rely on a measure of overall public transportation usage to evaluating the impacts of rail infrastructure on commute mode (Baum-Snow and Kahn (2000); Baum-Snow, Kahn and Voith (2005)). Baum-Snow, Kahn and Voith (2005) develops a model suggesting that much of the inner-city increase in rail usage comes from former bus users, while those in the outer suburbs come from former car commuters. Consistent with this result, they document larger changes in overall public transit usage in the suburbs relative to tracts near the CBD. I find a similar result, but also show that the underlying changes in rail and bus specific transit usage are consistent with the model's predictions. In addition, I supplement these results with ridership, expense, and asset information, by mode (rail versus bus), for each U.S. transit agency to investigate whether the observed reductions in bus usage reflect a reduction in bus supply. If transit agencies allocate resources to rail systems at the expense of their bus service, then this

[5] This study includes openings in Austin, Charlotte, Houston, Minneapolis, Nashville, Oceanside/Escondido, Phoenix, Salt Lake City, and Seattle. In addition, many recent expansions in other metropolitan areas included had not been incorporated into Baum-Snow, Kahn and Voith (2005). See Section 1.2.

reduction in supply could hurt those initially relying on bus transportation. I find no evidence of this crowding out hypothesis.

Third, I investigate the impacts of the introduction of a rail station on an important outcome that has not generally been addressed in similar studies; the time spent commuting. I document limited evidence that treatment leads to any reduction in commute time regardless of poverty rate or location relative to the CBD. This lack of evidence regarding time savings for commuters is certainly relevant to transit policy.

The remainder of this paper is organized as follows. Section 1.2 describes the data used in the analysis. Section 1.3 details the empirical methods. Section 1.4 presents the main results that describe the effects of rail transit on commute mode. Section 1.5 provides evidence to support the empirical assumptions of the research design. Section 1.6 examines intercity heterogeneity of the effects along with city level aggregate effects. Section 1.7 discusses migration as a mechanism and presents the metropolitan area aggregate estimates. Section 1.8 presents the impacts on commute time. Section 1.9 discusses the conclusions.

## 1.2 Data

To estimate the impacts of increased rail transit access on local commuting patterns, I use data derived from the 1970, 1980, 1990, and 2000 Decennial Census Summary Files along with the 2006-2010 American Community Survey 5-Year Estimate Summary Files. These data provide population and housing characteristics aggregated to the census tract level.[6] As the geographic boundaries of census tracts change over time, I employ two resources that rely on detailed spatial information for the location of housing units within census blocks to construct variables for each year that are consistent with 2010

[6]Census tracts consist of approximately 4,000 inhabitants and are "designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions" (U.S. Census).

census tract boundaries.

First, the GeoLytics Neighborhood Change Database 1970–2010 (NCDB) contains all of the demographic and housing characteristics used in the study along with the proportion of commuters whose primary mode of travel is by private automobile, public transit, and other commute mode.[7] Second, I use crosswalks provided by the Longitudinal Tract Database (LTDB) to construct more specific measures of the proportion of commuters by mode; public transit by rail versus bus. The LTDB also allows construction of the census tract median commute time.[8]

The NCDB contains a balanced panel of 73,057 census tracts for each decennial year between 1970 and 2010 (365,285 tract-by-year observations). Census tracts that remain farther than 4 miles from a rail station throughout the entirety of the study period are excluded along with tracts that were already within 1 mile of a rail station by 1970 leaving 14,173 census tracts (70,865 tract-by-year observations). All tract-by-year observations with missing or erroneous values for the outcomes of interest are dropped leaving an unbalanced panel of 70,274 tract-by-year observations (approximately 14,050 tracts per year) with which much of the empirical analysis is conducted.

The data set constructed with the LTDB contains an unbalanced panel of 320,278 tract-by-year observations. Restricting to tracts that are ever within 4 miles of a rail station and eliminating those that were within 1 mile of a rail station by 1970 results in 94,600 observations. Eliminating those with missing and erroneous commute mode (including all 1980 observations) results in 56,184 tract-by-year observations (approximately 14,050 tracts per year). Eliminating observations with missing commute time (including all 1970 observations) results in 56,106 tract-by-year observations (approximately 14,025

[7]See Table B.1 for full description of variables.

[8]LTDB found at US2010 Project http://www.s4.brown.edu/us2010/Researcher/Bridging.htm. Note that the 1980 census does not differentiate public transit usage by type, so these variables are unavailable for 1980. The 1970 census does not contain information on commute time, so median commute time is unavailable for 1970.

tracts per year).[9]

Three sources of spatial data are used to designate treatment status and to construct a measure of the distance from each tract to the city center. First, the exact longitude and latitude along with the date of opening for each station of all public, urban rail systems in the U.S. opened between 1970 and 2010 are scraped from Wikipedia which collects and compiles information gathered from local transit authority resources. The geographic information for each station is compared with General Transit Feed Specification (GTFS) files provided by each transit authority to ensure accuracy. Second, longitude and latitude coordinates for each metropolitan area's central business district (CBD) come from Holian and Kahn (2015). Finally, 2010 census tract boundary geographies are defined using the 2010 TIGER/Line Shapefiles. I calculate the spatial centroid for each tract and use this point to determine the distance to the nearest open rail station and CBD.

Table 1.1 lists counts for the number of rail stations, along with the earliest and most recent stations, opened between 1970 and 2010 for each metropolitan area.[10] In total, 1,431 stations across 30 metropolitan areas were opened during this time frame. The areas are scattered geographically across the U.S. Some of the oldest stations during the study period were added to the metro areas of Boston, Chicago, Cleveland, New York City, and Philadelphia, which all had well-established rail systems prior to 1970. Other early stations were opened in cities with less developed rail networks such as Washington DC and San Francisco. The most recent systems have opened in Albuquerque, Austin, Charlotte, Nashville, and Phoenix. The metro areas with the largest rail transit systems, as measured by the total number of stations as of 2010, are New York, Chicago, Philadelphia, and Boston. However, many of these stations were opened prior to 1970.

[9] I refer to all census tracts that were within 4 miles of a rail station by 2010, but farther than 1 mile from a rail station in 1970 as the full sample. Due to missing values, the number of observations associated with each variable may vary slightly throughout the empirical analysis.

[10] Stations are allocated to metropolitan areas (also refereed to as cities) using the 2010 Census designated Urban Areas. Some consolidation of urban area boundaries results in 30 distinct metropolitan areas.

The metro areas with the most new stations throughout the study period are Washington DC, the Greater Bay Area, and Los Angeles.

In the empirical analysis, census tracts are defined as treated if their spatial centroid falls within 1 mile of an open rail station. Figure 1.1 shows the cumulative number of treated tracts by year for each metropolitan area throughout the study period. The panels are arranged in decreasing order such that the metro areas with the largest systems (determined by the number of treated tracts as of 2010) can be found in Panel (a) and the smallest systems are found in Panel (d). The variation in timing and the number of treated tracts is diverse across the different metropolitan areas. In general, rail systems tend to experience an increase in the number of stations every 5-15 years. These increases represent new rail lines, or expansions on older lines, which provide access to neighborhoods that were initially without or had limited access in close proximity. Many cities have expanded their systems several times, while some have yet to expand their systems since they were initially opened.

Between 1970 and 2010, 3,000 census tracts became closer than 1 mile of a rail station. Los Angeles, the Bay Area, and Washington DC represent the largest share with 576, 351, and 309 tracts becoming treated, respectively. For Atlanta, Austin, Buffalo, Nashville, Oceanside/Escondido, San Diego, and St. Louis, all tracts that were within 1 mile by 2017 were treated between 1970 and 2010. For Albuquerque/Santa Fe, Baltimore, Charlotte, Dallas, Denver, the Bay Area, Los Angeles, Miami, Minneapolis, Phoenix, Pittsburgh, Portland, Sacramento, Salt Lake City, Seattle, and Washington DC, more than 50 percent of tracts treated by 2017 where treated during the study period.

Table 1.2 describes demographic, housing, and commuting characteristics for each census tract in 1970 (prior to any treatment). Column (1) presents the averages for each variable using the full sample. Columns (2) and (3) present the averages for tracts that were 1-4 miles (control group) and less than 1 mile (treatment group), respectively,

Table 1.1: Rail Systems By Metropolitan Area

| | Stations | First Year | Last Year |
|---|---|---|---|
| Albuquerque | 12 | 2006 | 2010 |
| Atlanta | 38 | 1979 | 2000 |
| Austin | 9 | 2010 | 2010 |
| Baltimore | 51 | 1980 | 1998 |
| Boston | 81 | 1971 | 2010 |
| Buffalo | 14 | 1984 | 1986 |
| Charlotte | 15 | 2007 | 2007 |
| Chicago | 68 | 1971 | 2006 |
| Cleveland | 7 | 1971 | 1999 |
| DC | 120 | 1976 | 2004 |
| Dallas | 66 | 1996 | 2010 |
| Denver | 35 | 1994 | 2006 |
| Houston | 16 | 2004 | 2004 |
| Los Angeles | 124 | 1990 | 2009 |
| Miami | 34 | 1984 | 2003 |
| Minneapolis | 26 | 2004 | 2010 |
| Nashville | 6 | 2006 | 2006 |
| New Haven | 10 | 1990 | 2002 |
| New York | 87 | 1971 | 2009 |
| Philadelphia | 39 | 1974 | 2004 |
| Phoenix | 33 | 2008 | 2008 |
| Pittsburgh | 50 | 1984 | 2004 |
| Portland | 90 | 1986 | 2010 |
| Sacramento | 53 | 1987 | 2007 |
| Salt Lake City | 36 | 1999 | 2008 |
| San Diego | 77 | 1981 | 2008 |
| San Francisco | 95 | 1972 | 2007 |
| San Jose | 74 | 1987 | 2005 |
| Seattle | 28 | 2000 | 2009 |
| St. Louis | 37 | 1993 | 2006 |
| Total | 1431 | 1971 | 2010 |

Column (1) describes the number of stations opened between 1970 and 2010 for each metropolitan area. Columns (2) and (3) display the year of the earliest and most recent expansion, respectively, during this time period for each metropolitan area. The last row presents the information for the total of all metropolitan areas.

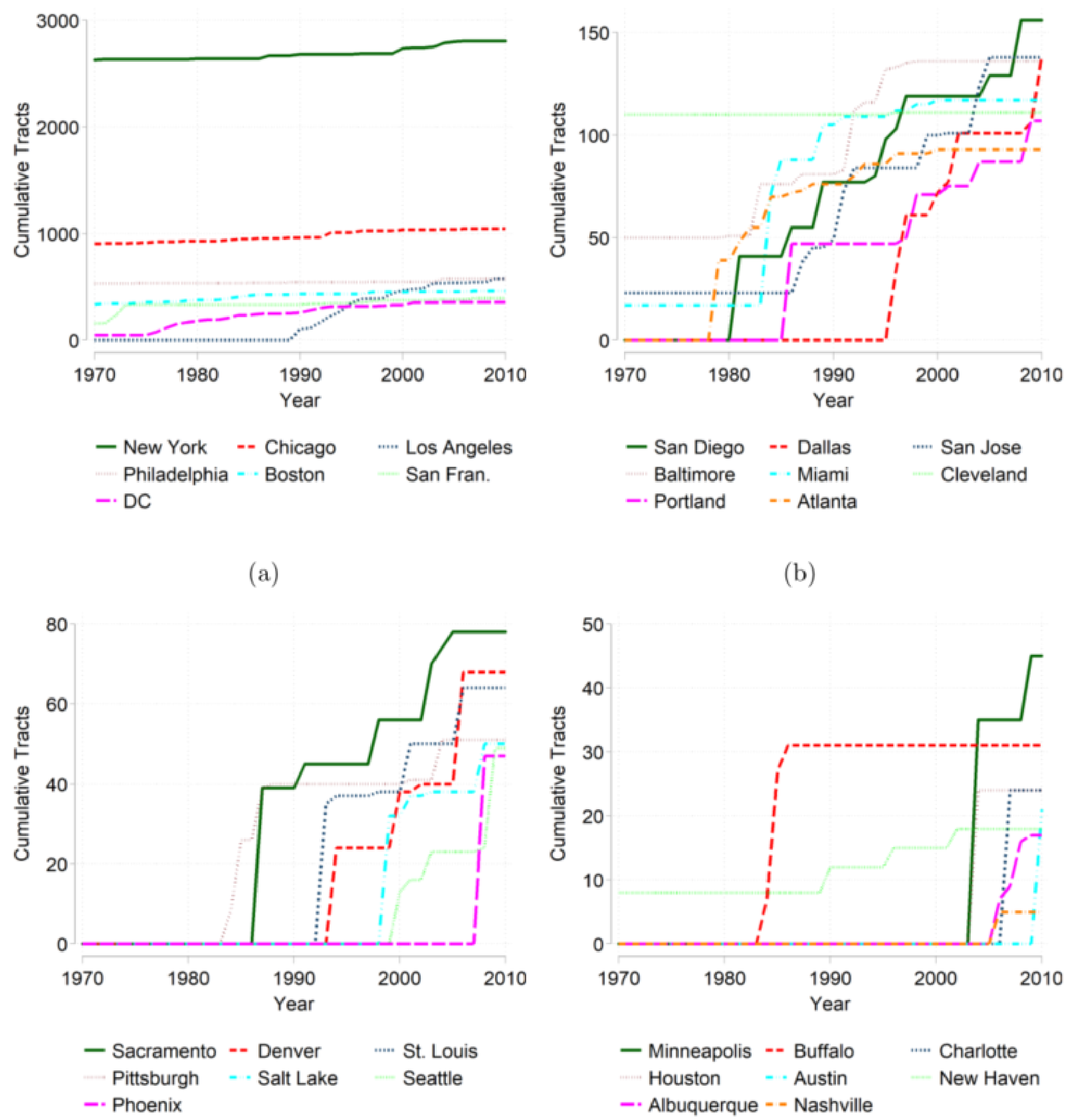



Figure 1.1: Number of Treated Tracts by Urban Area

Cumulative number of census tracts by 2010 Census designated Urban Area that are located within 1 mile of an open public rail station.

from a rail station by 2010. Column (4) provides the difference in means between the treatment and control group.

Census tract demographic and housing characteristics differ based on eventual treatment status. In particular, tracts that eventually become treated are more populous, with a higher population density and more households. They have a higher proportion of individuals with less than a high school education, a lower median family income, and a higher poverty rate relative to tracts that remain untreated. Fifteen percent of individuals in eventually treated tracts already use public transportation as their primary means of commuting relative to nine percent in the control group.[11] Housing rental prices are thirteen percent lower in the treatment group relative to the control group. Households are also more likely to be renting as opposed to owning their place of residence. There are approximately 11,173 and 3,000 tracts in the control and treatment group, respectively.

While the differences suggest a lack of comparability between the treatment and control group, the difference-in-differences research design uses the changes in outcomes across time to identify the causal effect. In addition, I employ several empirical procedures to provide evidence in support of my empirical specification.

To supplement my primary analysis, transit agency level information is collected from the National Transit Database maintained by the U.S. Federal Transit Administration. The data include operating and capital expense, ridership (unlinked passenger trips and passenger-miles), capacity (passengers at maximum service), service (passenger miles, directional route), and asset information, by transit mode (rail, bus) by year from 1993–1992.[12]. Some are unavailable for some of the years for some transit agencies. This data is used to evaluate four aspects of the main results; the technical feasibility of the observed results, whether the systems are operating at capacity, whether the ridership numbers

[11] If an individual uses multiple modes to commute, the primary commute mode is the mode used to cover the most distance in a given week.

[12] Some of these variables are available in 1992

Table 1.2: 1970 Summary Statistics by Eventual Treatment Status

| | (1) Total | (2) 1-4 miles | (3) $\leq$ 1 mile | (4) Diff |
|---|---|---|---|---|
| Population | 3182.17 | 3170.62 | 3224.68 | 54.06 |
| Pop. Density | 6732.44 | 5764.56 | 10297.84 | 4533.28*** |
| Dist. From CBD | 13.35 | 14.54 | 8.93 | -5.61*** |
| Prop. Public Transit | 0.10 | 0.09 | 0.15 | 0.06*** |
| Prop. Rail | 0.02 | 0.02 | 0.01 | -0.01*** |
| Prop. Bus | 0.07 | 0.06 | 0.13 | 0.07*** |
| Prop. Private Auto | 0.82 | 0.85 | 0.75 | -0.10*** |
| Prop. Other Commute | 0.08 | 0.07 | 0.11 | 0.04*** |
| Med. Commute Time† | 23.43 | 23.52 | 23.07 | -0.45*** |
| Prop. Male | 0.49 | 0.49 | 0.48 | -0.00*** |
| Prop. Black | 0.09 | 0.08 | 0.16 | 0.08*** |
| Prop. Native | 0.87 | 0.88 | 0.83 | -0.05*** |
| Prop. Same House | 0.47 | 0.48 | 0.44 | -0.04*** |
| Prop. Less HS | 0.41 | 0.39 | 0.46 | 0.07*** |
| Prop. HS | 0.33 | 0.33 | 0.30 | -0.04*** |
| Prop. Some College | 0.13 | 0.14 | 0.12 | -0.01*** |
| Prop. Bach./Grad. | 0.14 | 0.14 | 0.12 | -0.02*** |
| Unemp. Rate | 0.05 | 0.04 | 0.06 | 0.01*** |
| Prop. Poverty | 0.09 | 0.08 | 0.14 | 0.06*** |
| Family Income | 72601.36 | 75022.42 | 63700.38 | -11322.04*** |
| Prop. 0-17 | 0.34 | 0.35 | 0.30 | -0.05*** |
| Prop. 18-24 | 0.11 | 0.11 | 0.13 | 0.02*** |
| Prop. 25-29 | 0.07 | 0.07 | 0.08 | 0.01*** |
| Prop. 30-34 | 0.06 | 0.06 | 0.06 | -0.00*** |
| Prop. 35-44 | 0.12 | 0.12 | 0.11 | -0.01*** |
| Prop. 45-64 | 0.21 | 0.20 | 0.21 | 0.01*** |
| Prop. 65+ | 0.09 | 0.09 | 0.11 | 0.02*** |
| Num. Households | 1013.58 | 983.34 | 1125.01 | 141.67*** |
| Prop. Non-Family HHs | 0.19 | 0.16 | 0.28 | 0.12*** |
| Housing Units | 1058.88 | 1023.48 | 1189.29 | 165.81*** |
| Median Rent | 806.69 | 830.37 | 721.55 | -108.82*** |
| Prop. Occupied | 0.95 | 0.96 | 0.95 | -0.01*** |
| Prop. Renter Occupied | 0.37 | 0.33 | 0.52 | 0.19*** |
| Observations | 14173 | 11173 | 3000 | 14173 |

†Median commute time from 1980.
Median rent and family income in 2010 dollars.

agree with the results, and whether rail investment is acting to cannibalize bus service.

## 1.3 Methods

The goal of this study is to identify the causal effect of increased rail transit access on local commuting patterns. Using the opening of every public, urban rail station across the U.S. between 1970 and 2010, I exploit variation in rail access within census tracts over time. I compare changes in outcomes between tracts that became within 1 mile of an open rail station to comparable tracts that remained farther than 1 mile of a rail station. In the main specification, I restrict the sample to tracts that are ever within 4 miles of a rail station. This partially mitigates concerns that treated and untreated tracts are fundamentally different as tracts within 4 miles are likely a more comparable group than those outside of 4 miles. The general form of my staggered adoption difference-in-differences model is as follows:

$$y_{ict} = \beta D_{ict} + \delta_i + \gamma_{ct} + \epsilon_{ict} \tag{1.1}$$

$y_{ict}$ is the outcome of interest for census tract $i$ in county $c$ at year $t$. $D_{ict}$ is an indicator that equals 1 if tract $i$ in county $c$ at time $t$ has its centroid within 1 mile of an open rail station, and zero otherwise.[13] $\delta_i$ are census tract fixed effects which control for time invariant tract characteristics that affect $y_{ict}$. $\gamma_{ct}$ are county-by-year fixed effects. Their inclusion allows me to control for yearly shocks that are specific to individual counties. For example, if there is an increase in parking fees or public transit fares at the county level, $\gamma_{ct}$ would capture the resulting change in $y_{ict}$. $\epsilon_{ict}$ is the error term and is clustered at the county level.

[13] In the online appendix, I include estimates using a continuous measure of distance (log distance) to the nearest rail station. All results are qualitatively similar to those using Equation 1.1.

$\beta$ is the parameter of interest. In order for $\beta$ to identify the causal effect, two assumptions must be satisfied. First, in the absence of treatment, outcomes of treated tracts would have followed an identical trend to those of untreated tracts. I provide support of this assumption by examining pre-treatment differences in outcomes using the event study framework described below. I also perform two econometric procedures that alter the sample, making the treated and untreated groups more similar based on pre-treatment observable characteristics. These procedures are discussed in Section 1.5.

Second, commuting outcomes of untreated tracts must be unaffected by the introduction of rail transit. This assumption is assessed using a flexible empirical specification in which the impacts of rail transit can vary non-linearly in distance. This is discussed in Section 1.5.

In order to address the validity of the parallel trends assumption, I employ an event study framework characterized by the following model:

$$\begin{aligned} y_{ic\tilde{t}} &= \sum_{s \in \mathcal{T}} \beta_s \mathbb{1}(\tilde{t} = s) \cdot Treatment_i + \delta_i + \gamma_{ct} + \epsilon_{ict} \\ \mathcal{T} &= \{-4, -3, -2, 0, 1, 2, 3\} \end{aligned} \tag{1.2}$$

$\tilde{t}$ represents the year relative to treatment in 10 year increments. $\tilde{t} = -1$ represents the last decennial year prior to treatment, while $\tilde{t} = 0$ is the first decennial year in which tract $i$ is treated. The excluded relative year category is $\tilde{t} = -1$. $\mathbb{1}(\tilde{t} = s)$ is an indicator variable that equals 1 if $\tilde{t}$ equals $s$, and zero otherwise. $Treatment_i$ is an indicator for whether tract $i$ ever falls within 1 mile of a rail station.

The parameters of interest are the coefficients on the interaction between treatment status and the relative time: $\beta_{-4}, \beta_{-3}, \beta_{-2}, \beta_0, \beta_1, \beta_2$, and $\beta_3$. These coefficients describe how commuting outcomes of tracts that become treated differ over time relative to tracts that remain untreated. Under a parallel trends assumption, $\beta_0$ captures the effect of treatment at the first decennial year following the introduction of a rail station. $\beta_1, \beta_2$, and $\beta_3$ capture the dynamic effects of treatment in decennial years following the introduction.

The coefficients $\beta_{-4}, \beta_{-3}$, and $\beta_{-2}$ allow me to investigate differences in outcomes between the treated and untreated groups prior to the introduction of a rail station. Under the parallel trends assumption, these coefficients should be approximately equal to zero. I formally test the pre-treatment coefficient estimates for joint significance ($H_0 : \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$) to assess the validity of this identifying assumption.

## 1.4 Impact on Commute Mode

This section presents the estimated impacts of the introduction of an urban rail station on census tract level commute mode. To begin, I estimate the effects on the proportion of commuters by public transit and private automobile. Then, using information on the exact mode of public transit, rail versus bus, I provide a more detailed description of what is underlying the observed changes in public transit usage. Following the discussion of the national level average estimates, I explore the intercity heterogeneity across each urban area in my sample.

### 1.4.1 Impacts on Public Transit and Automobile Commuting

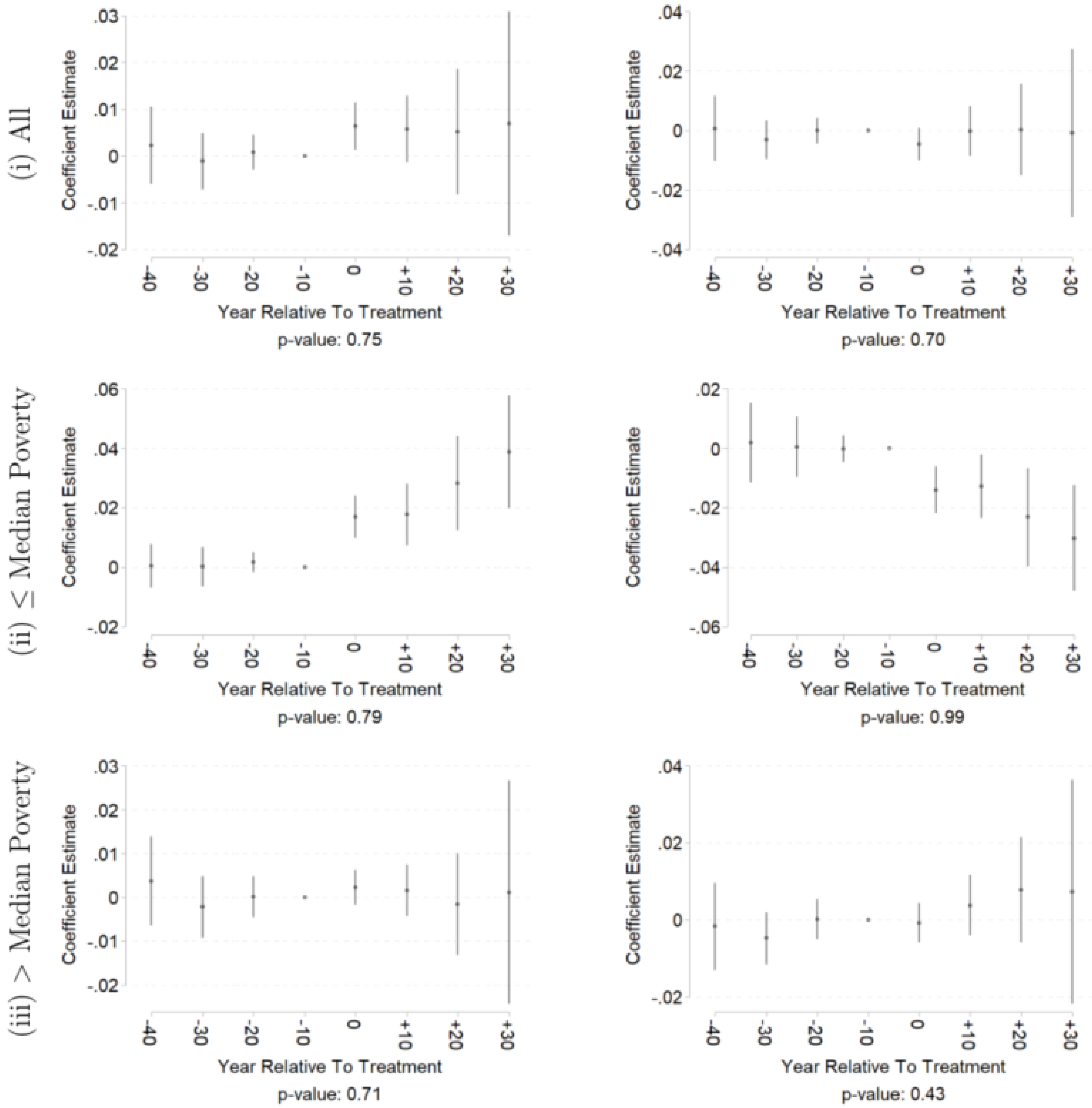
Figure 1.2 plots estimates of the parameters of interest along with 95 percent confidence intervals for the event study specification characterized by Equation 1.2. Panels (a) and (b) show estimates for the proportion of commuters whose primary mode is by public transit and by private automobile, respectively. Row (i) presents the results for the full sample (excluding tracts with missing values for the outcome), while Rows (ii) and (iii) display the estimates for tracts that had 1970 poverty rates below and above the median, respectively. Below each plot, I include the p-value for the test of whether the pre-treatment coefficient estimates are jointly equal to zero ($H_0 : \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$). The excluded relative time category is $\tilde{t} = -1$.

Visually, there are no differential trends in either outcome prior to treatment. All pre-treatment coefficient confidence intervals contain zero and the p-values for joint significance range between 0.43 and 0.99. This offers support of the parallel trends assumption required for the difference-in-differences framework to capture the causal effect. Following treatment, Row (i) shows a slight increase in public transit commuting with no obvious change in automobile usage. Row (ii) suggests that the changes are stronger in the below-median poverty subsample. Among these tracts, there is an approximate 2 p.p. increase in the decennial year immediately following treatment which appears to grow in subsequent years.

The apparent dynamic effects are of great interest as prior literature has suggested large differences between the short-run and long-run responses of neighborhood level commuting behavior following changes in rail service and quality. Notably, Voith (1991, 1997) provide theoretical and empirical evidence suggesting that short-run responses become amplified as transit authorities adapt to changes in ridership and the neighborhood composition shifts, further affecting ridership. In particular, they find long-run ridership to be twice as elastic in response to changes in service and quality relative to the short-run effects. However, their estimated mean lag parameter suggests that half of the lagged effect appears within only 1 year. Therefore, the decennial data available in the census is not ideal to distinguish short-run from long-run effects.

Indeed, the apparent dynamic effects in Figure 1.2 may actually reflect treatment effect heterogeneity across time as later relative year estimates are identified using tracts that were treated early in the study period. For example, the +20 and +30 year coefficients are identified using tracts treated prior to 1990. If earlier treated tracts experience a stronger response, then the apparent dynamic effects may be a consequence of the staggered-adoption empirical design. Restricting the sample of treated tracts to those before 1990, I examine similar event study plots (found in the online appendix). While the point estimates, particularly those in Row (ii), increase monotonically across time following treatment, the trend is far weaker and estimates cannot be statistically distinguished from one another. The effects are also stronger in this subsample across all post-treatment relative year coefficients. This suggests that the apparent dynamic effects are more likely a result of selection into early treatment.

Table 1.3 presents the estimates of Equation 1.1. Panel A includes tracts in all locations across each metropolitan area. Columns (1) and (2) display the estimates using the full sample of tracts, whereas the latter columns split the sample based on the 1970 tract level poverty rate relative to the median. From Column (1), census tracts that became closer than 1 mile of a rail station saw a statistically significant 0.66 percentage point (p.p.) increase in the proportion of public transit commuters relative to tracts that remained farther than 1 mile of a rail station. On an average proportion of 0.16, this represents a 4.1 percent increase in public transit commuting in treated tracts. The coefficient on automobile usage is negative, but small in magnitude and statistically indistinguishable from zero. While public transit commuting appears to increase slightly following the introduction of a rail station, there is little evidence of a change in automobile usage using the full sample.




(a) Proportion Public (b) Proportion Auto

Figure 1.2: Event Study - Proportion Public/Proportion Auto

Estimated coefficients and 95% confidence intervals $\beta_s$ in Equation 1.2. Standard errors clustered at the county level. Row (i) includes all census tracts in the sample (excluding missing tracts with missing values), while Row (ii) and Row (iii) restrict to tracts that had a 1970 poverty rate that was below-median and above-median, respectively. p-values for the joint significance of the pre-treatment coefficients ($H_0 : \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$) are found below each plot. The excluded relative year category is $\tilde{t} = -1$.

Table 1.3: Proportion Public and Auto Transit

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | | | $\leq$ Median Pov. | $\leq$ Median Pov. | $>$ Median Pov. | $>$ Median Pov. |
| | Public | Auto | Public | Auto | Public | Auto |
| Panel A: All Locations | | | | | | |
| $\leq$ 1 Mile | 0.00661** | -0.00267 | 0.0197*** | -0.0172*** | 0.000943 | 0.00395 |
| | (0.00320) | (0.00369) | (0.00511) | (0.00540) | (0.00258) | (0.00329) |
| Observations | 94,199 | 94,199 | 46,434 | 46,434 | 47,765 | 47,765 |
| Mean Outcome | 0.16 | 0.76 | 0.11 | 0.84 | 0.21 | 0.69 |
| Panel B: Outer City | | | | | | |
| $\leq$ 1 Mile | 0.0156*** | -0.0128** | 0.0208*** | -0.0188*** | 0.0107*** | -0.00671 |
| | (0.00418) | (0.00494) | (0.00559) | (0.00576) | (0.00323) | (0.00533) |
| Observations | 61,183 | 61,183 | 37,489 | 37,489 | 23,694 | 23,694 |
| Mean Outcome | 0.13 | 0.82 | 0.10 | 0.85 | 0.16 | 0.77 |
| Panel C: Inner City | | | | | | |
| $\leq$ 1 Mile | 0.00271 | 0.00403 | 0.0206*** | -0.0179** | -0.00162 | 0.00882** |
| | (0.00341) | (0.00408) | (0.00659) | (0.00715) | (0.00341) | (0.00394) |
| Observations | 33,016 | 33,016 | 8,945 | 8,945 | 24,071 | 24,071 |
| Mean Outcome | 0.24 | 0.66 | 0.15 | 0.79 | 0.27 | 0.61 |

Clustered standard errors (at county-level) in parentheses. Sample includes tracts that were every within 4 miles of a rail station by 2010. Outer city tracts lie farther than 7 miles from city CBD. Inner city tracts closer than 7 miles from CBD.

In Columns (3) and (4), I include tracts that had a pre-treatment poverty rate that was below the sample median. The results suggest relatively large effects for this subsample. Column (3) shows a significant 1.97 p.p. increase in the proportion of public transit commuters in response to treatment. This increase represents a 17.9 percent change on the average proportion of public transit commuters within these tracts. The estimate in Column (4) suggests a roughly equivalent reduction (1.70 p.p.) in automobile usage. Columns (5) and (6) present results for the above-median poverty subsample. The estimates are both small in magnitude and statistically insignificant suggesting no discernible change in the proportion of commuters by public transit or automobile following treatment.

The results of Panel A are consistent with Baum-Snow, Kahn and Voith (2005) suggesting that the largest changes in commute mode in response to rail transit are found in neighborhoods of relatively high socioeconomic status. However, the high speed of rail transit makes it an attractive option for those with more distant commutes. As many of the lower-poverty neighborhoods during this time period are found in regions of the city farther from the central business district (CBD), it is possible that the observed heterogeneity in Panel A reflects the spatial distribution of poverty across each metro area as opposed to any direct relationship to socioeconomic status. I investigate this possibility in Panels B and C by splitting the sample into outer-city and inner-city tracts based on the distance between each tract centroid to the CBD.[14]

In Panel B, the sample is restricted to tracts that are farther than 7 miles from the CBD. This subsample represents neighborhoods that have farther distance commutes to each metro area's focal point of business and commercial activity. Overall, tracts experience a significant 1.56 p.p. increase in public transit commuting and a 1.28 p.p.

[14]Event study plots for each subsample can be found in the online appendix. The pre-treatment coefficients are predominantly consistent with limited pre-trends.

reduction in automobile commuting following treatment. These estimates are much larger than those in Panel A, consistent with the notion that rail transit is more useful for those with longer distance commutes. For the below-median poverty, outer-city subsample, the estimates are very similar to those in Panel A. Regarding the above-median poverty subsample, Column (5) suggests a statistically significant 1.07 p.p. increase in public transit usage. However, there is no evidence of a change in automobile commuting for this group.

Panel C restricts to tracts that are within 7 miles of the CBD. The estimates in Columns (1) and (2) suggest limited overall impacts of rail transit for this inner-city subsample. However, the estimates for the lower-poverty group are remarkably similar to those in Panel B. For the higher-poverty subsample, the estimate on public transit usage is negative, but small in magnitude and statistically indistinguishable from zero. The positive estimate in Column (6) is surprising, but disappears after controlling for the slight upward pre-trend observed in the event study figure for this subsample.

Overall, the results suggest a large increase of approximately 2 p.p. in the proportion of public transit commuters in below-median poverty tracts, regardless of location relative to the CBD. This increase is accompanied by a slightly smaller magnitude decrease in the proportion of automobile commuters. For above-median poverty tracts, there is an increase in public transit commuting for tracts located farther than 7 miles from the CBD, but the magnitude is only half that of the lower-poverty subsample. There is limited evidence of a change in automobile commuting among these higher-poverty tracts.

### 1.4.2 Bus and Rail Commuting

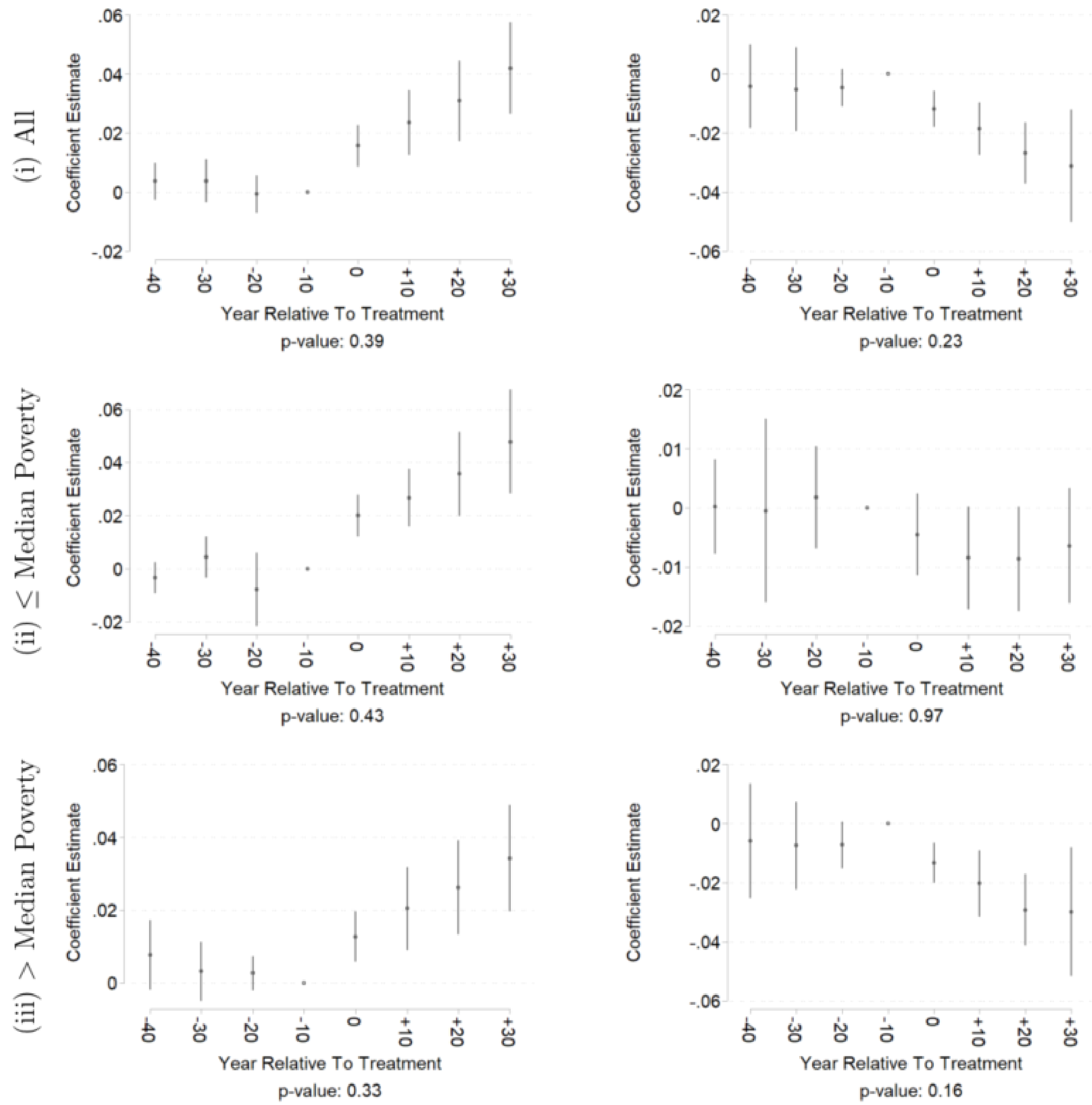
Using information on the specific mode of public transit, rail versus bus, I now dig deeper to understand what is underlying the observed changes to public transit usage. I investigate the possibility that increases in rail usage may be offset by reductions in bus usage leading to a limited impact on overall public transit commuting. This may be the case if the introduction of rail transit is accompanied by reductions in bus service or if individuals that once commuted by bus now travel predominantly by rail to reach their employment destination.

Figure 1.3 presents the event study plots using these more specific measures of public transit usage. There are no clear visual trends in outcomes prior to the introduction of a rail station. None of the pre-treatment estimates are statistically distinct from zero and the p-values for joint significance range from 0.16 to 0.97. Following treatment, there is an immediate increase in the proportion of rail commuters that appears to grow over time. The increasing effect in subsequent years is, again, driven by larger estimates among tracts treated earlier in the sample.[15] The impact is apparent in each poverty subsample, although the magnitude of the increase is larger for below-median poverty tracts. Bus commuting experiences a decrease in response to treatment, though the change seems to be primarily driven by above-median poverty tracts.

Table 1.4 presents the difference-in-differences estimates. From Column (1) of Panel A, treated tracts observe a statistically significant 1.75 p.p. increase in rail commuting following the introduction of a rail station. On an average proportion of 0.08, this represents a 21.9 percent increase in rail commuting among treated tracts. Column (2) shows that this increase in rail usage is accompanied by a significant 1.04 p.p. reduction in bus usage. This suggests that 59.4 percent of the increase in rail usage is offset by

[15] Event study plots for earlier treated tracts appear in the online appendix.

decreases in bus usage resulting in the small overall effect on public transit commuting observed in Section 1.4.1.




(a) Proportion Rail (b) Proportion Bus

Figure 1.3: Event Study - Proportion Rail/Proportion Bus

Estimated coefficients and 95% confidence intervals $\beta_s$ in Equation 1.2. Standard errors clustered at the county level. Row (i) includes all census tracts in the sample (excluding missing tracts with missing values), while Row (ii) and Row (iii) restrict to tracts that had a 1970 poverty rate that was below-median and above-median, respectively. p-values for the joint significance of the pre-treatment coefficients ($H_0 : \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$) are found below each plot. The excluded relative year category is $\tilde{t} = -1$.

Table 1.4: Proportion Bus and Rail Transit

| | (1) Rail | (2) Bus | (3) ≤ Median Pov. Rail | (4) ≤ Median Pov. Bus | (5) > Median Pov. Rail | (6) > Median Pov. Bus |
|---|---|---|---|---|---|---|
| Panel A: All Locations | | | | | | |
| ≤ 1 Mile | 0.0175*** | -0.0104*** | 0.0276*** | -0.00677 | 0.0111*** | -0.0102*** |
| | (0.00428) | (0.00332) | (0.00615) | (0.00474) | (0.00311) | (0.00374) |
| Observations | 75,328 | 75,328 | 37,121 | 37,121 | 38,207 | 38,207 |
| Mean Outcome | 0.08 | 0.08 | 0.06 | 0.05 | 0.09 | 0.12 |
| Panel B: Outer City | | | | | | |
| ≤ 1 Mile | 0.0168*** | -0.00119 | 0.0222*** | -0.00108 | 0.0104** | -0.000345 |
| | (0.00485) | (0.00338) | (0.00581) | (0.00409) | (0.00423) | (0.00359) |
| Observations | 48,942 | 48,942 | 29,980 | 29,980 | 18,962 | 18,962 |
| Mean Outcome | 0.07 | 0.05 | 0.06 | 0.04 | 0.08 | 0.07 |
| Panel C: Inner City | | | | | | |
| ≤ 1 Mile | 0.0178*** | -0.0142*** | 0.0356*** | -0.0125** | 0.0115*** | -0.0129** |
| | (0.00543) | (0.00487) | (0.0108) | (0.00573) | (0.00407) | (0.00517) |
| Observations | 26,386 | 26,386 | 7,141 | 7,141 | 19,245 | 19,245 |
| Mean Outcome | 0.09 | 0.14 | 0.06 | 0.08 | 0.10 | 0.16 |

Clustered standard errors (at county-level) in parentheses. Sample includes tracts that were every within 4 miles of a rail station by 2010. Outer city tracts lie farther than 7 miles from city CBD. Inner city tracts closer than 7 miles from CBD.

Splitting the sample by poverty rate and location relative to the CBD, the results suggest that the largest increases in rail usage take place in below-median poverty tracts, regardless of location. The effect on this group is particularly strong in the inner portion of the city. For higher-poverty tracts, the increase in rail usage is relatively stable with respect to location, but between 30-50 percent of the magnitude of the change in the lower-poverty sample.

Changes in bus transit commuting are limited to the inner portion of the city with both poverty groups experiencing approximately a 1.3 p.p. reduction in usage. There is no impact on bus usage in the outer portion of the city, regardless of initial poverty rate. The reduction in bus usage in the inner-city is enough to completely offset the increase in rail commuting for the above-median poverty group. For the lower-poverty group, this reduction is not strong enough to negate the large increase in rail commuting.

Overall, it appears that neighborhoods of higher socioeconomic status are more likely to shift towards rail commuting. The resulting increase in public transit commuting in these areas is accompanied by a reduction in private vehicle usage. However, less well-off neighborhoods see smaller increases in rail commuting with roughly equivalent declines in bus usage.

## 1.5 Assessing the Empirical Assumptions

### 1.5.1 Improved Comparability

The primary challenge to estimating the effects of expanded rail transit lies in finding a suitable counterfactual (Redding and Turner (2015)). In particular, rail infrastructure is not randomly assigned within a metropolitan area. Policymakers target areas with certain historical or forecast characteristics which may result in treated neighborhoods that are fundamentally different than untreated neighborhoods. Here, I address this problem by applying two methodological techniques that effectively alter the sample so that comparisons can be made between tracts that are more similar based on observable characteristics.

First, I employ a trimming procedure to construct a sample of treated and untreated tracts that share similar observable characteristics prior to treatment (Crump et al. (2009)). An indicator for whether a tract is ever treated is regressed on a series of demographic, commuting, and housing characteristics from 1970 (prior to any treatment) along with a set of county level fixed effects using a lasso probit regression model.[16] The model predicts treatment fairly well with an R-squared of 0.172. Estimates from this model are used to calculate the propensity score of becoming within 1 mile of a rail transit station. The sample is restricted to tracts with a propensity score between 0.1 and 0.9 and Equation 1.1 is estimated.

This trimming technique eliminates tracts that appear either very likely or very unlikely to be treated (22.2 percent of the sample). Of the remaining tracts, 2,760 come from the original treatment group while 8,267 come from the original control group meaning a reduction of 8.0 percent and 26.0 percent, respectively. Those excluded are less densely populated, farther from the CBD, and of relatively high socioeconomic status prior to treatment. The remaining set of tracts are much more similar based on pre-treatment characteristics, reducing the potential for omitted variable bias (trimmed summary statistics table available in the online appendix).

I also perform an inverse probability weighting (IPW) procedure as in Hirano, Imbens and Ridder (2003). I weight treated tracts by the inverse of their previously calculated propensity score, and untreated tracts by the inverse of one minus their propensity score. The procedure results in a sample where treated tracts that initially look like untreated

[16] Results are very similar using a standard probit model.

tracts and untreated tracts that initially look like treated tracts are weighted more heavily (IPW summary statistics table available in the online appendix).

Figure A.1 compares estimates from each robustness procedure with those from the main specification. For each outcome, I include the point estimate and 95 percent confidence interval. The first row of each figure displays estimates from the main specification, while subsequent rows display estimates from each robustness procedure. All trimming and IPW estimates are statistically indistinguishable from those using the unaltered sample.

Figures A.2 and A.3 display the same plots for the below-median poverty and above-median poverty subsamples, respectively. Again, the estimates are generally very similar to those in the main specification. Figure A.3 shows slight differences among above-median tracts using the IPW procedure. Specifically, the estimates suggest a small increase in public transit usage and a reduction in automobile usage. However, the estimates are not statistically different from those in the main specification. Overall, the results of each procedure suggest that using a more comparable sample of tracts does not change the results in a meaningful way. This provides support that the estimates in Section 1.4 are not a reflection of omitted variable bias brought on by endogenous selection into treatment.

### 1.5.2 Falsification Test For Selection

As an additional test for whether selection may be driving the results, I conduct two falsification tests which control for future station openings. If tracts that are chosen to receive a station have differential trends in commuting outcomes regardless of whether a station has actually been introduced, then the coefficient estimates for future treatment will reflect this selection. I estimate Equation 1.1 controlling for measures of future treatment. In seperate specifications, I control for two alternative definitions of future treatment: 1) whether tract $i$ at any point in future years will have a rail station within 1 mile and 2) whether tract $i$ became treated in the subsequent decennial year. If the introduction of a rail station is truly responsible for the effects in Section 1.4 the estimated difference-in-differences estimate will be unchanged by the inclusion of these controls, and the coefficient on future treatment should equal zero.

The difference-in-differences estimates, after controlling for future treatment, appear in Rows 4 and 5 of Figure A.1, A.2, and A.3. Controlling for future stations in Rows 4 and 5 do not change any of the estimates. For each variable within each subsample plot, the effects are very similar to those in the main specification. The online appendix presents the regression results with future station controls. The coefficient estimates for each future treatment indicator are predominantly small and insignificant. While some of the coefficients on rail and bus usage are significant or marginally significant, the estimates are much smaller than the difference-in-differences coefficients and of the opposite sign of the treatment effect. This suggests that these tracts may have been trending in the opposite direction than those estimated in the main results. I conclude that the falsification estimates are consistent with treatment driving the results as opposed to selection.

### 1.5.3 Distance Bin Specification

The difference-in-differences empirical strategy described by Equation 1.1 relies on a fixed distance threshold of 1 mile to define treatment. However, it is unclear that this threshold accurately separates which census tracts were and were not affected by the introduction of a rail station. In fact, it is plausible to imagine that the impacts would be strongest in tracts located in closest proximity and decline roughly monotonically in distance from the station.

To better understand the geographical extent of the effects, I augment the difference-in-differences framework allowing the treatment status of each tract to depend on the distance to the nearest open rail station. I restrict the sample to tracts that are ever within 20 miles of a station to allow for the potential that the impacts extend beyond the 4 mile threshold used for inclusion in my primary sample. I then define six distance bins: $\leq 1/2$ mile, 1/2-1 mile, 1-2 miles, 2-4 miles, 4-10 miles, and 10-20 miles. For each decennial year, I allocate tracts to a bin based on the distance to the nearest open rail station. This flexible specification allows for the effects of treatment to depend non-linearly on distance. The goal is to measure the effect of falling within distance bin $d$ on each commuting outcome relative to the reference category, tracts that fall farther than 20 miles from a rail station. The empirical specification is defined by Equation 1.3.

$$y_{ict} = \beta_1 D_{ict}^{\leq 1/2} + \beta_2 D_{ict}^{1/2-1} + \beta_3 D_{ict}^{1-2} + \beta_4 D_{ict}^{2-4} + \beta_5 D_{ict}^{4-10} + \beta_6 D_{ict}^{10-20} + \delta_i + \gamma_{ct} + \epsilon_{ict} \quad (1.3)$$

where $D_{ict}^{d}$ is an indicator that equals 1 if tract $i$ in county $c$ at year $t$ falls in distance bin $d$, and zero otherwise (e.g. $D_{ict}^{\leq 1/2}$ is an indicator for having an open station within 1/2 mile).

Overall, the results (found in the online appendix) are consistent with the notion of strong local effects that decrease monotonically in distance. Most of the estimates approach zero quickly, with small and insignificant effects farther than 1 mile from a new station. This provides support for the 1 mile threshold used in the main specification.

Two notable exceptions exist where the coefficients do not decline completely within 1 mile. First, for tracts located farther than 7 miles from the central business district, the magnitude of the estimates on rail and auto usage decline more slowly. I observe increases in rail usage and decreases in automobile usage for this subsample that extent up to 4 miles from a new station. This is perhaps a reflection of individuals that were initially driving into the central city now driving to a station and completing the rest of the trip by rail, avoiding congestion in the central portion of the city.

Second, reductions in bus usage in above-median poverty, inner-city tracts decline slowly in distance with statistically significant estimates up to 10 miles from a new rail station. This is perhaps suggestive that the introduction of rail infrastructure may be accompanied by local reductions in bus usage that can affect higher poverty tracts in moderate distances from new stations.

### 1.5.4 Heterogeneous Effects and Negative Weighting

The difference-in-differences specification characterized by Equation 1.1 is susceptible to the problem of treatment effect heterogeneity and negative weighting as described in de Chaisemartin and d'Haultfoeuille (2019). I apply their Wald-Time Corrected (WTC) estimator to examine the robustness of each estimate to this concern. The fifth row of Figure A.1 presents estimates based on this alternative estimator. The WTC point estimates are primarily very similar to those of the main specification. One exception is that the WTC estimate suggests a larger reduction in bus usage. This is particularly evident in the above-median poverty sample in Figure A.3. However, correcting for treatment effect heterogeneity does not appear to affect the main conclusions of the empirical analysis.

### 1.5.5 Improved Panel Balance

Figure 1.1 shows that the timing of treatment varies widely across tracts. As a result, the staggered adoption difference-in-differences and event study specifications identify parameters using an unbalanced panel. For tracts treated between 1971 and 1980, there exists commuting data from a single decennial year prior to treatment and four decennial years following treatment. Tracts treated between 2001-2010 have four decennial years prior to treatment and a single post-treatment decennial year.

I address this by estimating both specifications using a more balanced set of tracts based on when they were treated. Two subsamples of tracts are used. First, I restrict to tracts that became treated between 1991 and 2010. These tracts all have at least three decennial years prior to treatment and either one or two post-treatment decennial years. This subsample is particularly useful in assessing the robustness of the pre-trends analysis. The event study figures based on this restricted sample can be found in the online appendix and are primarily consistent with limited pre-trends. There are no clear visual trends prior to treatment and the p-values associated with the tests of joint significance yield mostly insignificant results.

Second, I restrict to tracts that became treated between 1971 and 1990. These tracts have at least three post-treatment decennial years of data and either one or two pre-treatment decennial years. While less useful in assessing pre-trends, estimates of Equation 1.1 using this sample, along with the 1991-2010 sample, are included in Figure A.1. The results are qualitatively similar to those of the main specification. Although, it appears that the estimates based on the 1971-1990 treatment sample suggest larger impacts on rail usage in response to treatment. I conclude that the results are robust to using a more balanced panel.

## 1.6 Cross-City Heterogeneity and Aggregate Effects

### 1.6.1 Cross-City Heterogeneity

The large number of cities expanding rail infrastructure during the study period provides an excellent opportunity to explore intercity heterogeneity. Here, I estimate Equation 1.1 restricting the sample to tracts located within each city. Splitting the sample by city reduces the statistical power of inference, particularly in cities that experienced a small number of station openings. In addition, as much of the development occurred in a small number of counties within each city (often a single county for many of the cities) it is difficult to cluster standard errors at the county level. However, by examining the point estimates across cities, there are some interesting trends that appear.

Figure 1.4 displays the difference-in-differences point estimates for the fraction commuting by public transit for each city. I include three estimates; one for the effect on all tracts within a given metro area and also the effect by tract poverty rate relative to the city-specific median. The estimates are organized by treatment effect in decreasing order. Similarly to Baum-Snow, Kahn and Voith (2005), I find that commuters near new stations in Washington DC and Boston experience some of the largest increases in public transit usage. Other notable cities with large gains to transit usage are Nashville, New Haven, San Francisco, and Portland.[17] On the opposite end of the spectrum, many of the sunbelt cities including Houston, Dallas, Austin, Atlanta, and Denver, along with Seattle, and some Mid-Western cities experience negligible (and even negative) changes in public transit commuting.

A striking feature of the estimates is the difference by poverty rate. In nearly every city, lower-poverty tracts experience more positive effects relative to above-median

[17]Nashville and New Haven are not precisely estimated and appear not to be driven by increases in rail usage.

poverty tracts. This fact is especially predominant in cities that did not see an overall increase in public transit usage. Of the fifteen cities that experienced the smallest gains to transit use, fourteen experienced larger gains in lower-poverty neighborhoods. This gap generally ranges between 1-6 p.p./ and suggests that many higher-poverty neighborhoods in these low-performing cities actually experienced a decline in public transit usage.

Figure 1.5 presents the same plot for automobile commuting. Among lower-poverty neighborhoods, much of the increase in public transit usage comes from former car commuters. However, there is little evidence that treatment leads to a reduction in auto usage in higher-poverty neighborhoods.

Using information on the exact mode of public transit, Figure 1.6 documents that rail commuting generally increases in treated tracts across all cities. Washington DC and Boston are again high performers in generating rail usage. It is noteworthy that changes in rail usage are far less dependent on the tract poverty rate. While there is substantial variation across cities, there is little variation within cities based on poverty. Figure 1.7 shows the bus specific transit results. These estimates highlight the source of much of the disparity between lower and higher-poverty neighborhoods. Among cities that saw a decline in bus commuting, much of the decrease was driven by above-median poverty neighborhoods.

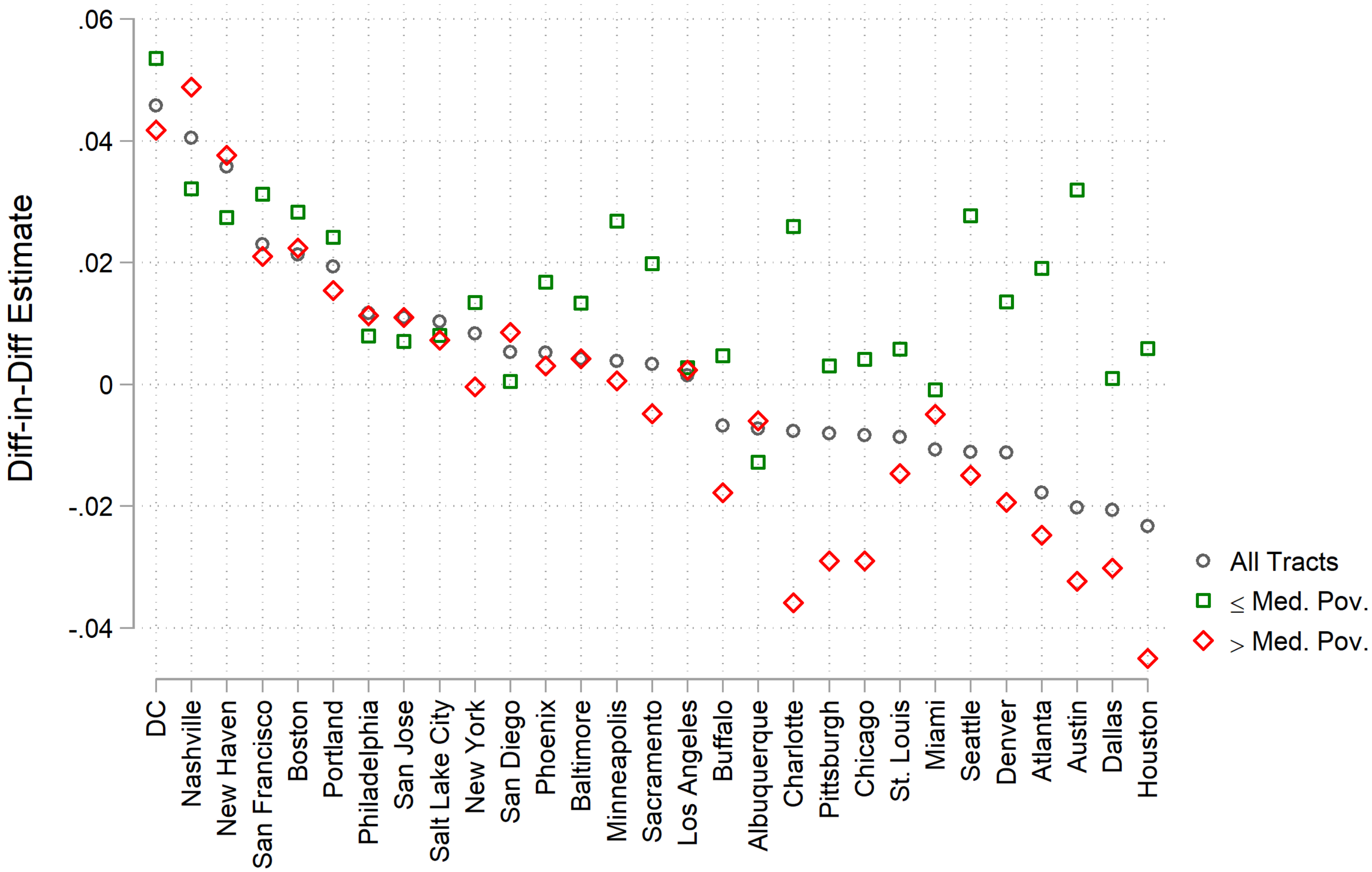



Figure 1.4: Prop. Public By Metro

Difference-in-differences point estimate by city. Grey circles represents estimates using the full sample within each city. Green squares and red diamonds represent those for the below and above-median poverty tracts, respectively, within each city. Median poverty is city-specific.

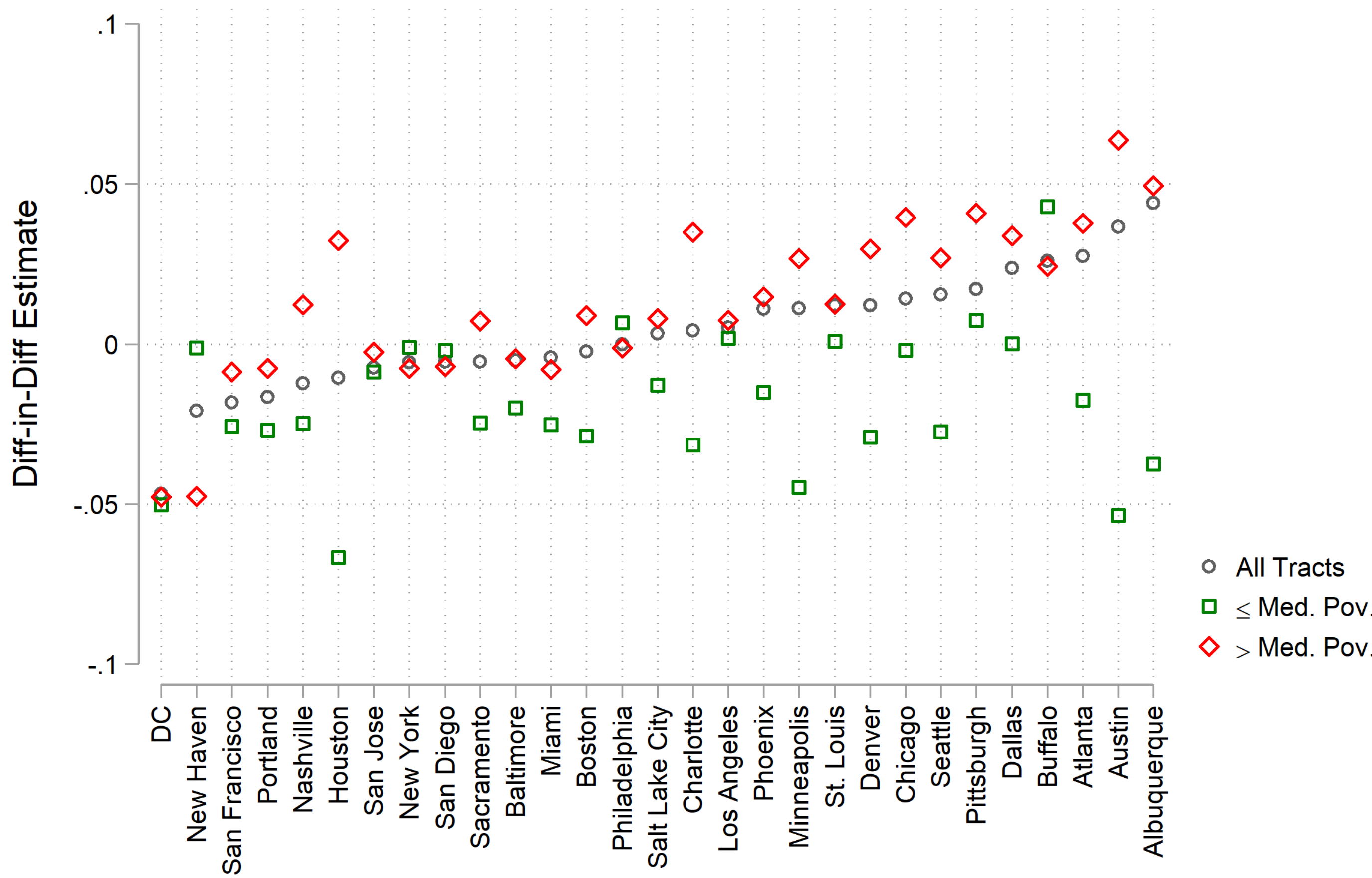



Figure 1.5: Prop. Auto By Metro

Difference-in-differences point estimate by city. Grey circles represents estimates using the full sample within each city. Green squares and red diamonds represent those for the below and above-median poverty tracts, respectively, within each city. Median poverty is city-specific.

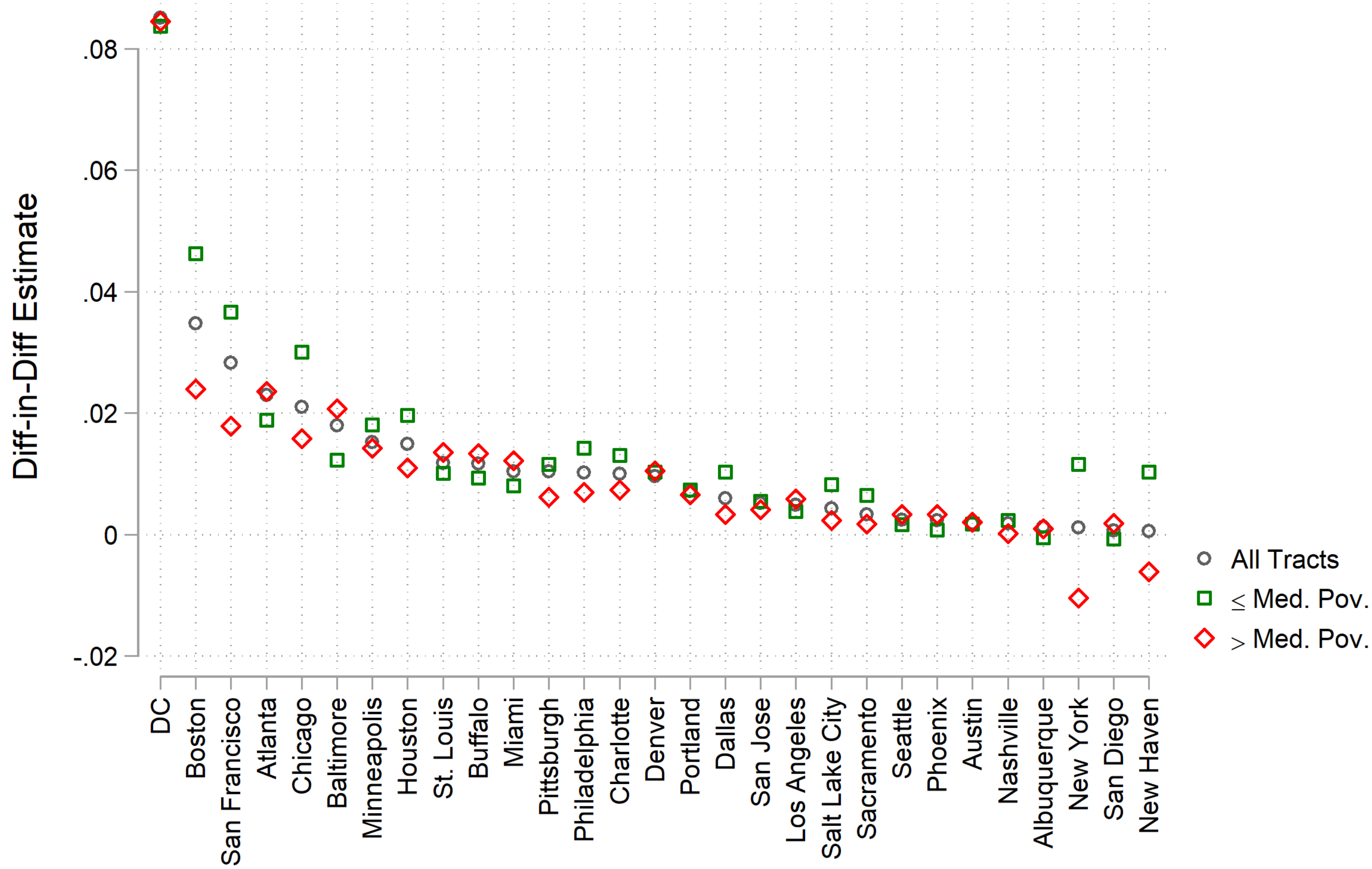



Figure 1.6: Prop. Rail By Metro

Difference-in-differences point estimate by city. Grey circles represents estimates using the full sample within each city. Green squares and red diamonds represent those for the below and above-median poverty tracts, respectively, within each city. Median poverty is city-specific.

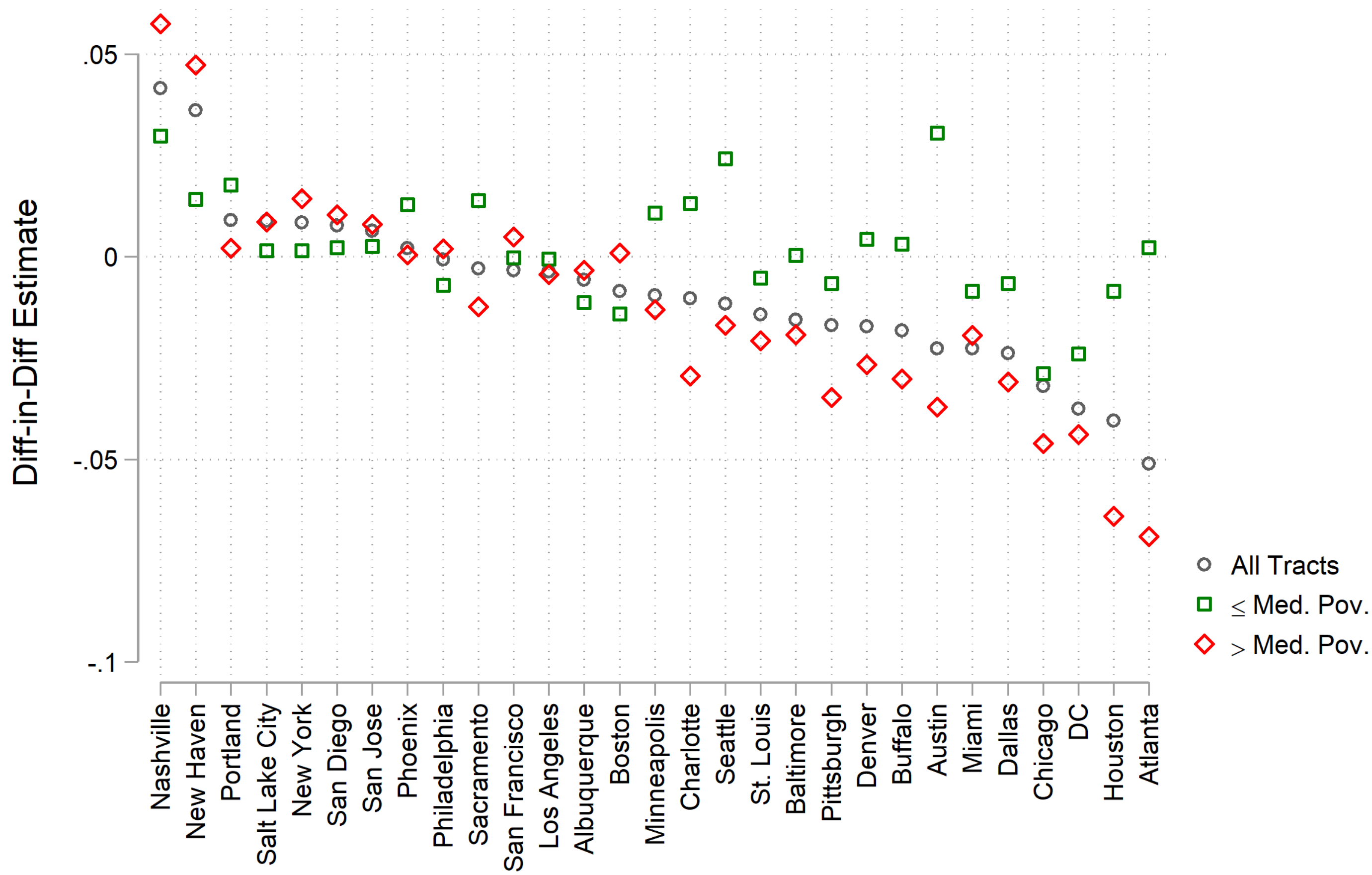



Figure 1.7: Prop. Bus By Metro

Difference-in-differences point estimate by city. Grey circles represents estimates using the full sample within each city. Green squares and red diamonds represent those for the below and above-median poverty tracts, respectively, within each city. Median poverty is city-specific.

The results indicate considerable variation in the effect of rail development. Washington DC and Boston appear to have successfully increased public transit usage in areas local to new stations, while nearly half of the cities experience null or even decreases to the transit share. Reductions in bus usage, particularly in above-median poverty neighborhoods offsets much of the increase in rail commuting in these cities. The evidence is consistent with the overall results that rail development is less effective at increasing transit usage in higher-poverty neighborhoods. The fact that public transit usage actually declines in higher-poverty neighborhoods within many of the cities is not particularly encouraging. If reductions in bus usage were accompanied by equivalent increases in rail usage within these neighborhoods, then this would perhaps reflect former bus users switching to rail. However, the results suggest that some reductions in bus share are not accompanied by large enough increases to rail usage providing evidence that perhaps rail could be crowding out bus transit.

### 1.6.2 Aggregate Effects

A primary objective of rail development is to increase public transit usage throughout an entire metropolitan area. In light of the small magnitude of the overall effects, it is not clear that these neighborhood level impacts translate into aggregate changes at the metropolitan level. I employ a difference-in-differences specification that exploits variation in the extent of rail transit within census designated urbanized areas over time to estimate these aggregate effects.

The results can help policymakers assess the effectiveness of rail transit in achieving the intended goal of increasing metropolitan public transit usage while decreasing automobile usage. If the impacts on commuting exist only at the local level, transit authorities may be hesitant to make costly investments into their rail systems.

Census tracts are allocated to 2010 Census designated urban areas (UA) based on the tract centroid relative to each UA boundary.[18] I then aggregate counts of individuals by commute mode to the UA-by-decennial year level and construct proportions. The number of open rail stations within each UA is used as a proxy for the level of rail infrastructure within each UA. The empirical model is characterized by Equation 1.4.

$$y_{it} = \beta D_{it} + \delta_i + \gamma_t + \epsilon_{it} \tag{1.4}$$

$y_{it}$ represents the outcome of interest for UA $i$ in decennial year $t$. $D_{it}$ is the number of open rail stations within UA $i$ at year $t$. $\delta_i$ and $\gamma_t$ represent UA and year fixed effects, respectively. $\epsilon_{it}$ represents the error term and is clustered at the UA level. For comparability, I restrict analysis to UAs that had at least one rail station open by 2010. In this difference-in-differences specification, $\beta$ represents the change in $y_{it}$ associated with an additional station opened within a UA boundary, holding UA and year constant.

Table A.1 in the appendix presents the results. Perhaps surprisingly, an increase in the number of rail stations is accompanied by a statistically significant reduction in the proportion of public transit commuters at the UA level. For an additional 10 stations (roughly equivalent to a rail line extension), UAs experience an average 0.274 p.p. decrease in public transit commuting. While the coefficient on rail usage is positive (though statistically insignificant), there is a significant reduction in bus usage that more than offsets the increase in rail commuting. There is no noticeable change in automobile commuting in response to increased rail infrastructure.

The fact that bus usage decreases more than rail usage increases is surprising and seems counterproductive to the goals of transit policy. Given that bus users are generally

[18] "An urban area will comprise a densely settled core of census tracts and/or census blocks that meet minimum population density requirements, along with adjacent territory containing non-residential urban land uses as well as territory with low population density included to link outlying densely settled territory with the densely settled core." (U.S. Census Bureau)

of lower socioeconomic status relative to rail users, the reduction in bus commuting may be particularly harmful to the poor. I conclude that policymakers should be wary of the expectation that rail development will necessarily be accompanied by overall increases in public transit usage.

## 1.7 Migration

### 1.7.1 Migration As A Mechanism

Rail transit development permanently alters the spatial distribution of transportation infrastructure within a metropolitan area. In response, household preferences regarding residential location are likely to shift. Those that find rail access beneficial will prefer to locate closer to new stations. Some may view rail transit negatively and prefer to locate farther from stations. As preferences change, rental and housing prices will adjust and a new spatial distribution of households will emerge.

The possibility of changes in neighborhood demography make it difficult to disentangle incumbent behavioral changes from neighborhood composition effects. For example, observed increases in rail usage may be due to current residents switching their commute mode or in-migration of likely rail users. Accordingly, the results of Section 1.4 are interpreted as the estimated effects on the equilibrium locus of commuting outcomes for a given neighborhood reflecting both changes in habits and composition.

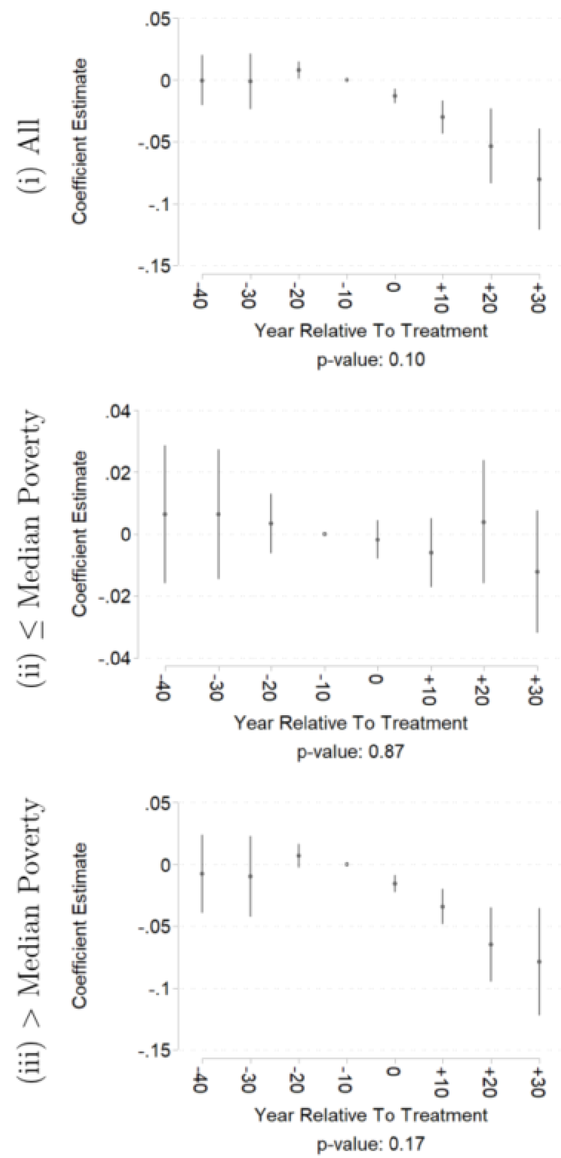
Migration complicates the question of who benefits most from expanded rail infrastructure. While policymakers may target certain communities when deciding where to place new stations, the intended users may not ultimately enjoy the amenity. In fact, if transit dependent households migrate away from treated neighborhoods (as a result of changes in preferences or prices), then proponents of rail transit may be misguided in their assertion that the poor benefit disproportionately.

I investigate neighborhood composition changes by estimating Equations 1.1 and 1.2 on an immutable tract level demographic characteristic, the proportion of black residents. During the study period, black commuters were 3.7-4.6 times more likely to commute by bus and 2.4-2.6 times more likely to commute by rail relative to non-black commuters. Under the assumption that these relative propensities are somewhat representative of the black individuals migrating in response to treatment, I can provide suggestive evidence of how these demographic changes may be influencing the results.

Figure 1.8 presents the estimates of Equation 1.2 using the tract level proportion of black residents. Prior to treatment, there is little evidence of differential trends between treated and untreated tracts. Following the introduction of a rail station, the proportion declines and the magnitude of the reduction grows in following years.

Table 1.5 presents the difference-in-differences estimates. On average, treated tracts experience a 2.23 percentage point decline in the proportion of black residents. The reduction is heavily driven by above-median poverty tracts. Estimating the effect on the log-number of residents by race (results in the online appendix), I find that the proportional change is due to out-migration of black residents in the higher poverty subsample. Under the assumption that these black residents are likely public transit users, their departure may be partially responsible for the limited effects on public transit usage among higher poverty neighborhoods.

For below-median poverty tracts, the demographic composition along this dimension appears less responsive to rail transit. However, inner-city neighborhoods experience a small decline in the proportion black which is driven by in-migration of non-blacks into treated neighborhoods. This may represent an influx of non-minority individuals that highly value the transit amenity leading to a particularly large increase in rail usage among this subsample.




(a) All Locations

Figure 1.8: Event Study - Proportion Black

Difference-in-differences point estimate by city. Grey circles represents estimates using the full sample within each city. Green squares and red diamonds represent those for the below and above-median poverty tracts, respectively, within each city. Median poverty is city-specific.

Table 1.5: Proportion Public and Auto Transit

| | (1) | (2) | (3) |
|---|---|---|---|
| | | $\leq$ Median Pov. | > Median Pov. |
| | Prop. Black | Prop. Black | Prop. Black |
| Panel A: All Locations | | | |
| $\leq$ 1 Mile | -0.0223*** | -0.00383 | -0.0227*** |
| | (0.00507) | (0.00377) | (0.00676) |
| Observations | 94,328 | 46,494 | 47,834 |
| Mean Outcome | 0.17 | 0.09 | 0.24 |
| Panel B: Outer City | | | |
| $\leq$ 1 Mile | -0.0132 | 0.00139 | -0.0230* |
| | (0.00950) | (0.00491) | (0.0138) |
| Observations | 61,266 | 37,531 | 23,735 |
| Mean Outcome | 0.13 | 0.08 | 0.21 |
| Panel C: Inner City | | | |
| $\leq$ 1 Mile | -0.0163 | -0.0132** | -0.00858 |
| | (0.0143) | (0.00602) | (0.0184) |
| Observations | 33,062 | 8,963 | 24,099 |
| Mean Outcome | 0.23 | 0.11 | 0.28 |

Clustered standard errors (at county-level) in parentheses. Sample includes tracts that were every within 4 miles of a rail station by 2010. Outer city tracts lie farther than 7 miles from city CBD. Inner city tracts closer than 7 miles from CBD.

Overall, the results suggest that the estimated impacts on commute mode are at least partially driven by changes in population composition, particularly in higher poverty neighborhoods. If policymakers intend to target certain types of communities with rail infrastructure, the results highlight migration as an unintended consequence that could undermine such efforts. Regardless, the estimates in Section 1.4 help to understand the causal effects of rail access at the census tract level.

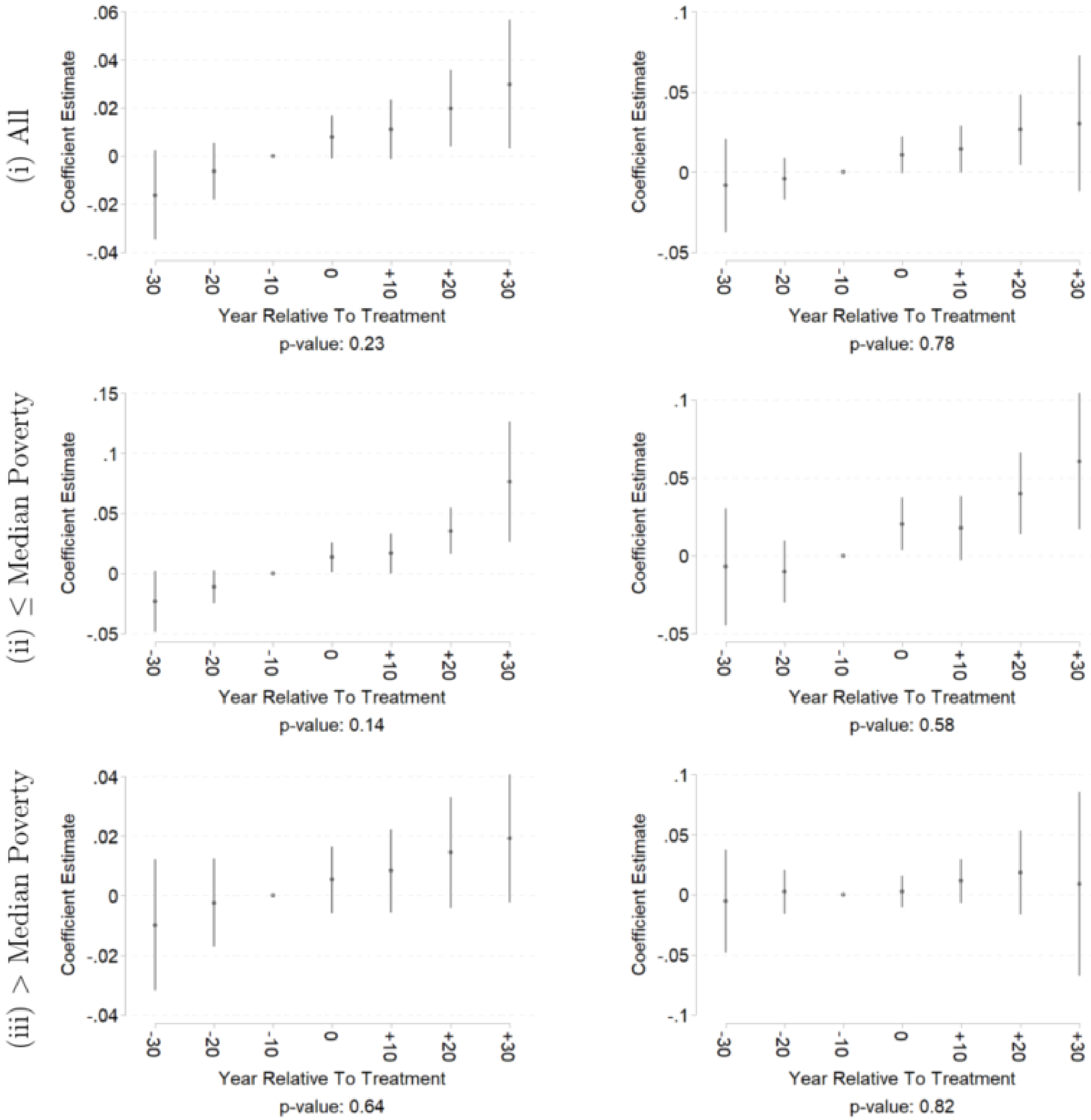
## 1.8 Impact on Commute Time

A potential benefit of rail transit development may be a reduction in the commute time for those residing in close proximity. The high speed and right-of-way benefits of rail travel may allow users a quicker way to navigate rush hour congestion relative to private automobile or bus travel. Any meaningful daily time savings would certainly be relevant to discussions regarding public transit policy. Here, I investigate this possibility by estimating the effect of rail transit access on the census tract level median one-way commute time.

Panel (a) of Figure 1.9 shows the event study plots on the log of the median commute time. As the 1970 Census Summary Files do not contain information on commute time, there are only, at most, three decennial years of data prior to treatment. The coefficients in each plot suggest that the outcomes of treated and untreated tracts follow slightly different trends prior to the introduction of a rail station. Visually, there is a small upward trend in the coefficients that appears driven by the below-median poverty subsample (p-value 0.14). The presence of pre-trends raises concern regarding the difference-in-differences empirical strategy.

In order to mitigate this problem, I apply the inverse probability weighting procedure discussed in Section 1.5. The event study plots using these weights are found in Panel (b). Visually, the procedure reduces the presence of pre-trends with p-values ranging from 0.58-0.82.

Table 1.6 displays the difference-in-differences estimates. Columns (1)-(3) present the unweighted estimates while Columns (4)-(6) show those using the inverse probability weights. In Panel A, Column (1) shows a marginally significant 0.201 minute average increase in one-way commute times following treatment. On an average of 26.91, this represents a 0.75 percent increase. Assuming a round trip commute, five days per week, 50 working weeks per year, this corresponds to a 1.68 hour increase in the time spent commuting. Similarly to the results on commute mode, the impacts appear to be strongest in below-median poverty tracts.




(a) Unweighted (b) Inverse Probability Weighted

Figure 1.9: Event Study - Median One-Way Commute Time (Log Minutes)

Estimated coefficients and 95% confidence intervals $\beta_s$ in Equation 1.2. Standard errors clustered at the county level. Row (i) includes all census tracts in the sample (excluding missing tracts with missing values), while Row (ii) and Row (iii) restrict to tracts that had a 1970 poverty rate that was below-median and above-median, respectively. p-values for the joint significance of the pre-treatment coefficients ($H_0 : \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$) are found below each plot. The excluded relative year category is $\tilde{t} = -1$.

Table 1.6: Median One-Way Commute Time (Log Minutes)

| | Unweighted | | | Inverse Probability Weighted | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Commute Time | ≤ Median Pov. Commute Time | > Median Pov. Commute Time | Commute Time | ≤ Median Pov. Commute Time | > Median Pov. Commute Time |
| Panel A: All Locations | | | | | | |
| ≤ 1 Mile | 0.0106** | 0.0188*** | 0.00519 | 0.0130** | 0.0217*** | 0.00489 |
| | (0.00450) | (0.00651) | (0.00528) | (0.00581) | (0.00785) | (0.00766) |
| Observations | 75,261 | 36,935 | 38,326 | 56,106 | 29,539 | 26,567 |
| Mean Outcome | 3.25 | 3.24 | 3.26 | 3.18 | 3.20 | 3.17 |
| Panel B: Outer City | | | | | | |
| ≤ 1 Mile | 0.0179*** | 0.0248*** | 0.00735 | 0.0119 | 0.0208** | 0.00315 |
| | (0.00554) | (0.00699) | (0.00746) | (0.00727) | (0.00853) | (0.0124) |
| Observations | 48,886 | 29,801 | 19,085 | 37,734 | 23,737 | 13,997 |
| Mean Outcome | 3.27 | 3.26 | 3.27 | 3.20 | 3.22 | 3.18 |
| Panel C: Inner City | | | | | | |
| ≤ 1 Mile | 0.00736 | 0.0116 | 0.00557 | 0.0153* | 0.0337 | 0.00529 |
| | (0.00613) | (0.0104) | (0.00673) | (0.00906) | (0.0210) | (0.00679) |
| Observations | 26,375 | 7,134 | 19,241 | 18,372 | 5,802 | 12,570 |
| Mean Outcome | 3.22 | 3.16 | 3.24 | 3.13 | 3.10 | 3.15 |

Clustered standard errors (at county-level) in parentheses. Sample includes tracts that were every within 4 miles of a rail station by 2010. Outer city tracts lie farther than 7 miles from city CBD. Inner city tracts closer than 7 miles from CBD.

The results in Columns (4)-(6) are predominantly very similar to those in the unweighted specification. Overall, the results suggest an increase in commute time following treatment. The increase appears to be largest in areas where public transit usage increased the most. I interpret the results as a lack of evidence that the introduction of rail infrastructure led to a decrease in commute time for tracts local to new rail stations. However, when estimating the impacts at the metropolitan area aggregate level using Equation 1.4, I find a small, but statistically significant decrease in median commute time in response to an increase in the number of rail stations.

## 1.9 Conclusion

Population growth and decentralization were pervasive elements of the U.S. urban landscape throughout much of the 20$^{\text{th}}$ century. These factors placed an enormous amount of stress on transportation networks which continues to afflict many metropolitan areas. As policymakers debate the merits of potential transit investments to mitigate this stress, empirical evidence regarding past transportation infrastructure development can be invaluable.

This paper examines the impacts of all urban rail expansions throughout the U.S. between 1970 and 2010 to understand the effects on commuting patterns. Four key empirical facts emerge. First, changes in neighborhood level commute mode resulting from increased rail access are far stronger in areas of higher socioeconomic status. These neighborhoods experience a high take-up of rail commuting which is accompanied by reductions in private automobile usage. I find no evidence that neighborhoods of lower socioeconomic status experience changes in public transit commuting.

Second, most of the increase in rail commuting is accompanied by reductions in bus usage. This result is particularly strong in poorer neighborhoods with reductions in bus usage completely offsetting any increases in rail usage. The decline in bus commuting extends up to ten miles from newly opened rail stations. At the metropolitan area aggregate level, increasing the number of total stations is accompanied by a reduction in bus commuting that exceeds the increase in rail commuting. As a result, public transit commuting actually declines as cities expand their rail systems. As bus users are disproportionately poor, this result is not particularly encouraging to the notion that rail development can be especially valuable in areas of low socioeconomic status.

Third, I find evidence of systematic migration in response to rail development. I show that black individuals migrate out of higher poverty neighborhoods that receive a new station. This out-migration is likely at least partially responsible for the limited change in public transit usage following treatment in these areas. I also document that non-blacks migrate into lower poverty, inner-city neighborhoods that receive a station. This may reflect non-minority households that highly value access to the new transit amenity.

Finally, I find no evidence of commute time savings in neighborhoods located near newly opened stations. In fact, the estimates suggest a small increase in commute times, particularly in areas that experience an increase in public transit usage. However, there is a small decrease in commute time observed at the metropolitan area aggregate level.

The evidence suggests that while rail development can increase public transit usage, it does so only at the local level and primarily in well-off neighborhoods. Policymakers should be wary of the expectation that rail transit will reduce private vehicle dependence and increase mobility among poorer communities.

# Chapter 2

# Rail Development, Job Access, and Employment

## 2.1 Introduction

The recent announcement of the Biden-Harris American Jobs Plan (AJP) underscores a resurgent focus of U.S. policy on physical infrastructure. Under the plan, $2.3 trillion will be invested over the next decade with a large portion dedicated to improving American transportation networks. A proposed $85 billion of the total is to be spent upgrading and expanding the nation's public transit infrastructure, roughly doubling the amount of federal dollars available to transit agencies. By investing these new funds into bus and rail capital projects, the AJP will seek to improve public transit's ability to meet rider demand and extend service into new communities. However, alternative spending options will force agencies to make critical decisions concerning effective use of the funds.

Beginning in the early 1970s, public rail transit became an increasingly popular tool for policymakers looking to better connect locations within urban areas. Between 1968 and 2017, 32 metropolitan areas opened or expanded public, commuter-oriented rail systems leading to the opening of more than 1,700 stations.[1] As the fundamental purpose of such investment is to lower the cost of travel between connected locations, proponents often contend that development can improve access to employment among targeted communities (Garrett (2004)). The systems provide a faster, and arguably more reliable, option for commuters relative to traditional bus service. By developing rail systems, policymakers may significantly improve a worker's ability to access centers of employment. In addition, the opening of a new station may be accompanied by transit-oriented development leading to an increase in local labor demand. For the city's residents that rely on public transit and proximity to access opportunities for employment, the benefits to rail development may be substantial.

A central theme of the AJP is to develop infrastructure intended to improve opportunity among vulnerable populations. High levels of public transit usage among low-skilled, low-income, and minority residents (Clark (2017)) suggest that investment could serve as a powerful tool to this aim. The spatial distribution of residence and employment across many U.S. metropolitan areas became increasingly dispersed throughout the latter half of the twentieth century (Baum-Snow (2010); LeRoy and Sonstelie (1983)). Those wealthy enough to afford cars increasingly abandoned living in central cities, electing to reside in the outer suburbs where land was comparatively cheap. By contrast, low-income and often minority households faced a range of barriers to relocating towards the city's periphery.[2] As centers of employment decentralized, a lack of geographic mobility among the urban poor left them unable to access an increasing proportion of jobs. This spatial mismatch between residential and employment location has been posited as a driver of urban income and racial inequality (Brueckner and Zenou (2003); Kain (1968)).

[1] These 32 metropolitan areas account for roughly 44 percent of the U.S. population (U.S. Census (2010)).

[2] E.g., racial redlining (Kain (1968)), dependence on public transit (Glaeser, Kahn and Rappaport (2008), and financial and time constraints exacerbated by a high incidence of domestic violence, mental health problems, holding multiple jobs, and raising children as a single parent (Bergman et al. (2019)).

This paper serves to investigate the extent to which public, rail transit development can be an effective tool in improving the economic well-being of targeted households. Using a novel panel of person and household level data, I exploit variation in rail access arising from the large number of stations opened across the U.S. between 1968 and 2017. I compare changes in outcomes for households located near newly opened rail stations to those of households that remained more distant. This staggered adoption difference-in-differences framework provides estimates of the causal effect of rail development on worker employment status and labor income.

I address the possibility of endogenous residential sorting in response to expanded rail infrastructure by using an intent-to-treat (ITT) approach. I define treatment status based on whether a household's location from two years prior received a station in the current year. As the choice of residential location two years before a station's opening is less influenced by the future transit amenity, this mitigates the concern that selection into treatment may be contaminating the results. Additionally, the ITT status can be used to instrument for actual treatment status to construct two-stage least squares (2SLS) estimates of the impacts.

The ITT results suggest that household heads living within 2 miles of a new rail station experience an average 2.73 percentage point increase in the probability of employment. Instrumenting for actual treatment status indicates a 3.68 percentage point increase. Positive employment effects are observed immediately following a station's opening, but increase slightly in magnitude until stabilizing at an elevated level three years after treatment. The change is accompanied by an insignificant average 28.5 log point (roughly 33 percent) increase in labor income (2SLS 38.4 log point) which also takes several years to fully materialize.[3] Interestingly, the employment effects are strongest in households that do not own a private automobile and tend to be located in neighborhoods

[3] 2017 dollars adjusted for inflation.

near the city center with a higher proportion low-skill residents.

Identification of the causal effect rests on the validity of a parallel trends assumption under which the outcomes of households that received a nearby rail station would have followed an identical trend to those of households that remained more distant. I investigate this assumption using an event study framework where I examine the existence of pre-treatment differences in outcomes between treated and untreated households. The estimates indicate no evidence of differential trends in outcomes prior to treatment. I also explore the robustness of the results to two econometric procedures which modify the sample so that comparisons are made between more similar households. The results are robust to each of these approaches.

I explore two mechanisms through which rail development may improve employment. First, the new transit amenity reduces the cost of travel between connected locations. As a result, transit-dependent workers can expand the geographic extent in which they search for employment and may face a more viable commute to previously difficult to access areas. The improved accessibility to centers of employment may shorten the duration of joblessness (Andersson et al. (2018)) and provide higher paying opportunities relative to before rail development. These impacts may be amplified by network effects through other individuals residing in the same neighborhood (Bayer, Ross and Topa (2008); Hellerstein, McInerney and Neumark (2011)).

Second, treated neighborhoods may experience significant changes in local economic conditions following the opening of a station. Prior literature documents a wide variety of impacts surrounding rail development including changes to the demographic composition (Glaeser, Kahn and Rappaport (2008); Heilmann (2018); Kahn (2007); McQuilkin (2020)), rental prices and housing values (Baum-Snow and Kahn (2000); Billings (2011)), and local business activity (Credit (2018); Schuetz (2015); Yen (2020)) in areas near new stations. If these changes favorably affect the local labor market faced by workers, then those residing near new transit may experience improved employment. In complementary analysis, I examine changes in local employment counts to explore this mechanism.

The ability for transit policy to improve employment has received increasing attention in recent years. Pang (2017) links subway infrastructure to increased labor force participation among low-skilled workers at the city-level. Tyndall (2019) shows that light rail expansion in four metropolitan areas increases employment at the census tract-level. Yen (2020) uses expansions across Los Angeles to show that labor force participation and employment improves at the tract-level and suggests that increased job density in treated areas is the source of the improvement. However, these prior studies employ cross-sectional data aggregated up to some spatial unit.

This paper provides three primary contributions to this literature. First, I employ person-level panel data which allows for a more detailed analysis of the underlying changes that drive the spatially aggregated estimates available from previous studies. While aggregated estimates suggest improved employment resulting from public transit infrastructure, it is possible the estimates are being driven by changes to neighborhood composition. If areas receiving rail access experience an influx of individuals with a higher probability of employment relative to incumbents, then aggregate estimates may be misinterpreted as rail access improving individual-level employment. With several studies suggesting changes in neighborhood demographic composition following rail development (Glaeser, Kahn and Rappaport (2008); Heilmann (2018); Kahn (2007); McQuilkin (2020)), this scenario is plausible. My results provide a more direct measure of how worker employment responds which may be more informative from a policy perspective.

Second, by including the opening of all commuter-oriented rail stations between 1968 and 2017, my estimates represent a comprehensive national analysis of the impacts of U.S. rail development. The more than 1,700 stations opened during this period represent roughly half of the stations currently open across the U.S. This builds on prior literature which has focused on a single metropolitan area (Heilmann (2018); Severen (2018); Tsivanidis (2018)) or a subset of those included in this study (Baum-Snow and Kahn (2000); Baum-Snow, Kahn and Voith (2005); Glaeser, Kahn and Rappaport (2008); Kahn (2007); McQuilkin (2020)). Furthermore, my estimates incorporate many of the most recent rail expansions which were excluded from prior studies. For this reason, the estimates may be more pertinent to policy discussion regarding future development.

Third, I investigate the relative importance of each proposed mechanisms using the person-level panel as well as administrative data on employment counts for each neighborhood over time. I document that local labor counts respond quite differently depending on the location of the neighborhood with those near the city center experiencing growth in the number of jobs, particularly in the food and services sector. This suggests that transit-oriented development around rail infrastructure may at least partially explain the observed employment effects.

The remainder of this paper is organized as follows. In Section 2.2, I describe the data. Section 2.3 details my empirical strategy. Section 2.4 presents the main results. Section 2.5 examines the results of several alternative and robustness specifications. Section 2.6 discusses the mechanisms. Section 2.7 concludes.

## 2.2 Data

To estimate the impacts of rail transit development, I utilize data derived from the Panel Study for Income Dynamics (PSID). This longitudinal survey, which spans a period of fifty years, contains a rich set of individual- and household-level characteristics that forms the basis for much of my empirical analysis. Beginning in 1968, the PSID included 4,802 households drawn from a nationally representative sample and an oversample of low-income families. These households, along with descendent households resulting from family members splitting to form new family units, were surveyed annually from 1968–1997 and biannually from 1999–2017. By the most recent wave in 2017, the survey had grown to 9,607 households.

The PSID includes information on employment, income, education, and demographic characteristics for members of each surveyed household. As the outcomes of the household head play a central role in determining a family's economic well-being, my empirical analysis focuses on this member.[4] The survey also includes household-level information on expenditure, asset ownership, mobility, and housing. Importantly, the restricted use PSID Geocode Match File identifies the census block for the residential location of each household during each wave.[5] This precise measure of location allows me to construct a proxy for rail transit access which is detailed below.

The PSID sample consists of 295,685 household-by-year observations for 31,280 unique households. With forty waves of the survey conducted between 1968-2017, an average of 7,392 households are observed in a given year. Figure 2.1 displays the number of households by year. In general, the sample grows larger over time as descendents from families split to form new households. Between 1990-1999, several refresher samples were added to improve representation among recent immigrant groups.[6] In 1996, roughly 2,000 households descended from the original sample were dropped due to funding limitations.

The average household appears in 10 waves with attrition rates ranging between 2.5-3 percent in a given wave. About half of the attrition is due to non-response, whereas the remaining half is split between death of an observation and the inability to track households after they move. While correlations are detected between demographic variables

[4]Household head is predominantly synonymous with primary earner.

[5]Location information is unavailable in the Geocode Match File for 1969, as well as even years 1998–2016 due to the biannual nature of the survey during this period. During these years, I employ the PSID Between-Wave Moves File to identify each household's residential location as described in the online appendix.

[6]The Latino sample added in 1990 was dropped in 1995 due to funding limitations.

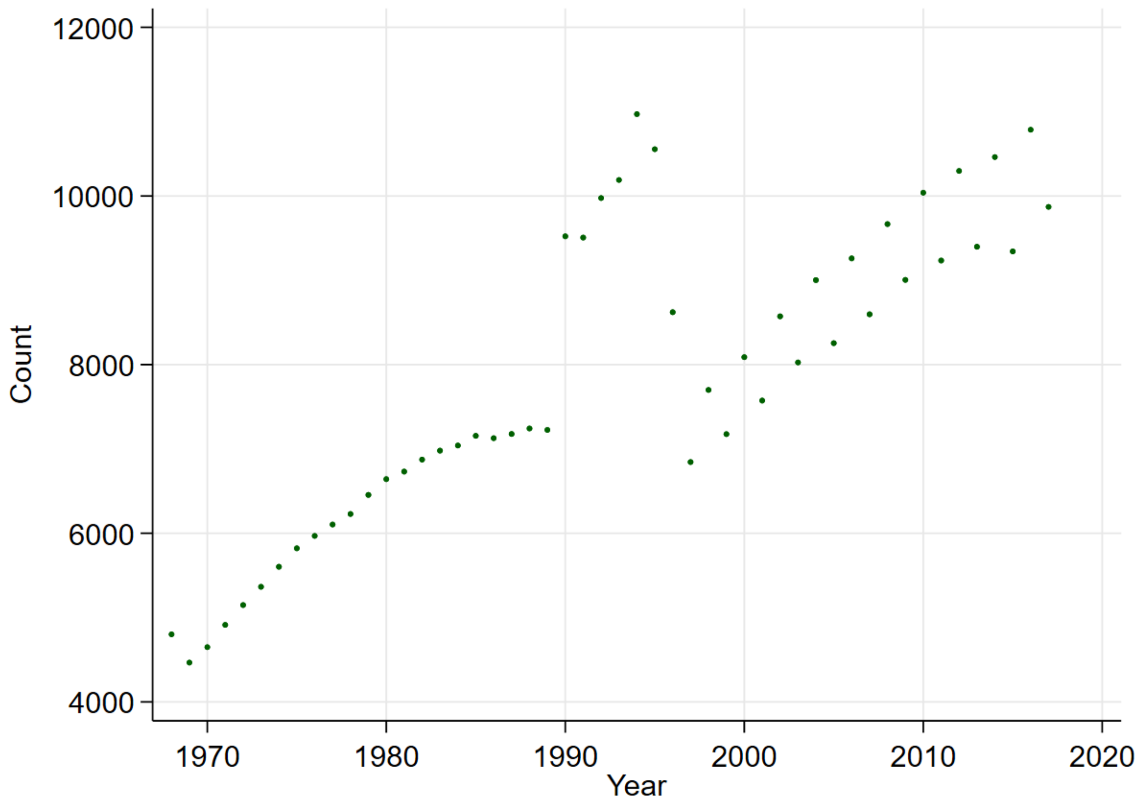



Figure 2.1: Number of Households By Year

In 1990, roughly 2,000 Latino households were added to improve representation among major immigrant groups. However, a lack of funding quickly lead to these households being dropped in 1995. In 1997, about 2,200 households derived from the original 1968 sample and their descendents were dropped due to high growth among this group. In 1997 and 1999, about 500 immigrant families were added to improve representation among immigrants.

and the probability of attrition, the predictive power is relatively low. This suggests that selective attrition is unlikely to introduce significant bias to my estimates (Fitzgerald, Gottschalk and Moffitt (1998)).

For neighborhood-level characteristics, I merge the PSID with GeoLytics Neighborhood Change Database 1970–2010 (NCDB). Derived from the 1970, 1980, 1990, and 2010 Decennial Census Summary Files along with the 2006–2010 American Community Survey 5-Year Estimate Summary Files, these data contain demographic and housing characteristics aggregated to the census-tract level. I merge the PSID to the NCDB based on the most recent past decennial year (e.g. PSID years 1980–1989 will have characteristics from the 1980 NCDB).[7] These tract-level characteristics serve to help explore the differential impacts on households based on the type of neighborhood they reside in.

Two sources of spatial data are required to assign treatment status. First, I construct a data set that includes the exact geographic coordinates and date of opening for all commuter rail stations currently open across the U.S.[8] The data is compiled using information scraped from Wikipedia and supplemented with details collected from local transit authorities. The spatial component of each station is compared with General Transit Feed Specification files provided by each transit authority to ensure accuracy.

The data is comprehensive, containing all 3,460 stations currently open across the U.S.[9] Figure 2.2 presents a map providing a geographic depiction of the data. Panel (a) displays stations open in 1968, while Panel (b) includes those open by 2017. The difference is striking. In 1968, extensive rail systems were restricted to cities in the Northeast and Chicago, with a very limited number across the rest of the country. By 2017, the U.S. had seen major development including throughout many of the major cities on the West Coast along with several cities in the Midwest, Mountain States, and Sun

[7]For PSID years 1968 and 1969, I use characteristics from the NCDB 1970.
[8]This includes light rail, heavy rail, subway, rapid transit, and commuter rail.
[9]As of 2017.

Belt. In addition, already large systems in Chicago and across the Northeast experienced considerable expansion during this period. Table 2.1 describes the number of stations opened within each city between 1968 and 2017. During this period, 1,717, or roughly half of currently open stations, opened for public use. Across the 32 cities that expanded rail transit during this time, the largest expansions occurred in Washington DC, Los Angeles, and the Greater Bay Area.

Figure 2.3 summarizes the evolution of rail transit within each metropolitan area by plotting the number of cumulative open stations within each city by year. The panels are arranged in decreasing order such that metro areas with the largest systems (as measured by the number of open stations by 2017) are found in Panel (a) while the smallest systems appear in Panel (d). The variation in timing and the number of stations is diverse across the different metropolitan areas. In general, rail systems tend to experience an increase in the number of stations every 5-15 years. These correspond to new rail lines, or expansions on older lines, which provide access to neighborhoods not previously served. Many cities have expanded their systems several times, while some have yet to expand their systems since they were initially opened.

The second source of spatial data includes the geographic boundaries for each of the more than 11 million 2010 U.S. Census blocks.[10] For each block, I calculate the spatial centroid and measure its distance to the nearest open rail station for each year. This distance is merged with the PSID using block identifiers, providing a detailed measure of rail access for each household by year.

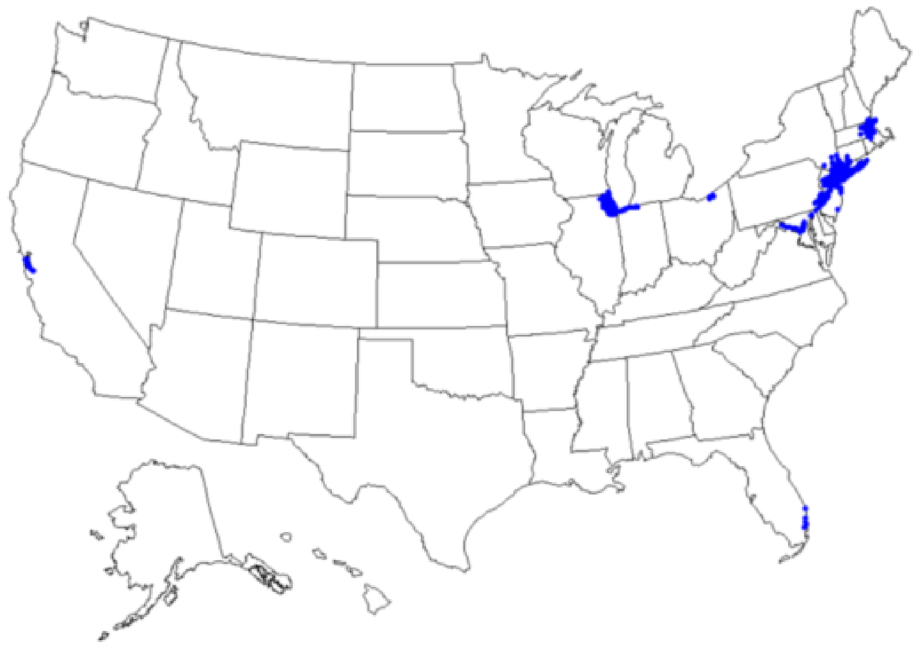
An additional source of spatial data is derived from Holian and Kahn (2015) which defines estimated coordinates for the central business district (CBD) of each U.S. metropolitan statistical area. These coordinates represent the focal point of commercial activity

[10] A census block is the smallest unit by which the U.S. Census tabulates population data. In urban areas, it is roughly equivalent to a city block, divided on each side by roads.

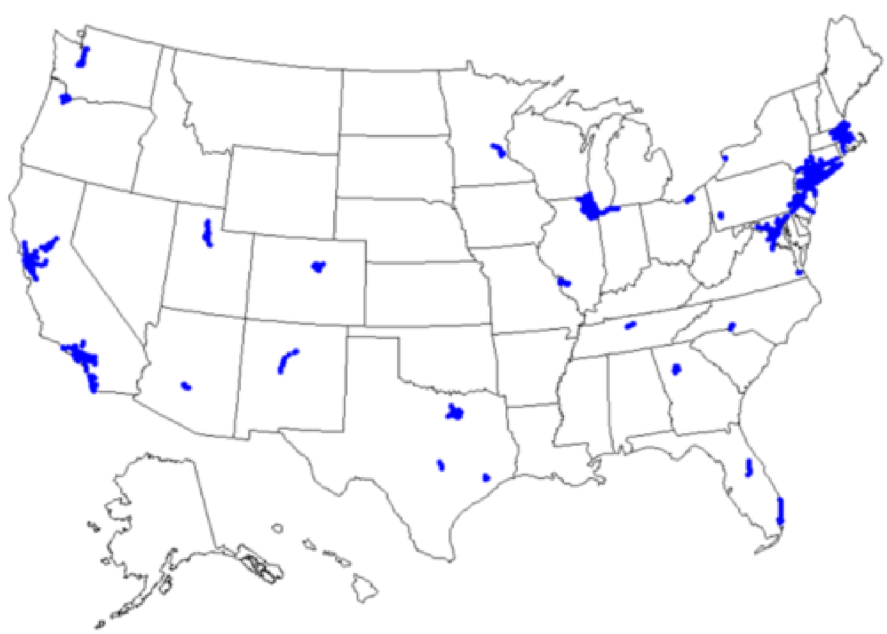
Table 2.1: Rail Systems By Metropolitan Area

| | Stations | First Year | Last Year |
|---|---|---|---|
| Albuquerque/Santa Fe | 15 | 2006 | 2017 |
| Atlanta | 38 | 1979 | 2000 |
| Austin | 9 | 2010 | 2010 |
| Baltimore | 45 | 1983 | 1998 |
| Boston | 88 | 1971 | 2017 |
| Buffalo | 14 | 1984 | 1986 |
| Charlotte | 20 | 2007 | 2015 |
| Chicago | 87 | 1969 | 2017 |
| Cleveland | 11 | 1968 | 2015 |
| Dallas | 80 | 1996 | 2016 |
| Denver | 61 | 1994 | 2017 |
| Greater Bay Area | 190 | 1972 | 2017 |
| Greater NYC Area | 76 | 1968 | 2017 |
| Greater Philadelphia Area | 75 | 1968 | 2016 |
| Houston | 39 | 2004 | 2017 |
| Los Angeles | 154 | 1990 | 2017 |
| Miami | 36 | 1984 | 2012 |
| Minneapolis | 44 | 2004 | 2014 |
| Nashville | 6 | 2006 | 2006 |
| New Haven | 15 | 1990 | 2013 |
| Norfolk | 11 | 2011 | 2011 |
| Oceanside/Escondido | 15 | 2008 | 2008 |
| Orlando | 12 | 2014 | 2014 |
| Phoenix | 40 | 2008 | 2016 |
| Pittsburgh | 52 | 1984 | 2012 |
| Portland | 102 | 1986 | 2015 |
| Sacramento | 52 | 1987 | 2015 |
| Salt Lake City | 66 | 1999 | 2013 |
| San Diego | 61 | 1981 | 2005 |
| Seattle | 34 | 2000 | 2016 |
| St. Louis | 37 | 1993 | 2006 |
| Washington DC | 132 | 1976 | 2015 |
| Total | 1717 | 1968 | 2017 |

Column (1) describes the number of stations opened between 1968 and 2017 for each metropolitan area. Columns (2) and (3) display the year of the earliest and most recent expansion, respectively, during this time period. The last row presents the information for the total of all metropolitan areas.



(a) Open Stations: 1968



(b) Open Stations: 2017

Figure 2.2: Number of Households By Year

Each dot represents an open public commuter rail station.

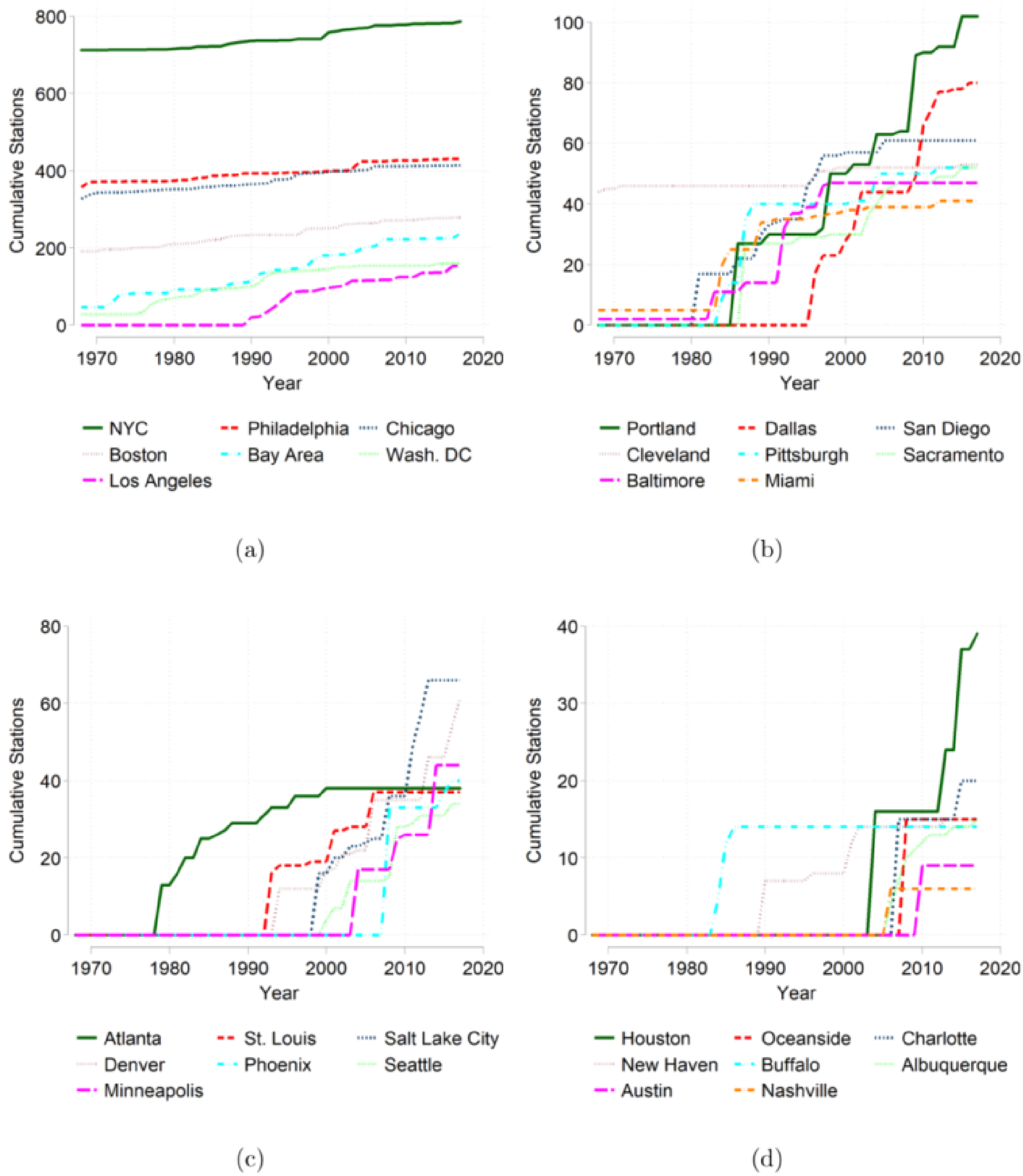



Figure 2.3: Number of Households By Year

Cumulative number of public commuter rail stations by metropolitan area. Panels arranged in decreasing order by metropolitan transit system size.

within each metropolitan area. Using this measure, I am able to explore how the effects of rail development differ based on a household's distance to the city center.

Table 2.2 presents summary statistics describing characteristics of each household-by-year observation in the PSID. The number of observations varies depending on missing values. In particular, commuting information is available only for 1968–1986 while car ownership is available for 1968–1986 and 1999–2017, resulting in far fewer observations for each of these variables. However, for the primary outcomes on employment and income, the degree to which missing values affect the sample size is minor.

Approximately 16 percent of the sample lives within 2 miles of a station throughout the duration of the study. On average, household heads earn $38,516 in labor income each year and work 1572 hours.[11] These measures are slightly lower than the overall U.S. average during this time period, a result of the oversample of low-income families. My primary measure of employment is an indicator that equals 1 if the worker earned a strictly positive annual income and zero otherwise. This measure allows for the possibility that rail access can affect those outside of the labor force. By this measure, 78 percent of household heads are employed throughout the study period.

Household heads are on average 44 years old with just over 12 years of education. 69 percent are male and 56 percent are married. 34 percent are black, much higher than the overall U.S. population. Based on information from 1968–1986, roughly 81 percent of household heads commute to work by private automobile while 10 percent use public transit which is roughly consistent with urban commuting outcomes across the U.S. Each household is located an average 17.32 miles from a city CBD and tend to live in neighborhoods that have a higher proportion of black residents and those below the poverty line relative to the national average for urban areas.

Complementary analysis is conducted using the Longitudinal Employer-Household

[11] All monetary variables in 2017 inflation adjusted dollars.

Table 2.2: Descriptive Statistics

| Variable | Obs. | Mean | Std.Dev. |
|---|---|---|---|
| $<$ 2 Miles of Station | 295,685 | 0.16 | 0.37 |
| Employed (Annual Labor Income $>$ 0) | 294,383 | 0.78 | 0.41 |
| Annual Labor Income | 294,383 | 38,516 | 57,958 |
| Hourly Wage | 281,963 | 19.48 | 25.74 |
| Annual Hours Worked | 294,382 | 1572 | 1,058 |
| Age | 298,329 | 43.73 | 16.63 |
| Male | 298,383 | 0.69 | 0.46 |
| Black | 296,217 | 0.34 | 0.48 |
| Married | 298,372 | 0.56 | 0.50 |
| Educational Attainment | 258,643 | 12.28 | 2.97 |
| Rental Price | 118,024 | 633 | 474 |
| Renter | 281,028 | 0.44 | 0.50 |
| Moved | 293,443 | 0.23 | 0.43 |
| Owns Car | 200,474 | 0.80 | 0.40 |
| Number Cars | 189,386 | 1.41 | 1.08 |
| Distance to CBD | 292,720 | 17.32 | 17.44 |
| Proportion Black | 254,861 | 0.26 | 0.34 |
| Proportion $<$ High School | 254,849 | 0.32 | 0.20 |
| Proportion Below Poverty | 254,861 | 0.16 | 0.13 |
| Commute: Public Transit | 68,232 | 0.10 | 0.30 |
| Commute: Auto | 68,232 | 0.81 | 0.39 |
| Commute: Walk | 68,232 | 0.05 | 0.22 |
| Commute: Miles | 68,899 | 11.03 | 12.55 |
| Commute: Minutes | 71,163 | 43.45 | 35.40 |

Includes all household-by-year observations in PSID. Individual-level variables summarized for household head. Income, wage, and rental price in 2017 dollars. Employed defined by whether individual has annual labor income greater than zero. Neighborhood level proportions black, below poverty, and less than high school taken from most recent decennial census associated with the year of each observation. Trip to work information available between 1969–1986 (commute mode also missing 1982). Car ownership information available between 1968–1986 and 1999–2017.

Dynamics Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics (WAC) files. These data contain administrative counts of workers by census block of employment. This information is used to evaluate how local labor markets respond to the introduction of a station, allowing examination of the neighborhood change mechanism.

## 2.3 Methods

My objective is to estimate the impact of rail transit development on individual-level employment outcomes. Using the opening of all public, commuter rail stations across the U.S. between 1968 and 2017, I exploit variation in rail access within households over time. I compare changes in outcomes for households that became closer than 2 miles of a rail station to those of households that remained more distant. Variation in treatment status arises due to new rail stations opening in neighborhoods where households were previously located.

Identification requires that households remaining farther than 2 miles from a station serve as a suitable counterfactual for treated households, had they never received treatment. I examine the validity of this parallel trends assumption by estimating pre-treatment differences in outcomes between treated and untreated households using an event study empirical specification characterized by Equation 2.1.

$$y_{it} = \sum_{s=-T}^{T} \beta_{\tilde{t}} \times 1(\tilde{t} = s) \times 1(\text{treated}_i) + \delta_i + \gamma_t + \epsilon_{it} \tag{2.1}$$

$y_{it}$ denotes the outcome of interest for household $i$ at year $t$. $\tilde{t}$ represents the year relative to the introduction of a rail station for treated households. $\tilde{t} = -1$ denotes the final year in which a household in the treatment group resided farther than 2 miles from a station.

$\tilde{t} = 0$ denotes the first year in which the family resides near a station. The excluded year is $\tilde{t} = -1$. $1(\text{treated}_i)$ is an indicator for whether household $i$ is part of the treatment group. $1(\tilde{t} = s)$ is an indicator for the year relative to treatment, where $s$ ranges between $-T$ and $T$. Treated observations outside of these bounds are excluded as the number of observations becomes increasingly limited for households farther from $\tilde{t} = 0$. $T = 5$ is used throughout my primary analysis.

$\delta_i$ and $\gamma_t$ are household and year fixed effects, respectively. The inclusion of household fixed effects controls for time-invariant household characteristics that may affect $y_{it}$, so that identification comes from comparisons within the same family unit as their access to rail transit changes. Year fixed effects control for annual shocks that effect the outcomes of all households. $\epsilon_{it}$ represents the error term and is clustered at the household-level.

The coefficients of interest, $\beta_{\tilde{t}}$, describe the difference in outcomes between the treatment group and the control group in relative year $\tilde{t}$, conditional on household and year fixed effects. In order to evaluate the validity of the parallel trends assumption, I test for joint significance of the pre-treatment coefficients $H_0: \ \beta_{-5} = \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$. Rejection of the null suggest that households in each group are experiencing differential trends in outcomes prior to the introduction of a rail station. A failure to reject is consistent with the assumption of parallel trends.

An additional benefit to estimating Equation 2.1 is the ability to examine the dynamic effects of receiving a rail station. The coefficients $\beta_0, \ \beta_1, \ ... \ , \ \beta_5$ describe differences in outcomes in the years following the opening of a station. Estimating these parameters provides detail on the time by which outcome $y_{it}$ adjusts following treatment. As the mechanisms by which rail development impact labor market outcomes may take time to materialize, these coefficients shed light on the potential for delayed effects.

Equation 2.2 characterizes the general form of my staggered adoption difference-in-differences model.

$$y_{it} = \beta D_{it} + \delta_i + \gamma_t + \epsilon_{it} \tag{2.2}$$

The coefficient of interest, $\beta$, describes the relative change in outcomes between the treatment and control group following the introduction of a rail station. In order for $\beta$ to identify the causal effect of rail development, I require the parallel trends assumption evaluated with Equation 2.1 be satisfied.

I also require that the timing of a station's opening does not spuriously correlate with other sharp neighborhood changes that could contaminate the difference-in-differences estimates. Any neighborhood change that occurs perfectly contemporaneously with the opening of a station represents a threat to identification. However, the event study pre-trend estimates measure differences in outcomes up to the year immediately prior to the station's opening. Combined with the widespread, and sometimes extreme, changes to the exact date of opening of rail lines due to unexpected financial, engineering, or political complications, it is difficult to imagine such sharp changes that are not a consequence of expanded rail infrastructure.

To better address endogenous sorting of households into treatment, my primary estimates of Equation 2.2 use an intent-to-treat (ITT) definition of treatment status. In order to define whether household $i$ is treated at time $t$, I focus on the location of residence from time $t-2$. Using this location, I determine whether a rail station was introduced within 2 miles at time $t$, regardless of whether household $i$ has moved to a new location. A household's residential location two years prior to the introduction of a rail station is unlikely to have been selected based on a joint relationship of receiving rail access and the observed employment outcome in the current year. As a result, this measure of treatment status reduces the likelihood that endogenous residential sorting could be biasing the results.

This ITT definition of treatment status is a suitable instrument of actual treatment status. Due to the strong relationship between a household's location two years prior and that in the current year, the ITT instrument satisfies the requirement of being relevant. In order to be valid, it is required that the instrument can only affect changes in outcomes in the years following rail development through its relationship with actual treatment status. The second requirement is not testable, but it is difficult to envision how prior residential location could influence a change in employment in the year in which a rail station is introduced outside of being correlated with actual treatment.

## 2.4 Employment Effects

### 2.4.1 Main Results

Figure 2.4 presents event study estimates using Equation 2.1 to explore the impact of rail development on employment status. Employment is measured as an indicator equal to one if the household head earns a positive annual labor income, and zero otherwise. A household is defined as treated the year it falls closer than 2 miles to a new rail station.

The horizontal axis represents the year relative to treatment for each treated household. Each point in the figure corresponds to $\hat{\beta}_t$, the estimated difference in outcomes between the treated and untreated group, conditional on household and year fixed effects. I include a 95 percent confidence interval for each estimate, illustrated with an error bar. All estimated differences are relative to that of the reference year, one year prior to the station's opening.

The number above each point estimate provides a count for the number of treated households that are used to identify the associated coefficient. For example, there are 371

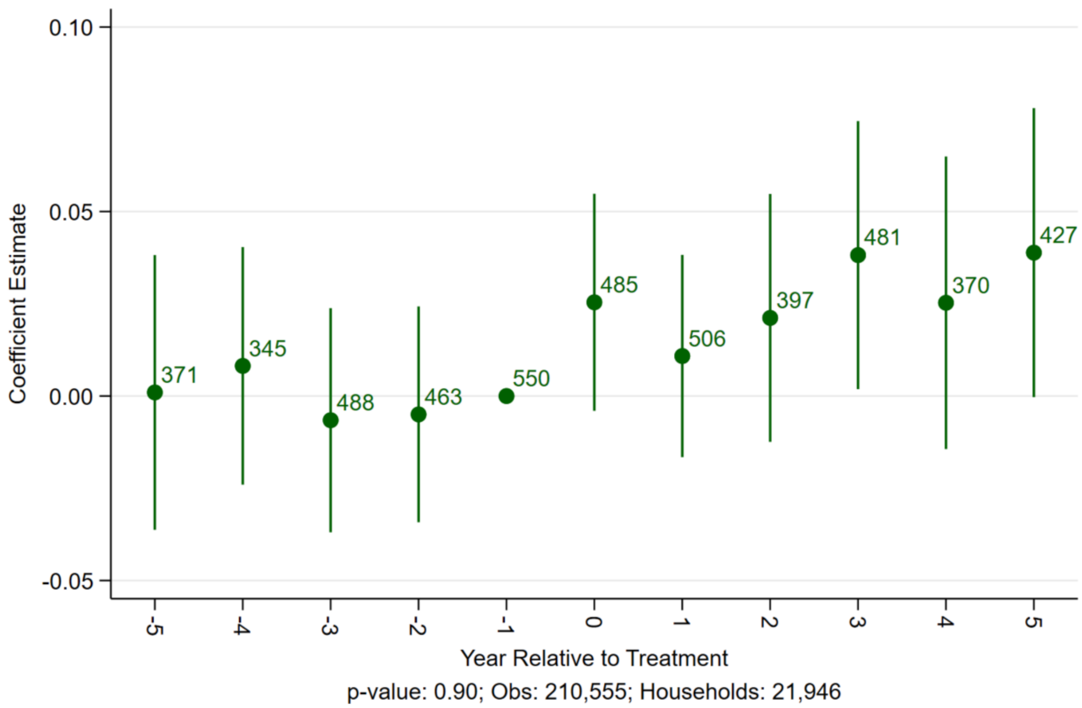



Figure 2.4: Event Study: Employment Status

Employment status defined by whether annual labor income is strictly positive. Estimated coefficients and 95% confidence intervals $\beta_{\tilde{t}}$ in Equation 2.1. Standard errors clustered at the household level. Outcome represents whether the household head earns positive labor income. p-values for the joint significance of the pre-treatment coefficients ($H_0 : \beta_{-5} = \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$) are found below each plot. The excluded relative year category is $\tilde{t} = -1$. Points are labeled by the number of observations falling into the treatment group for the associated relative year.

treated households with an observed employment status 5 years prior to their station's introduction. In years close to treatment, the number of treated observations is larger (484-548). As the relative year becomes farther from zero, the number of treated observations associated with each coefficient estimate decreases. Due to the biannual nature of the PSID following the 1997 wave, the number of treated households decreases monotonically in alternating years pre- and post-treatment. I investigate the robustness of the estimates to improving the balance of the panel in Section 2.5 by making requirements on the minimum number of pre- and post-treatment observations.

Below the event study figure, I include the p-value associated with a test for joint significance of the pre-trend estimates: $H_0 : \beta_{-5} = \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$. This allows for formal evaluation of the parallel trends assumption required for causal interpretation of the results. I also include the number of observations and the number of households.

Prior to treatment, there is little evidence of differential trends between groups. All estimated confidence intervals include zero and there is no apparent visual trend in the point estimates. The p-value associated with joint significance is equal to 0.90, providing support for the parallel trends assumption. Following the introduction of a rail station, treated household heads see an immediate increase in the probability of employment. In years zero through two, the estimates range between 1-2.5 percentage points (p.p.). For years three through five, treated households are roughly 2.5-4 p.p. more likely to be employed relative to households in the control group.

I estimate Equation 2.2 using the intent-to-treat definition of treatment status described in Section 2.3. By using household location from two years prior to the station's opening, this approach mitigates concerns of endogenous residential sorting. Column (1) of Panel A in Table 2.3 presents the difference-in-differences ITT estimate. The result suggests that rail access resulting from the opening of a new station leads to a statistically significant 2.73 p.p. increase in the probability of employment. In Column (2), I use the

ITT status as an instrument for actual treatment status. The first stage estimate can be found in Panel C, which shows a strong relationship between ITT status and actual treatment status (F-statistic equal to 2,777). The IV estimate suggests that rail access leads to a significant 3.68 p.p. increase in the probability of employment.

Figure 2.5 displays the estimated event study coefficients for the impact on the log of annual labor income. The pre-treatment coefficients suggest no trend in outcomes prior to the introduction of a station. All estimates are not statistically distinct from zero and the p-value for joint significance equals 0.74, providing no evidence against the parallel trends assumption. The post-treatment coefficients match the same upward trend observed in Figure 2.4. There is a small apparent increase in log labor income immediately following treatment, and the impact increases slightly until stabilizing in years three through five. Panel B of Table 2.3 presents the ITT and IV estimates. The ITT estimate suggests that treatment leads to a statistically significant 28.5 log point (33.1 percent) increase in labor income. The IV approach increases this to 38.4 log points (47.2 percent). The estimates appear to be a reflection of the employment effects described above.

If rail infrastructure is to improve the well-being of targeted workers, the observed increase in employment probability and labor income from past rail development is encouraging. The ITT and IV point estimates are meaningful, suggesting that workers near new stations experience employment gains comparable to those of other place-based policies Freedman (2013); Ham et al. (2011). It is noteworthy that the effects appear immediately following treatment, but continues to increase in subsequent years. The mechanisms, whether increased geographic mobility or neighborhood change, could plausibly take time to fully affect workers. For example, a worker may initially not understand the full extent of the station's amenity value as they familiarize themselves with the expanded set of locations accessible through the system. Alternatively, it might take years for the local labor market to achieve equilibrium following a station's introduction as

Table 2.3: Employment and Income Estimates

| | (1) ITT | (2) IV |
|---|---|---|
| Panel A: Employment Status | | |
| $\leq$ 2 miles | 0.0273** | 0.0368** |
| | (0.0112) | (0.0152) |
| Observations | 195,666 | 195,327 |
| Households | 20,683 | 20,645 |
| Panel B: Log Labor Income | | |
| $\leq$ 2 miles | 0.285** | 0.384** |
| | (0.116) | (0.157) |
| Observations | 195,689 | 195,350 |
| Households | 20,692 | 20,654 |
| Panel C: First Stage | | |
| $\leq$ 2 miles | | 0.743*** |
| | | (0.0140) |
| F-Stat | | 2,777 |
| Observations | | 195,530 |
| Households | | 20,654 |

Panels A and B display estimated staggered adoption difference-in-differences estimates using intent-to-treat and 2SLS empirical specifications of Equation 2.2. Employment status defined by whether annual labor income is strictly positive. Labor income in 2017 dollars. Panel C shows the estimated first stage including the F-statistic associated with significance of the instrument in the first stage.

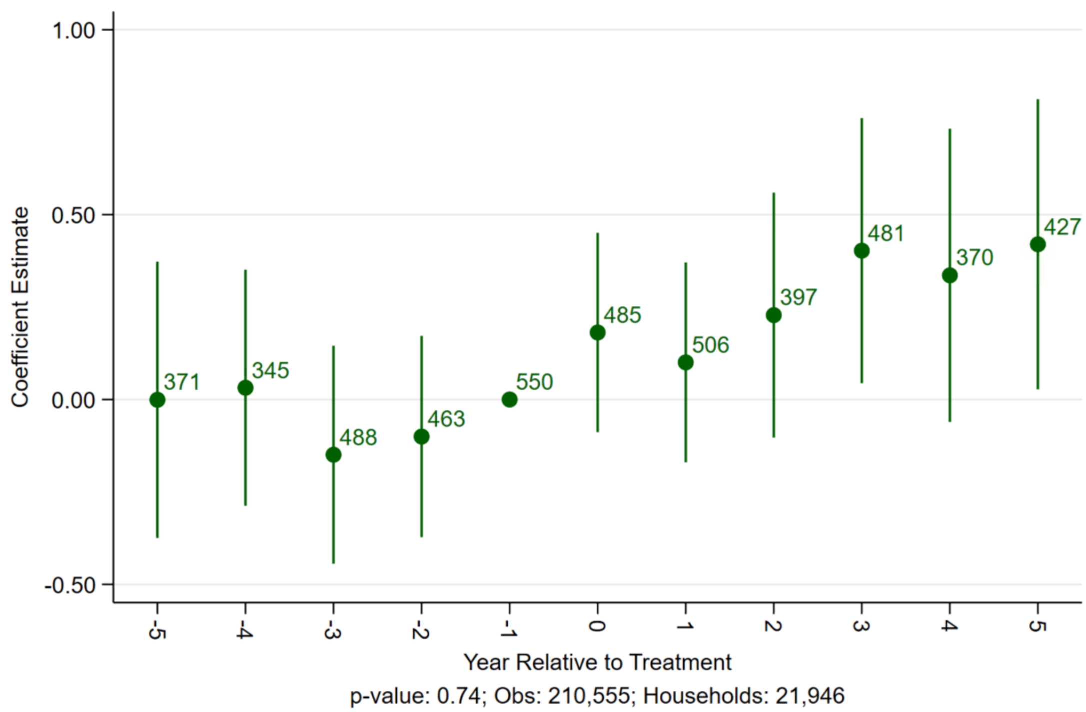



Figure 2.5: Event Study: Log Labor Income

Log of labor income measured in 2017 inflation adjusted dollars. Estimated coefficients and 95% confidence intervals $\beta_{\tilde{t}}$ in Equation 2.1. Standard errors clustered at the household level. Outcome represents whether the household head earns positive labor income. p-values for the joint significance of the pre-treatment coefficients ($H_0 : \beta_{-5} = \beta_{-4} = \beta_{-3} = \beta_{-2} = 0$) are found below each plot. The excluded relative year category is $\tilde{t} = -1$. Points are labeled by the number of observations falling into the treatment group for the associated relative year.

new businesses open and adapt to changes in local demand. In Section 2.6, I explore additional evidence regarding each of these potential mechanisms.

### 2.4.2 Effects By Household

With equity playing a central focus of the American Jobs Plan, it is important to understand the types of households that have benefited from past rail development. The spatial mismatch literature postulates that minority households and low-skilled workers are most likely to be affected by limited job access. However, while these populations are indeed disproportionately heavy public transit users, much of this use comes from bus as opposed to rail ridership (Clark (2017)).

In Table 2.4, I estimate Equation 2.2 restricting the sample based on race and educational attainment. Column (1) shows the estimate using the full sample. Columns (2) and (3) display those for black and non-black workers, respectively. While the point estimates on employment status and labor income for black workers are higher relative to non-blacks, the difference is insignificant. Columns (4) and (5) describe the effects by whether the worker has a high school diploma. Again, there is no apparent difference in the employment impact based on education. The estimates by subgroup suggest a roughly 2-3 p.p. increase in employment probability and 25-35 log point increase in labor income, across each of these demographic characteristics.

Table 2.4: Employment and Income Estimates - Household Subgroups

| | (1) Full Sample | (2) Black | (3) Non-Black | (4) < H.S. | (5) $\geq$ H.S. | (6) No Car | (7) Car |
|---|---|---|---|---|---|---|---|
| Panel A: Employment Status | | | | | | | |
| $\leq$ 2 miles | 0.0273** | 0.0302* | 0.0221 | 0.0308 | 0.0308*** | 0.0964* | 0.0180 |
| | (0.0112) | (0.0175) | (0.0140) | (0.0263) | (0.0114) | (0.0578) | (0.0168) |
| Observations | 195,666 | 62,381 | 133,260 | 55,550 | 138,777 | 17,975 | 116,985 |
| Households | 20,683 | 6,723 | 13,941 | 5,580 | 14,619 | 2,073 | 11,327 |
| Panel B: Log Labor Income | | | | | | | |
| $\leq$ 2 miles | 0.285** | 0.280 | 0.257 | 0.268 | 0.343*** | 0.826 | 0.155 |
| | (0.116) | (0.176) | (0.152) | (0.267) | (0.120) | (0.552) | (0.186) |
| Observations | 195,666 | 62,381 | 133,260 | 55,550 | 138,777 | 17,975 | 161,985 |
| Households | 20,683 | 6,723 | 13,941 | 5,580 | 14,619 | 2,073 | 11,327 |

Panels A and B display estimated staggered adoption difference-in-differences estimates using the intent-to-treat specification of Equation 2.2. Employment status defined by whether annual labor income is strictly positive. Labor income in 2017 dollars. Columns (2) and (3) restrict to households with black versus non-black head, respectively. Columns (4) and (5) restrict to household heads with no high school diploma versus those with a high school diploma, respectively. Column (6) limits to households that do not in general own a car and requires the treatment group do not own a vehicle in the year they are treated. Column (7) limits to households that always own a car and requires treated households do own a vehicle in the year they are treated.

In Columns (6) and (7), I split the sample based on whether the household owns a private vehicle. As rail infrastructure may likely be particularly useful for those lacking the alternative option of driving to work, we might expect a particularly strong effect among those without a car. In Column (6), I restrict the sample to individuals that live in households with limited car ownership.[12] The point estimates on employment status and labor income are roughly triple those of the full sample. This suggests, perhaps unsurprisingly, that household car ownership is an important dimension by which employment impacts differ. The effects among those owning a car are small and indistinguishable from zero.[13] The results imply that transit investment can be helpful among households of differing demographics, but can be especially important to those constrained by a lack of car ownership.

### 2.4.3 Effects By Neighborhood

I next explore the effects depending on characteristics of the neighborhood receiving the station. For policymakers that must decide where rail infrastructure should be placed, these results describe how residents of certain types of neighborhoods have been differentially affected by past development. Table 2.5 documents the impacts on subgroups of the sample based on neighborhood-level characteristics. In Column (2), I estimate Equation 2.2 for households predominantly residing within 5 miles from the city central business district.[14] The point estimate for these inner-city neighborhoods is higher (although not statistically different) than the effect in the full sample. This especially true in Panel A which displays the effects on employment status. Column (3) restricts to households

[12]Car ownership changes over time, so I allow limited car ownership. Non-Car Owners are 1) households that own a car in less than 25% of survey waves they are observed and 2) only those that do not own a car in the year that they are treated.

[13]Car Owners are 1) households that own a car in every survey wave they are observed and 2) only those that own a car in the year that they are treated.

[14]Stations are also located within 5 miles from the CBD.

located between 5-20 miles from the city CBD. The effect is roughly equal to that of the full sample. The results provide some evidence of a stronger employment effect in inner-city locations.

Table 2.5: Employment and Income Estimates - Neighborhood Subgroups

| | (1) Full Sample | (2) Inner-City | (3) Outer-City | (4) Black | (5) Non-Black | (6) $<$ H.S. | (7) $\geq$ H.S. |
|---|---|---|---|---|---|---|---|
| Panel A: Employment Status | | | | | | | |
| $\leq$ 2 miles | 0.0273** | 0.0406* | 0.0273* | 0.0295* | 0.0156 | 0.0421** | 0.0116 |
| | (0.0112) | (0.0227) | (0.0146) | (0.0158) | (0.162) | (0.0168) | (0.0145) |
| Observations | 195,666 | 34,096 | 86,723 | 88,783 | 94,814 | 93,871 | 89,463 |
| Households | 20,683 | 3,916 | 9,413 | 9,754 | 9,494 | 9,315 | 9,882 |
| Panel B: Log Labor Income | | | | | | | |
| $\leq$ 2 miles | 0.285** | 0.359 | 0.329** | 0.271* | 0.182 | 0.431** | 0.132 |
| | (0.116) | (0.230) | (0.149) | (0.182) | (0.174) | (0.170) | (0.151) |
| Observations | 195,666 | 34,096 | 86,723 | 88,783 | 94,814 | 93,871 | 89,463 |
| Households | 20,683 | 3,916 | 9,413 | 9,754 | 9,494 | 9,315 | 9,882 |

Panels A and B display estimated staggered adoption difference-in-differences estimates using the intent-to-treat specification of Equation 2.2. Employment status defined by whether annual labor income is strictly positive. Labor income in 2017 dollars. Column (2) restricts to households predominantly located in neighborhoods within 5 miles from the city central business district (CBD) and requires treated households are treated in a neighborhood within 5 miles from the CBD. Column (3) restricts to households predominantly located in neighborhoods analogously restricted to 5-20 miles from the CBD. Column (4) restricts to households predominantly located in neighborhoods with an above-median proportion of black individuals and requires treated households are treated in neighborhoods with above-median proportion black. Column (5) restricts to households analogously restricted to below-median black neighborhoods. Column (6) restricts to households predominantly located in neighborhoods with an above-median proportion of individuals with less than high school diploma and requires treated households are treated in neighborhoods with an above-median proportion of individuals with less than high school diploma. Column (7) restricts to households analogously restricted to below-median proportion less than high school neighborhoods.

In Columns (4) and (5), I split the sample based on the proportion of black residents. In Column (4), I require that households generally locate in neighborhoods with an above-median proportion black.[15] Column (5) presents estimates for below-median proportion black neighborhoods. The point estimate is larger for households predominantly located in black neighborhoods relative to non-black neighborhoods. Finally, I split the sample similarly based on the proportion of adults with a high school diploma. In Column (6), the point estimate suggests a large effect in neighborhoods with a larger share of non-diploma residents. The estimates in Column (7) suggest minimal impacts in neighborhoods with a higher share of high school graduates. Table 2.5 provides evidence that residents of neighborhoods located near the city center and neighborhoods with a larger fraction of low-skill workers are likely to see greater employment gains.

### 2.4.4 Effects Among Movers Versus Stayers

Table 2.6 breaks the estimates based on the number of years following treatment that the household continues to live in the same census block as treatment. In Column (2), among households that move within 2 years following a station's opening, the effect is largest (5.12 p.p. increase in employment probability). For households that remain for between 3-4 years, the effect diminished slightly (3.95 p.p.). Among households that continue to live in the same neighborhood 5 or more years subsequent to treatment, the effect becomes small and insignificant.

[15] Stations also open in above-median black neighborhoods.

Table 2.6: Employment and Income Estimates - Movers Vs. Non-Movers

| | (1)<br>Full Sample | (2)<br>Move ≤ 2 years | (3)<br>Move W/in 3-4 years | (4)<br>Stay ≥ 5 years |
|---|---|---|---|---|
| Panel A: Employment Status | | | | |
| ≤ 2 miles | 0.0273** | 0.0512*** | 0.0395** | 0.008 |
| | (0.0112) | (0.0144) | (0.0183) | (0.0161) |
| Observations | 195,666 | 192,739 | 192,076 | 192,841 |
| Households | 20,683 | 20,208 | 20,043 | 20,079 |
| Panel B: Log Labor Income | | | | |
| ≤ 2 miles | 0.285** | 0.509*** | 0.356** | 0.092 |
| | (0.116) | (0.146) | (0.191) | (0.170) |
| Observations | 195,666 | 192,739 | 192,076 | 192,841 |
| Households | 20,683 | 20,208 | 20,043 | 20,079 |

Panels A and B display estimated staggered adoption difference-in-differences estimates using the intent-to-treat specification of Equation 2.2. Employment status defined by whether annual labor income is strictly positive. Labor income in 2017 dollars. Column (2) restricts to treated households that moved from the rail neighborhood within 2 years of treatment. Column (3) restricts to households that remained in the rail neighborhood 3-4 years following treatment. Column (4) restricts to households that remained in the rail neighborhood at least 5 years following treatment.

This monotonic relationship between length in treatment block following treatment suggests that part of the observed employment response is due to those that do not continue to live near transit. However, it is important to note that the choice to continue living in the same block is endogenous to many other factors that correlate with economic well-being. For this reason, it is difficult to compare each of these estimates. However, if the sample in Column (4) in restricted to households without a private vehicle (not shown), the point estimate becomes nearly identical to that of the full non-car owner group in Column (6) of Table 2.4. Again, it appears that an important dimension to who benefits to rail development is the presence of alternative transportation options.

## 2.5 Robustness

Here I investigate the robustness of my main estimates to provide evidence that the effects captured in Section 2.4 truly represent the causal effects of rail transit.

### 2.5.1 Alternative Treatment Thresholds

For both employment and labor income, I explore the geographic extent of the impact by varying the distance threshold used to define treatment. First, I restrict the treated group to households that became within 1 mile of a station. Under this specification, I exclude those in the 1-2 mile distance ring from the control group as the employment outcomes among households in this distance range appear to be positively affected. Second, I use households that became between 2-3 miles of a station as the treatment group, while excluding those between 0-2 miles.

The estimates based on these alternative definitions of treatment and control are found in Table 2.7. Column (1) presents the main estimates using the 2 mile threshold. Column (2) presents the impact on households that fell within 1 mile of a station. Point

Table 2.7: Alternative Treatment Distance Threshold

| | (1) Main | (2) 1 Mile | (3) 2-3 Mile |
|---|---|---|---|
| Panel A: Employment Status | | | |
| $\leq$ X miles | 0.0273** | 0.0291* | .0149 |
| | (0.0112) | (0.0170) | (0.0128) |
| Observations | 195,666 | 196,293 | 183,603 |
| Households | 20,683 | 20,838 | 19,021 |
| Panel B: Log Labor Income | | | |
| $\leq$ X miles | 0.285** | 0.245 | 0.141 |
| | (.116) | (0.175) | (0.127) |
| Observations | 195,666 | 196,293 | 183,603 |
| Households | 20,683 | 20,838 | 19,021 |

Panels A and B display estimated staggered adoption difference-in-differences estimates using intent-to-treat specification Equation 2.2. Employment status defined by whether annual labor income is strictly positive. Labor income in 2017 dollars. Column (1) provides estimate using a 2 mile distance treatment threshold. Column (2) provides estimate using a 1 mile distance treatment threshold, excluding households between 1-2 miles from the control group. Column (3) provides estimates using a 2-3 mile treatment ring, excluding households between 0-2 miles from the treatment group.

estimates are roughly equal to those in Column (1). While the estimates lack power, they are still positive employment and labor income. For the 2-3 mile treatment ring, the point estimates are much smaller relative to Column (1) and not statistically different from zero. In fact, the estimate on labor income is slightly negative under this specification. The results suggest a limited impact of rail transit development farther than 2 miles from a new station.

### 2.5.2 Improved Comparability of Households

The estimates in Section 2.4 rely on a parallel trends assumption where the outcomes of treated households would have followed an identical trend to those of untreated households, in the absence of rail transit. While the event study pre-treatment estimates are consistent with this assumption, I apply two methodological techniques that alter the sample so that comparisons are made between households that are more similar based on observable characteristics.

First, I use a trimming procedure to construct a sample of treated and untreated households that appear more similar based on observable characteristics prior to treatment Crump et al. (2009). An indicator for whether the household ever falls within 2 miles of a new rail station is regressed on a set of demographic, household, and neighborhood characteristics using a logit regression model. Estimates from this model are used to calculate the propensity score of treatment. The sample is restricted to households with a propensity score between 0.1 and 0.9 and Equation 2.2 is estimated. The procedure drops households that appear either very likely or very unlikely to be treated (85 percent of the sample, primarily from the control group). The remaining households are much more similar based on pre-treatment observable characteristics, reducing the potential for omitted variable bias.

I also utilize an inverse probability weighting (IPW) procedure as in Hirano, Imbens and Ridder (2003). I weight treated households by the inverse of the propensity score, and untreated households by the inverse of one minus the propensity score. The procedure results in a sample where treated households that initially look like untreated households and untreated households that initially look like treated households are weighted more heavily.

Panel (a) of Figure B.1 presents the estimated impact of rail transit on employment status from the main specification along with each robustness procedure. The point estimate and 95 percent confidence interval are shown for each. The first row displays the estimate from the main specification, while subsequent rows shows those from each robustness procedure. The trimmed and IPW point estimates are very similar to that of the main specification, and both are statistically indistinguishable. Panel (b) of Figure B.1 displays the same figure for labor income. Again, the point estimates are quite similar to those in the main specification suggesting a positive effect of rail access on employment. I conclude that the estimates from each improved comparability procedure are consistent with those from Section 2.4.

### 2.5.3 Improved Panel Balance

Due to the staggered treatment of households across years, the varied duration in which households are observed, and the biannual nature of the survey from 1997–2017, the estimates in Section 2.4 are identified using an unbalanced panel. While all treated households are observed in at least one prior year and one subsequent year following the opening of a rail station, the likelihood that a household is observed decreases in years farther from treatment. This is highlighted by the number of observations used to identify each point in Figures 2.4 and 2.5.

I address panel imbalance by estimating Equation 2.2 using a sample of tracts in which I require a certain number of pre- and post-treatment observations. Due to the biannual nature of the PSID in latter years, I make restrictions accounting for the fact that many households by design will not be surveyed every consecutive year pre- and post-treatment. These estimates are found in Rows 4 and 5 of each panel of Figure B.1. For Balanced Sample 1, I required that households were observed at least either 3 or 4 years pre-treatment, and 3 or 4 years post-treatment. For Balanced Sample 1, I required that households were observed at least either 4 or 5 years pre-treatment, and 4 or 5 years post-treatment. [16] While the standard errors increase under these restrictions, the point estimates are relatively similar to those using the main sample. I conclude that the main results are robust to using a more balanced panel.

### 2.5.4 Sample Weights

The PSID is constructed using two initial samples derived from the Survey Research Center and Survey of Economic Opportunity. Due to unequal selection probabilities for households between the two samples, and to address for selection of later immigrant additions to the survey, the PSID includes a set of sample weights. These weights address for both unequal selection probabilities across samples and differential attrition. Estimates using these weights are presented in Row 6 of each panel of Figure B.1. The application of these weights has no little impact on the estimates. The point estimate on employment status is virtually identical to the unweighted specification. For labor income, the point estimate and standard error is larger. However, the estimate is not statistically different from that in the main specification.

### 2.5.5 Wald Time-Corrected Estimates

The event study specification characterized by Equation 2.1 is susceptible to the problem of treatment effect heterogeneity and negative weighting described in de Chaisemartin and d'Haultfoeuille (2019). I apply a Wald-Time Corrected (WTC) estimator of the event study coefficients to examine the robustness of each estimate to this concern. Figure B.2 plots both the event study coefficients using Equation 1.2 on the left and the WTC event study estimates on the right. While the estimates vary somewhat

[16] It is extremely unlikely that a household skips a wave of the survey and returns in future years.

relative to those in the main specification, the results are qualitatively similar. Prior to treatment, there is no obvious differential trend in outcomes between the treatment and control group. Following the introduction of a rail station, treated households experience an increase in the probability of employment which levels off between 3-5 years following treatment.

Figure B.3 presents the sample plot for labor income. Prior to the stations introduction, the estimates are consistent with limited pre-trends. Following treatment, there is a fluctuation in the estimates before they stabilize at an elevated level between 3-5 years following treatment. Overall, the results are qualitatively similar to those in Section 2.4.

## 2.6 Discussion of Mechanisms

This section explores two proposed mechanisms underlying the estimated employment effects in Section 2.4. Under the mobility mechanism, workers benefit from the improved access to an expanded set of employment opportunities throughout the city. Alternatively, the neighborhood change mechanism posits that improved local labor market conditions near new stations help drive the employment effects.

### 2.6.1 Commuting Outcomes

Under the mobility mechanism, rail development acts to lower the cost of intracity travel, improving the ability for workers to access new centers of employment. If this mechanism is playing a role, we might expect to observe changes in worker commuting characteristics. The 1968–1986 PSID waves include information on commute mode (by public transit, private vehicle, and walking) as well as one-way distance traveled and one-way commute time. While this subset of years only partially covers the full study period, it is worth exploring these outcomes to evaluate the existence of the mobility mechanism.

If workers that had previously walked to work can now use the train, we might expect an increase in public transit commuting among affected households. Alternatively, former car-commuters may find rail service preferable to driving and might also increase public transit usage among treated households. However, if workers that had previously ridden the bus to work switched to rail, then this would not change their overall public transit commuting.

Panels A, B, and C of Table 2.8 present estimates of Equation 2.2 on an indicator for whether an individual primarily uses public transit, a private vehicle, or walks to get to work, respectively. Column (1) displays the estimate using the 2 mile treatment radius providing no evidence of changes to commute mode. In Column (2), I restrict the treatment group to households within 1 mile of a station. Again, there is no detectable change. The results suggest limited changes to these commuting measures.

It is difficult to speculate the reason for the insignificant commute mode response. One possibility is that the introduction of a rail station does not meaningfully affect individual commuting habits in the sample. An alternative is that new rail users had already commuted by bus. McQuilkin (2020) documents that increases in rail commuting following a station's introduction can extend up to 2 miles from a station and that reductions in bus commuting can extend even farther. In many neighborhoods, the increase in rail usage is completely offset by reductions in bus usage. In either case, the lack of evidence also reflects the small sample size after restricting the treatment group to those observed 1968–1986, prior to when many cities opened their first stations.

If former bus users do indeed switch to rail usage, then we might expect changes in commute distance or time. Rail service often travels at far higher speeds relative to bus which may reduce the time taken to commute a similar distance, or allow a worker to increase their feasible commute distance. Table 2.9 displays the estimated impact on

Table 2.8: Commute Mode

| | (1) ITT: 2 Mile | (2) ITT: 1 Mile |
|---|---|---|
| Panel A: Commute: Public Transit | | |
| $\leq$ X miles | -0.0179 | 0.0163 |
| | (0.0235) | (0.0277) |
| Observations | 47,534 | 47,385 |
| Households | 7,350 | 7,359 |
| Panel B: Commute: Auto | | |
| $\leq$ X miles | .0058 | $-0.0403$ |
| | (0.0292) | (0.0375) |
| Observations | 47,534 | 47,385 |
| Households | 7,350 | 7,359 |
| Panel C: Commute: Walk | | |
| $\leq$ X miles | -0.00062 | $-0.0118$ |
| | (0.0113) | (0.0163) |
| Observations | 47,534 | 47,385 |
| Households | 7,350 | 7,359 |

Each Panel display the estimated staggered adoption difference-in-differences estimates using intent-to-treat empirical specifications of Equation 2.2. Panels A and B describe the impacts on an indicator for whether the head's primary means of commute is public transit and private automobile, respectively. Panel C displays the estimated impact on the one way commuting distance in miles. Column (1) uses a 2 mile distance threshold to define treatment. Column (2) provides estimate using a 1 mile distance treatment threshold, excluding households between 1-2 miles from the control group.

Table 2.9: Commute Distance and Time

| | (1) ITT: 2 Mile | (2) ITT: 1 Mile |
|---|---|---|
| Panel A: Commute Distance | | |
| $\leq$ X miles | 1.413* | 0.734 |
| | (0.747) | (1.086) |
| Observations | 49,163 | 49,011 |
| Households | 7,256 | 7,265 |
| Panel B: Commute Time | | |
| $\leq$ X miles | -1.113 | $-5.561^{*}$ |
| | (2.177) | (3.272) |
| Observations | 49,583 | 49,405 |
| Households | 7,536 | 7,546 |

Each Panel display the estimated staggered adoption difference-in-differences estimates using intent-to-treat empirical specifications of Equation 2.2. Panels A and B describe the impacts on an indicator for whether the head's primary means of commute is public transit and private automobile, respectively. Panel C displays the estimated impact on the one way commuting distance in miles. Column (1) uses a 2 mile distance threshold to define treatment. Column (2) provides estimate using a 1 mile distance treatment threshold, excluding households between 1-2 miles from the control group.

one-way commute distance (miles) and time (minutes). Again, the small sample size is limiting. I find weak evidence that one-way commute distances increases in households within 2 miles from a new station and a marginally significant decrease in commute time for households located within 1 mile.

Overall, the trip to work information in the PSID makes it difficult to identify whether increase geographic mobility is underlying the observed employment effects. The limited sample size restricts my ability to examine changes in commuting characteristics by subgroup such as examining the change in commute distance only among public transit users. Presenting convincing evidence regarding this mechanism remains a gap that requires more extensive commuting information that may be available in other sources of data.

### 2.6.2 Local Labor Market Effects

Prior literature has documented a variety of changes in neighborhoods receiving a rail station. If these changes favorably affect the labor market from the workers' perspective, then residents of treated neighborhoods may experience improved employment. For example, increased foot-traffic resulting from rail ridership or other transit-oriented development may lead to the opening or expansion of establishments to capitalize on increased consumer demand. If this is the case, then food and restaurant or service industry jobs are those that may be particularly affected.

I investigate this possibility using the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics (WAC) files which include administrative employment counts for the annual number of workers employed within each census block between 2002–2017. Similarly to the discussion of commuting outcomes, these data only partially spans my full study period. However, the number of census blocks affected is large enough to better conduct statistical inference.

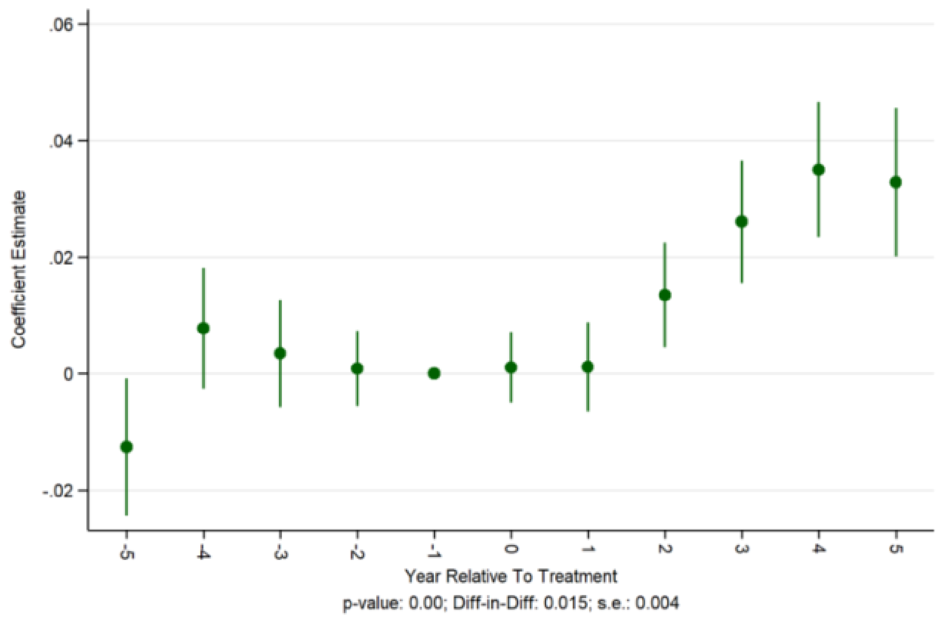
Figure 2.6 displays event study estimates to explore changes in log employment counts following a station's opening. A distance threshold of 1 mile is used to define treatment, and block fixed effects (as opposed to household fixed effects) are included. To address the neighborhood change mechanism, I include only jobs classified by the NAICS as accommodation and food service. In Panel (a), I restrict the sample to blocks that are located within 5 miles of the city CBD. While the p-value for joint significance is suggestive of significant pre-trends, much of this comes from 5 years prior to the station's opening. Beginning two years after a station has opened, there is a significant increase in the number of jobs that continues to grow until stabilizing four years after treatment. The difference-in-differences estimate suggests the opening of a station leads to a 1.5 percent increase in the number of jobs.

In Panel (b), I include blocks located 5–20 miles from the CBD. In this outer city sample, there appears to be a roughly equivalent decrease in the number of jobs. This distinct difference in effect between inner and outer city rail stations highlights the importance of location within city in evaluating the impacts of rail transit. In inner city locations, I find evidence of increased employment activity suggesting that the observed employment effects in Section 2.4 may be at least partially a result of neighborhood change. In particular, the difference between the estimates in Columns (2) and (3) of Panel A could be explained by the differential impacts identified in the LODES data.

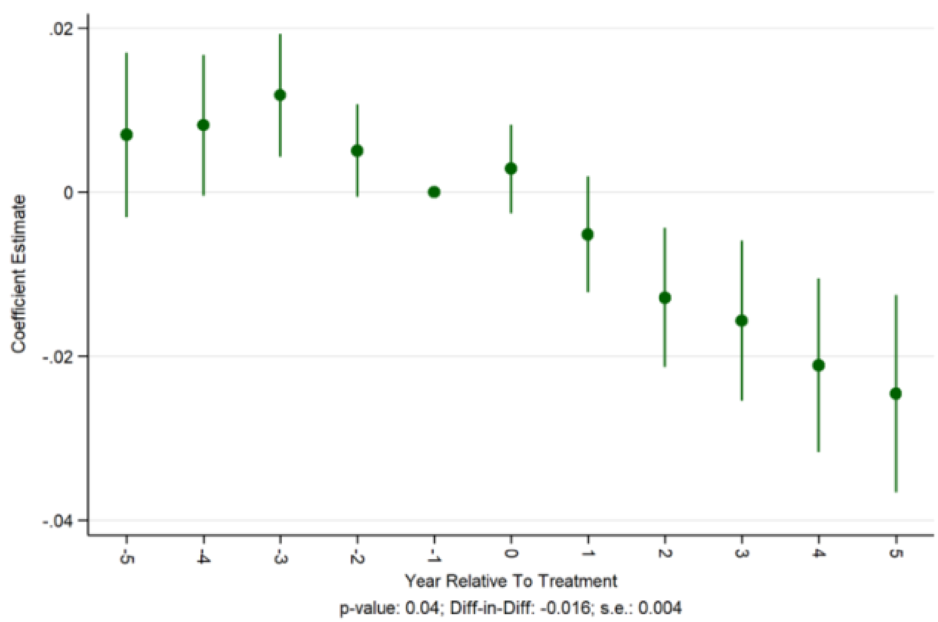
## 2.7 Conclusion

As the U.S. prepares to dramatically increase funding for public transit infrastructure projects under the American Jobs Plan, evidence regarding the potential consequences of alternative spending options will be of great value to helping to guide effective use of the funds. The ability for transit to improve access to employment, particularly among low-income and minority communities, is certain to play a role in weighing the costs and benefits of development. However, empirical evidence regarding this proposed benefit is surprisingly sparse considering the substantial cost and relative permanence of transit infrastructure.

The results of this paper shed light on the proposed labor market benefits of public




(a) Inner City




(b) Outer City

Figure 2.6: LODES: Log Employment Counts - Accommodation and Food Services

Log employment counts from LODES, accommodation and food services jobs only. Estimated coefficients and 95% confidence intervals $\beta_{\tilde{t}}$ in Equation 2.1. Standard errors clustered at the census block level.

commuter rail infrastructure investment. Using novel person-level data and exploiting variation in rail access arising from the large number of stations opened between 1968 and 2017, I am able to identify the worker-level employment consequences of past U.S. rail development. I find that those located in neighborhoods receiving a new rail station experience meaningful gains to employment relative to households that remained more distant. The impacts are comparable to other traditional place-based policies targeting disadvantaged communities.

The impacts depend heavily on household and neighborhood characteristics. Of particular importance, whether or not a household owns a car is a strong determinant of the strength of the effect. Perhaps unsurprisingly, those that do not own a vehicle see the largest benefits relative to car owners. The effects do not appear to vary much based on worker demographics such as race or education. However, workers residing in neighborhoods with a higher fraction of racial minorities and lower levels of education experience larger gains to employment relative to non-minority and more educated neighborhoods.

Households located near the city center are more likely to benefit from rail development relative to those located in the outer suburbs. Administrative data on employment counts suggests that the differential effect may be a result of an increase in employment opportunities among inner city neighborhoods receiving a station. This fact is consistent with notion of transit-oriented or transit-proximate development in which public transit can work to stimulate local economic activity. Evidence regarding whether rail access improves geographic mobility which can also improve opportunity is less clear. While researchers should continue to seek out additional data sources the may better describe the mechanisms driving the employment gains, it is important that future transit infrastructure development decisions incorporate the empirical facts identified in this paper.

# Chapter 3

# Real Estate Responses To Rail Development

## 3.1 Introduction

Public, urban rail development serves to enhance targeted communities' ability to navigate a given metropolitan area. Accordingly, the amenity value of improved transit access resulting from new development should be capitalized into local residential and commercial property values. However, past U.S. rail development has been accompanied by a wide range of intended and unintended consequences that may also influence local property values.[1] In addition, these consequences have been shown to vary extensively based on city and neighborhood characteristics, along with engineering and planning factors associated with the rail system.

In this paper, we use comprehensive data on U.S. rail development between 1996–2017 to investigate the impact of transit infrastructure on home values. Using the opening

[1]E.g., changes in local crime (Ihlanfeldt (2003)), demography Glaeser, Kahn and Rappaport (2008); Heilmann (2018); Kahn (2007); McQuilkin (2020), business activity and employment opportunities (Schuetz (2015); Tyndall (2019); Yen (2020)), and land use (Atkinson-Palombo (2010)).

of every rail station across all 32 metropolitan areas expanding rail systems during this period, we provide estimates the national level effects of receiving a rail station on local home values. We also heterogeneity of the effects both within and across cities. Estimates suggest a significant 3-7.5 percent increase in the value of single family homes located near new stations. The effects appear well before a station opens suggesting that expected future transit access can drive home prices upward.

We employ a staggered-adoption difference-in-differences approach comparing housing values in zip codes that fell into close proximity of a new rail station against those that remained farther. The annual data for the twenty-two year study period allows for detailed examination of dynamic effects using an event study empirical model.

An extensive literature exists documenting improved local transit infrastructure on properties values. However, the methods, data, and set of years and cities used to conduct empirical analysis have varied considerably making comparisons across studies difficult. Many papers have focused on rail development within a single city, often for a single rail line, to estimate the effects (Atkinson-Palombo (2010); Billings (2011); Gibbons and Machin (2005); Heilmann (2018); McMillen and McDonald (2004); Voith (1993)). Other studies have incorporated rail development in a set cities; Baum-Snow and Kahn (2000) looks at rail openings in five major cities that expanded transit in the 1980s while Kahn (2007) looks at fourteen cities that saw rail development between 1970–2000.

Most of these studies use either associative cross-sectional hedonic pricing models, or temporal first differences and difference-in-differences based approaches comparing values at some point in time before versus after a rail line has opened. Yet, Billings (2011) argues that proper identification of the effects requires knowing not only the price gradient in distance to transit infrastructure, but also information on property value trends of control and treatment homes. Additionally, McMillen and McDonald (2004) highlights the importance of anticipatory effects as home values may adjust upon expected future transit access. Even with the more sophisticated difference-in-difference approach, comparisons of home values at two (or even several) points in time may not be appropriate.

While prior studies have generally estimated a positive effect of rail development on home values, the impacts appear to vary considerably both across and within cities. But it is difficult to determine whether the discrepancy is driven by differences in methodology and data versus actual differences in the effects. Only within Baum-Snow and Kahn (2000) and Kahn (2007) are consistent empirical methods used to examine cross-city heterogeneity. Along with Atkinson-Palombo (2010), these three studies document within-city heterogeneity. This paper seeks to build on the literature by constructing consistent estimates with improved data and empirical methods.

We provide two primary contributions to the literature concerning home values and transit development. First, the data is extensive relative to prior studies including analysis on a substantial number of stations opened across nearly all U.S. cities that currently contain a rail system. As prior studies often focus on a single city and even a single rail line expansion, our estimates my be less influence by city level idiosyncrasies providing a better measure of the overall national level effects. In addition, we can examine how thew effects vary both within and across metropolitan areas with better statistical precision and consistent methodology.

Second, most previous studies rely on few observations across time, limiting the ability to examine the dynamic effects of rail development. As the impacts may emerge immediately following the announcement of a new rail line, continuing to change during construction and after opening, these dynamic effects are of great interest. The relative high frequency of the home value data spanning a period of over twenty years provides an excellent opportunity to explore these dynamic effects.

The remainder of this paper is organized as follows. In Section 3.2, I describe the rail infrastructure and housing index data. In Section 3.3, I detail the empirical methods. Section 3.4 presents the zip code level estimated effects. Section 3.5 concludes.

## 3.2 Data

Housing data comes from the Zillow Home Value Index (ZHVI) which provides estimates for the value of a typical home in the 35$^{\text{th}}$ to 65$^{\text{th}}$ percentile by zip code.[2] The data is monthly, spanning the period of January 31, 1996 through February 28, 2021. We keep each July 31 observation and drop years 2018–2021 leaving an annual panel for each zip code between 1996–2017.[3] Estimates are inflation-adjusted to 2017 dollars.

Detailed information on the exact location and date of opening for all urban rail stations opened throughout the United States between 1996–2017 is scraped from Wikipedia. All subway, light rail, and commuter rail systems are included. Table 3.1 presents the number of stations opened, along with the first year and last year of an opening, for each 2010 Census designated Urban Area throughout the study period. In total, 912 stations opened across 31 cities during the twenty-one year period. Of the 31 cities, 19 of them opened 20 or more new stations. Portland and Los Angeles stand out with the most significant expansions (72 and 67 new stations, respectively). Roughly half of the cities opened a station in the first three years (1997–1999) while more than half opened a station in the final three years (2015–2017). This results in variation within many of the sample cities for a large portion of the sample duration.

Figure 3.1 displays the cumulative number of stations for each city across time. Panel (a) shows the urban areas with the largest number of stations by 2017 while Panel (d) displays those for the smallest. There is considerable variation in the number of stations

[2] ZHVI is available at the neighborhood level. We currently use zip code values due to the availability of zip code spacial data. Future estimates will employ Zillow ZTrax data which provides home values at the property level.

[3] Future estimates will employ monthly estimates and extend the study period to include years 2018–2021.

Table 3.1: Rail Systems By Urban Area

| | Stations | First Year | Last Year |
|---|---|---|---|
| Albuquerque | 15 | 2006 | 2017 |
| Atlanta | 2 | 2000 | 2000 |
| Austin | 9 | 2010 | 2010 |
| Baltimore | 8 | 1997 | 1998 |
| Boston | 45 | 1997 | 2017 |
| Charlotte | 20 | 2007 | 2015 |
| Chicago | 23 | 1997 | 2017 |
| Cleveland | 3 | 1997 | 2015 |
| DC | 20 | 1997 | 2015 |
| Dallas | 63 | 1997 | 2016 |
| Denver | 49 | 2000 | 2017 |
| Houston | 39 | 2004 | 2017 |
| Los Angeles | 67 | 1997 | 2017 |
| Miami | 5 | 1998 | 2012 |
| Minneapolis | 44 | 2004 | 2014 |
| Nashville | 6 | 2006 | 2006 |
| New Haven | 4 | 2002 | 2013 |
| New York | 57 | 2000 | 2017 |
| Norfolk | 11 | 2011 | 2011 |
| Orlando | 12 | 2014 | 2014 |
| Philadelphia | 27 | 1997 | 2015 |
| Phoenix | 40 | 2008 | 2016 |
| Pittsburgh | 12 | 2001 | 2012 |
| Portland | 72 | 1997 | 2015 |
| Sacramento | 26 | 1998 | 2015 |
| Salt Lake City | 66 | 1999 | 2013 |
| San Diego | 27 | 1997 | 2008 |
| San Francisco | 51 | 1997 | 2017 |
| San Jose | 36 | 1997 | 2012 |
| Seattle | 34 | 2000 | 2016 |
| St. Louis | 19 | 1998 | 2006 |
| Total | 912 | 1997 | 2017 |

Column (1) describes the number of stations opened between 1996 and 2017 for each metropolitan area. Columns (2) and (3) display the year of the earliest and most recent expansion, respectively, during this time period for each metropolitan area. The last row presents the information for the total of all metropolitan areas.

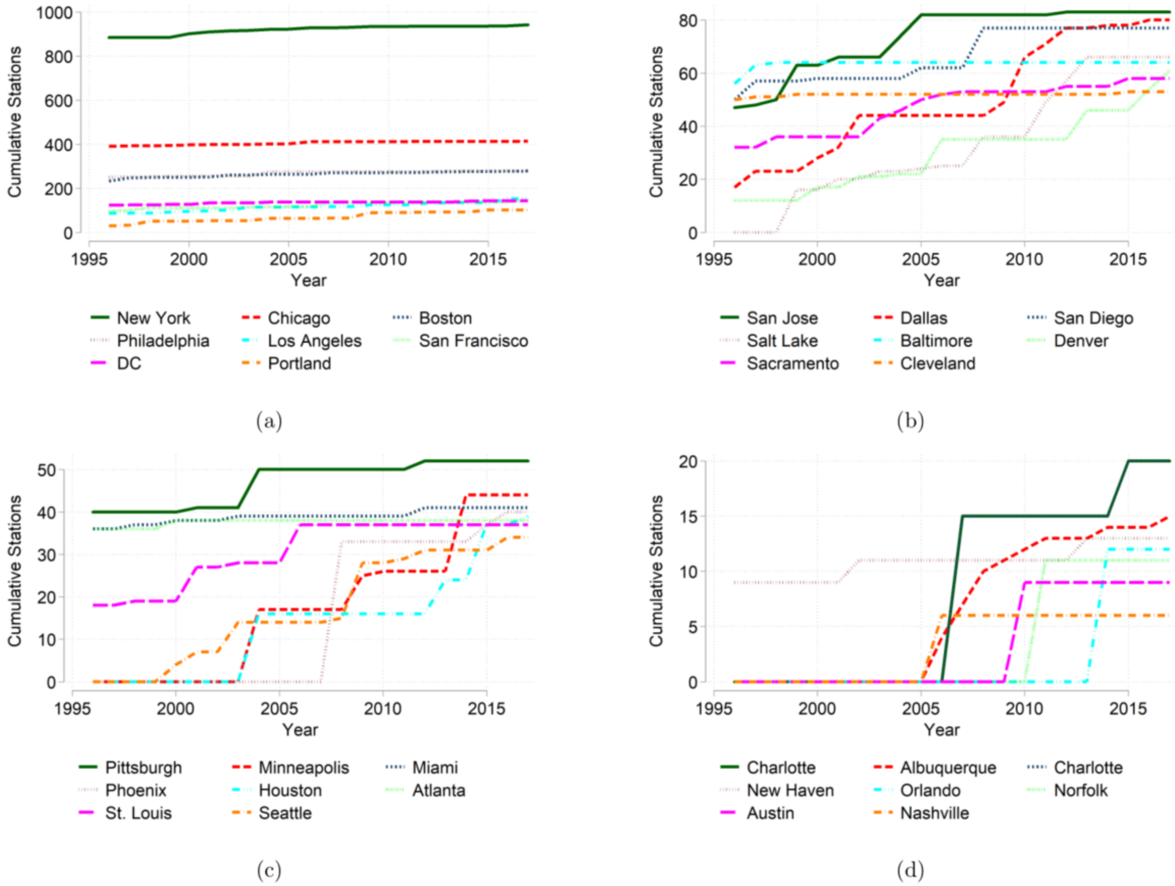



Figure 3.1: Number of Treated Tracts by Urban Area

Cumulative number of open rail stations by 2010 Census designated Urban Area.

across most of the cities in the sample. The increases represent the opening of new rail lines which bring rail access to initially distant neighborhoods.

Zip code geographic information is collected from 2010 zip code boundaries. For each zip code, I calculate the spatial centroid and measure the distance to the nearest rail station for each year to construct a measure of past, present, and future rail access in order to define treatment. I keep all zip codes that eventually fall within 4 miles of a station. Treatment is defined by whether a zip code lies within 1 mile of a station.

Figure 3.2 displays the values of houses across the study period split by eventual

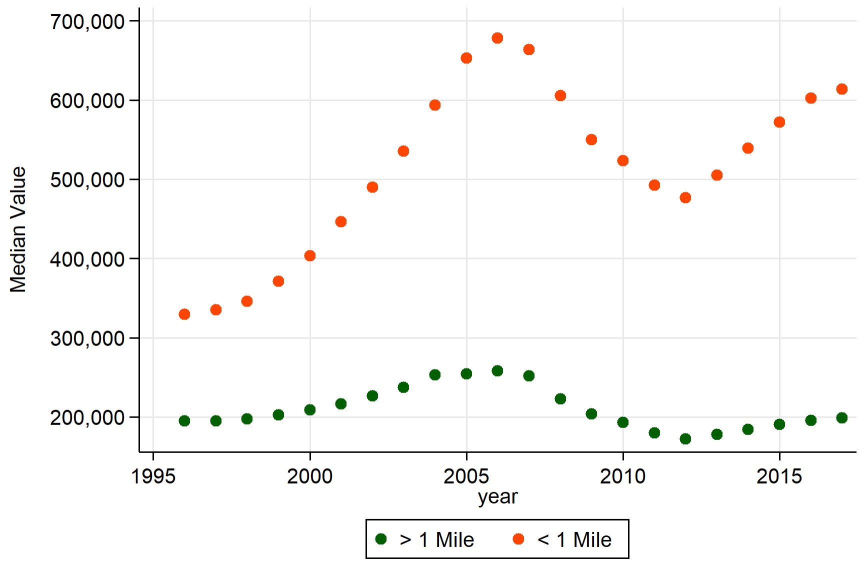



Figure 3.2: Housing Values By Year

Housing value estimates by year, by treatment status. Sample includes all zip codes that were within 4 miles of a station in 2017.

treatment status. Homes located in zip codes with centroids closer than 1 mile to a rail station are, on average, valued much higher relative to those that never receive treatment. They are also far more volatile experiencing a larger effect of the build-up and burst of the 2007 housing bubble. In Figure 3.3, I plot the 2017 values based on the distance to the nearest rail station. There is a near-monotonic negative relationship between home values and distance to the nearest station. Interestingly, a large approximate $100,000 jump exists between those located within two versus one mile to a station. This motivates using a 1 mile threshold to define treatment status.

Table 3.2 presents the average home values across the study period for each city in the sample. In Column (1), we include values for all zip codes that eventually fell within 4 miles of a station. Columns (2) and (3) break the sample based on eventual treatment status. At the bottom of the table, averages for the full sample are included.

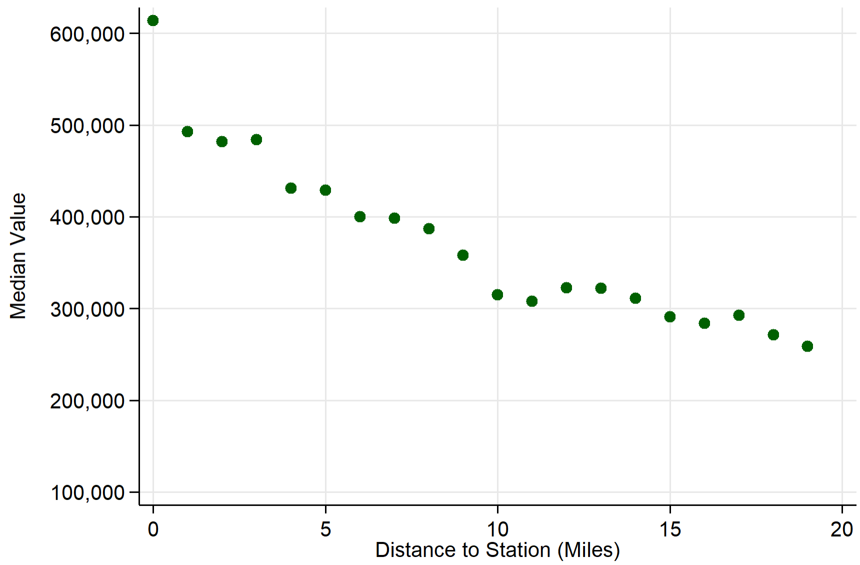



Figure 3.3: Housing Values By Distance To Station (2017)

Housing value estimates by distance to nearest station.

On average, home prices are higher in eventually treated zip codes. However, for 13 cities, home values are higher in outside of the 1 mile treatment threshold. In Section 3.4, we examine treatment effect heterogeneity across cities.

## 3.3 Methods

I estimate the impacts of rail development on local housing values using a staggered-adoption differences-in-differences empirical specification. Home values in zip codes that became closer than 1 mile to a station are compared to those that remain farther than 1 mile, before and after the station opens. To improve comparability between treated and untreated zip codes, I restrict the sample to zip codes that were eventually located within 4 miles of a station by 2017.

Identification of the causal effect relies on a parallel trends assumption where home

Table 3.2: Home Values By City - Eventual Treatment Status

| | Total | 1-4 miles | $\leq$ 1 mile |
|---|---|---|---|
| Albuquerque | 218,649 | 228,004 | 176,320 |
| Atlanta | 343,126 | 346,051 | 337,656 |
| Austin | 360,125 | 357,621 | 370,091 |
| Baltimore | 264,015 | 275,567 | 235,082 |
| Boston | 510,158 | 418,100 | 651,881 |
| Buffalo | 123,999 | 93,347 | 189,390 |
| Charlotte | 316,296 | 303,830 | 362,004 |
| Chicago | 330,606 | 296,574 | 365,680 |
| Cleveland | 140,329 | 146,430 | 131,396 |
| DC | 484,892 | 463,208 | 532,130 |
| Dallas | 223,924 | 223,600 | 225,284 |
| Denver | 339,213 | 338,641 | 340,810 |
| Houston | 294,130 | 298,259 | 284,112 |
| Los Angeles | 527,004 | 536,065 | 498,236 |
| Miami | 404,125 | 421,907 | 333,163 |
| Minneapolis | 235,900 | 233,149 | 245,785 |
| Nashville | 246,460 | 256,886 | 215,877 |
| New Haven | 444,953 | 381,103 | 562,701 |
| New York | 598,059 | 533,642 | 663,167 |
| Norfolk | 215,049 | 180,826 | 290,027 |
| Orlando | 227,825 | 209,669 | 278,827 |
| Philadelphia | 268,704 | 273,845 | 263,225 |
| Phoenix | 209,137 | 195,014 | 243,032 |
| Pittsburgh | 168,343 | 167,135 | 171,118 |
| Portland | 367,119 | 365,212 | 369,374 |
| Sacramento | 309,451 | 315,538 | 293,977 |
| Salt Lake City | 276,829 | 285,010 | 253,029 |
| San Diego | 563,502 | 627,940 | 427,044 |
| San Francisco | 804,251 | 776,058 | 848,764 |
| San Jose | 909,312 | 985,143 | 773,781 |
| Seattle | 395,830 | 401,167 | 366,476 |
| St. Louis | 210,141 | 192,678 | 244,746 |
| Total | 455,773 | 426,633 | 504,694 |
| Observations | 64,490 | 40,416 | 24,074 |

Column (1) describes the mean home value index for the full sample by city. Column (2) restricts to zip codes that remained untreated by 2017. Column (3) restricts to zip codes that were treated by 2017.

values in zip codes that remained farther than 1 mile of a station can serve as a suitable counterfactual for those that became closer than 1 mile, had the station never been built. I evaluate this assumption using an event study empirical framework, comparing treated and untreated zip codes prior to the opening of the station. Equation 3.1 characterizes the empirical model.

$$\begin{aligned} y_{ic\tilde{t}} &= \sum_{s \in \mathcal{T}} \beta_s \mathbb{1}(\tilde{t} = s) \cdot Treatment_i + \delta_i + \gamma_{ct} + \epsilon_{ict} \\ \mathcal{T} &= \{-15, -14, ..., 13, 14\} \end{aligned} \tag{3.1}$$

where $y_{ict}$ equals the (log) home value for zip code $i$ in county $c$ at relative year $\tilde{t}$. $\tilde{t} = -1$ represents the last year prior to treatment, while $\tilde{t} = 0$ is the first decennial year in which zip code $i$ is treated. The excluded relative year category is $\tilde{t} = -1$. $\mathbb{1}(\tilde{t} = s)$ is an indicator variable that equals 1 if $\tilde{t}$ equals $s$, and zero otherwise. $Treatment_i$ is an indicator for whether zip $i$ ever falls within 1 mile of a rail station. $\delta_i$ are zip code fixed effects $\gamma_{ct}$ are county-year fixed effects. Errors are clustered at the zip code-level.

Prior literature indicates that anticipatory effects are likely to impact home values prior to the actual opening of a station (McMillen and McDonald (2004)). As neighborhoods experience changes to their expectations regarding future local transit development, home values will potentially respond well before the station opens. The event study estimates can be helpful in determining the extent to such anticipation. In addition, the post-treatment coefficients help describe any dynamic effects of receiving rail access.

To summarize the pre- and post-treatment differences in home values between treated and untreated zip codes, I estimate a difference-in-differences model characterized by

Equation 1.1.

$$y_{ict} = \beta D_{ict} + \delta_i + \gamma_{ct} + \epsilon_{ict} \tag{3.2}$$

where $D_{ict}$ represents the treatment indicator for zip codes that fall withing 1 mile of station.

We select fifteen years prior to treatment and 14 years following treatment as our window of observation. As changes in expected future transit access could extend many years prior to a station opening, this large study period helps describe these potential anticipatory effects. However, in relative years far from zero, there are fewer treated zip codes contributing to identification. To examine the robustness of the estimates, we make restrictions on the number of pre- and post-treatment years and estimates Equations 3.1 and 3.2.

## 3.4 Zip Code Level Estimates

Figure 3.4 presents the estimates of Equation 3.1 using the full sample of zip codes with centroids eventually falling within 4 miles of a rail station. We include point estimate for $\beta_s$ for $s = -15, -14, ..., 14$ along with 95 percent confidence interval error bands. The excluded relative year category is one year prior to a station's opening.

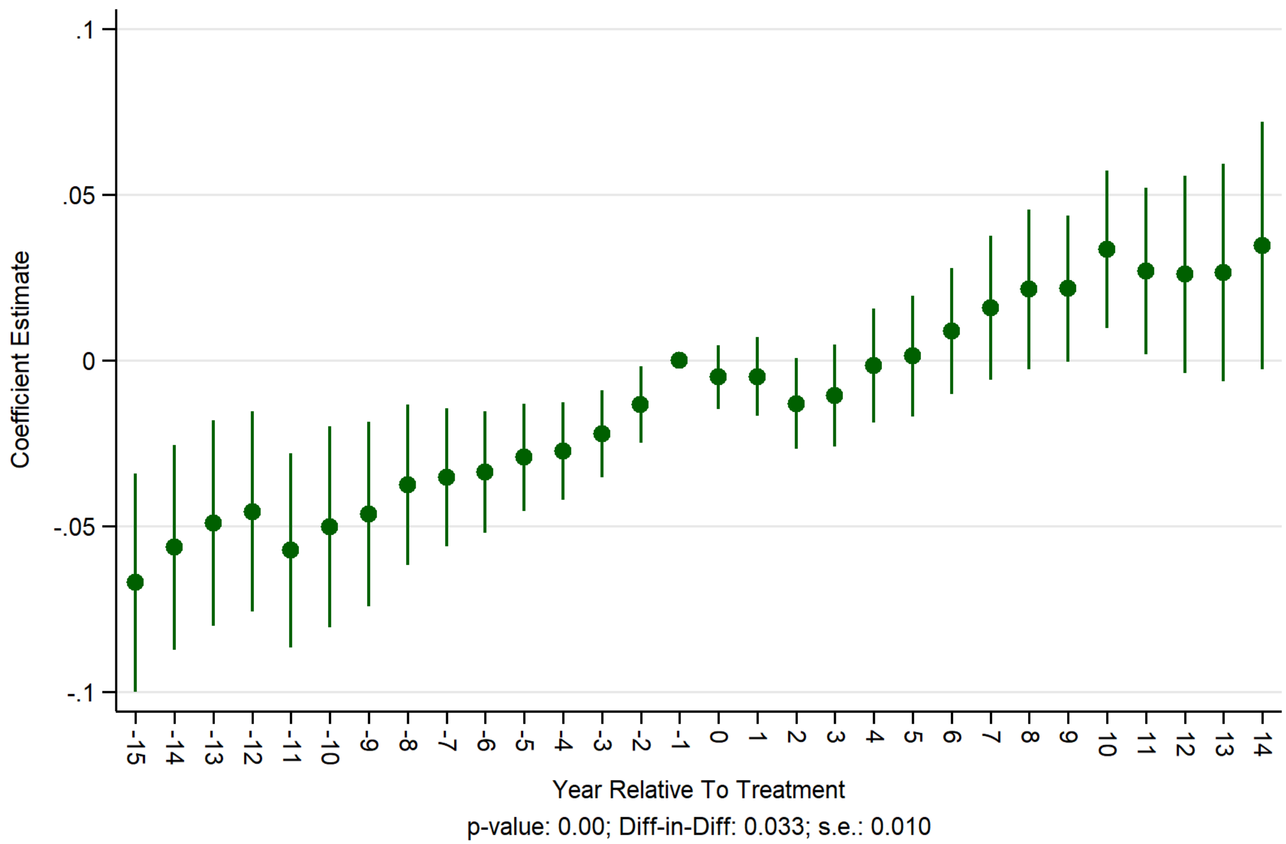



Figure 3.4: Event Study Results - Log Value

Event study estimates of impact of rail station on housing values.

Fifteen to ten years prior to opening, there is no noticeable trend in the estimates and all coefficients are statistically indistinguishable. In relative years 10–1, there is a monotonic increasing trend in the estimates. In total, home values appreciate by approximately 5 percent during this 9 year period. The estimates suggest that home values in treated zip codes are increasing relative to those in the control group well before a station opens. This is consistent with the notion that expected future rail infrastructure can influence home values up to roughly 10 years prior to opening. After the station opens, there is a small dip in home values for the first two years, until an upward trend reappears from relative years 2–10. For relative years 10 onward, the estimates stabilize at the new elevated level approximately 3-4 percent higher than at the time of treatment.

In Table 3.3, we present estimates of Equation 3.2. Column (1) includes the full sample. The estimate suggests that homes located in zip codes receiving a station within one mile experience a significant 3.31 percent increase in value following a stations opening. This estimate is of similar magnitude to previous studies. In Columns (2) and (3), we split the sample based on the 1996 home index value relative to the city-specific median. Column (2) shows that among initially higher home value zip codes, the effect is slightly smaller (though not significantly different). Column (3) shows that homes in initially lower value zip codes experience a significant 4.72 percent increase following treatment.

As the pre-treatment coefficients in Figure 3.4 display a clear upward trend from ten years preceding the stations opening, it is likely that defining treatment in the year in which a station opens does not accurately capture the effect of rail development on home values. This raises concern that previous studies that rely on relatively infrequent data for only couple or few year years may not capture the true effect. However, as the time between announcement and a new rail line opening is roughly 10 years, the estimates in Figure 3.4 and Table 3.3 may still indicate a positive causal effect on home

Table 3.3: Diff-in-Diff - Log Home Values

| | (1) | (2) | (3) |
|---|---|---|---|
| | | $\geq$ Med. Value | $<$ Med. Value |
| VARIABLES | Log Value | Log Value | Log Value |
| $\leq$ 1 Mile | 0.0331*** | 0.0217* | 0.0472*** |
| | (0.00968) | (0.0126) | (0.0146) |
| Observations | 44,960 | 21,617 | 21,604 |
| R-squared | 0.895 | 0.924 | 0.911 |

Estimates of Equation 3.2. Column (1) presents the estimate using the full sample of zip codes that were located within 4 miles from a rail station by 2017. Column (2) restricts to only zip codes that had a home value index greater than the city specific median in 1996. Column (3) restricts to zip codes that had an index less than the city specific median in 1996.

values. Further, if home values begin to adjust 10 years prior to a stations opening, then difference-in-differences estimates based on Equation 3.2 would understate the full effect. In order to better address anticipatory effects, future research will require information on the announcement of rail development.

In Figure 3.5, I display the difference-in-difference estimates restricting the sample to each metropolitan area. The estimates are organized in decreasing order by city level treatment effect. We include each point estimate along with the 95 percent confidence interval. The point estimates are predominantly positive with 6 of 28 cities experiencing significant increases. Charlotte, Phoenix, St. Louis, and Los Angeles see the largest increase of between 7-21 percent. Treated zip codes in Pittsburgh and Nashville see significant declines, although these estimates are identified with a small number of station openings. The figure highlights the intercity heterogeneity of home value responses to rail development and suggests that estimated effects in prior literature may be highly specific limiting the external validity.

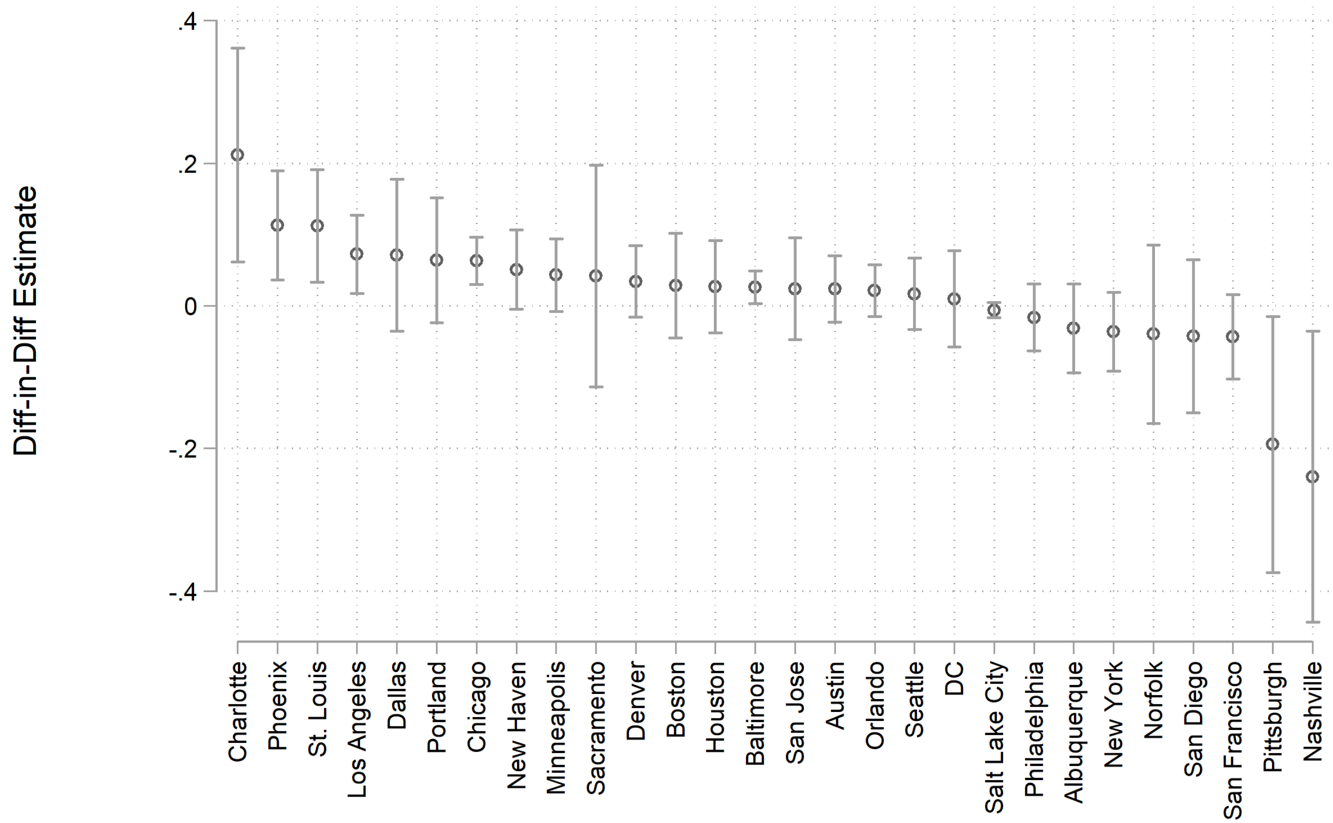



Figure 3.5: Treatment Effect By Metro

Difference-in-Differences estimates by metropolitan area. Point estimates and 95% confidence intervales. Standard errors clustered at the zip code level.

## 3.5 Conclusion

Estimating how property values respond to transit development represents a critical component to understanding the overall impacts of transit on targeted communities. If changes in home values, which feed into rental prices, alter the ability of incumbent or migrating residents to afford living in certain areas, then the efficacy relating to the proposed objectives of transit investment may be reduced. Prior literature has pieced together data and empirical methodology to produce estimates of the impacts of certain transit expansions. However, weaknesses relating to the potential for dynamic effects along with considerable heterogeneity between and within cities make these estimates less useful when weighing future transit development.

This paper uses comprehensive data that includes annual home value information across nearly all U.S. rail transit cities throughout a 21 year period to estimate the impacts of rail transit development. We find meaningful, positive estimated effects that are roughly consistent with those in previous studies, but highlight concerns that these estimates may not be accurately capturing the true effect. We also demonstrate an great amount of variation in the effect by city suggesting that prior estimates may be difficult to apply to cost benefit calculation relating to newly proposed transit development. Future research should focus on using improved methodology along with more complete data to provide better estimates of how new development will impact home values in affected areas.

# Appendix A

# Appendix: Chapter 1

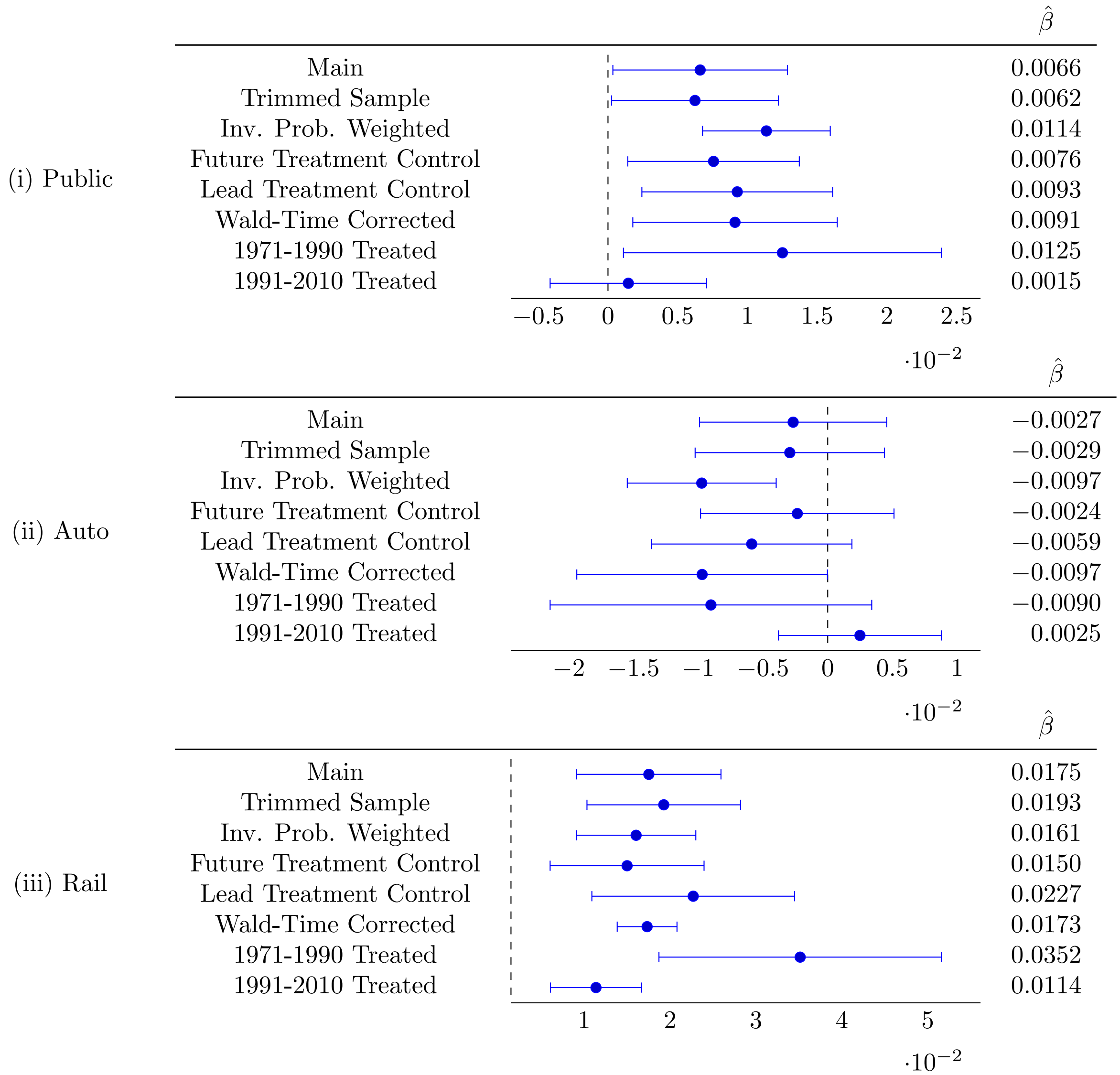



Figure A.1: Robustness Specifications - Full Sample (continued on next page)

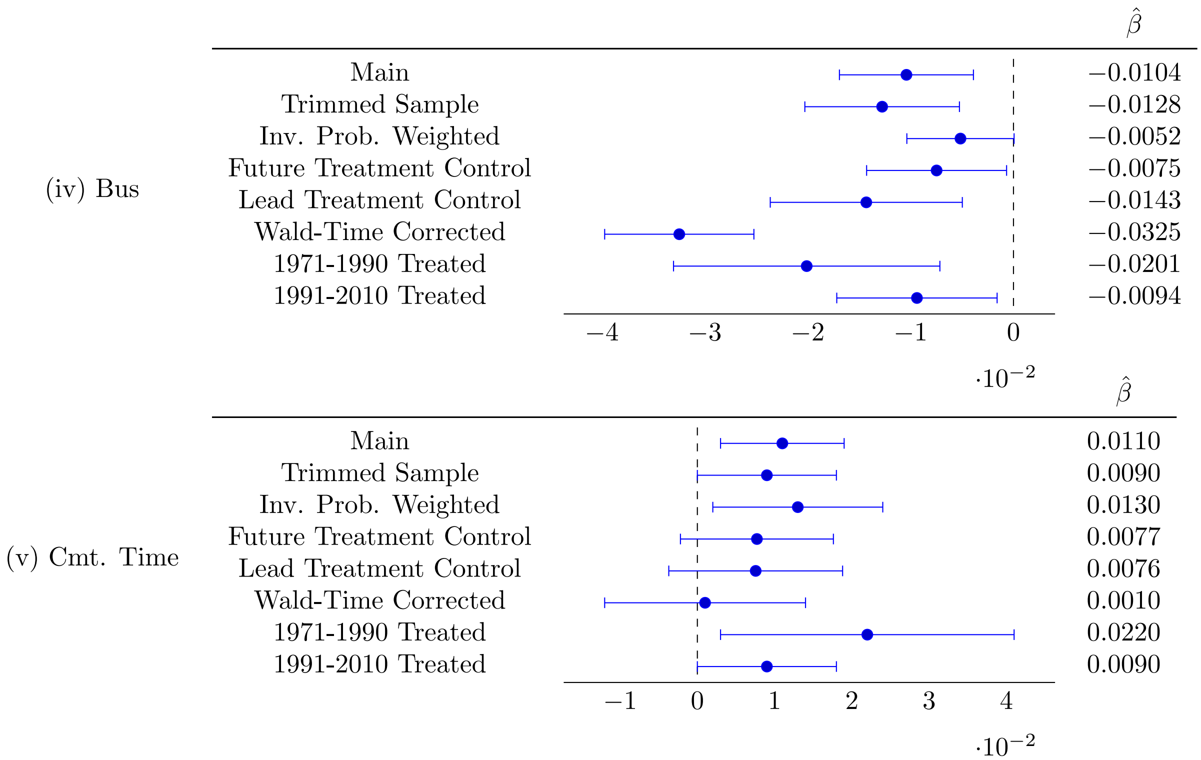



Figure A.1: Robustness Specifications - Full Sample

Estimated coefficient and 95% confidence interval for Equation 1.1 and each robustness specification described in Section 1.5.

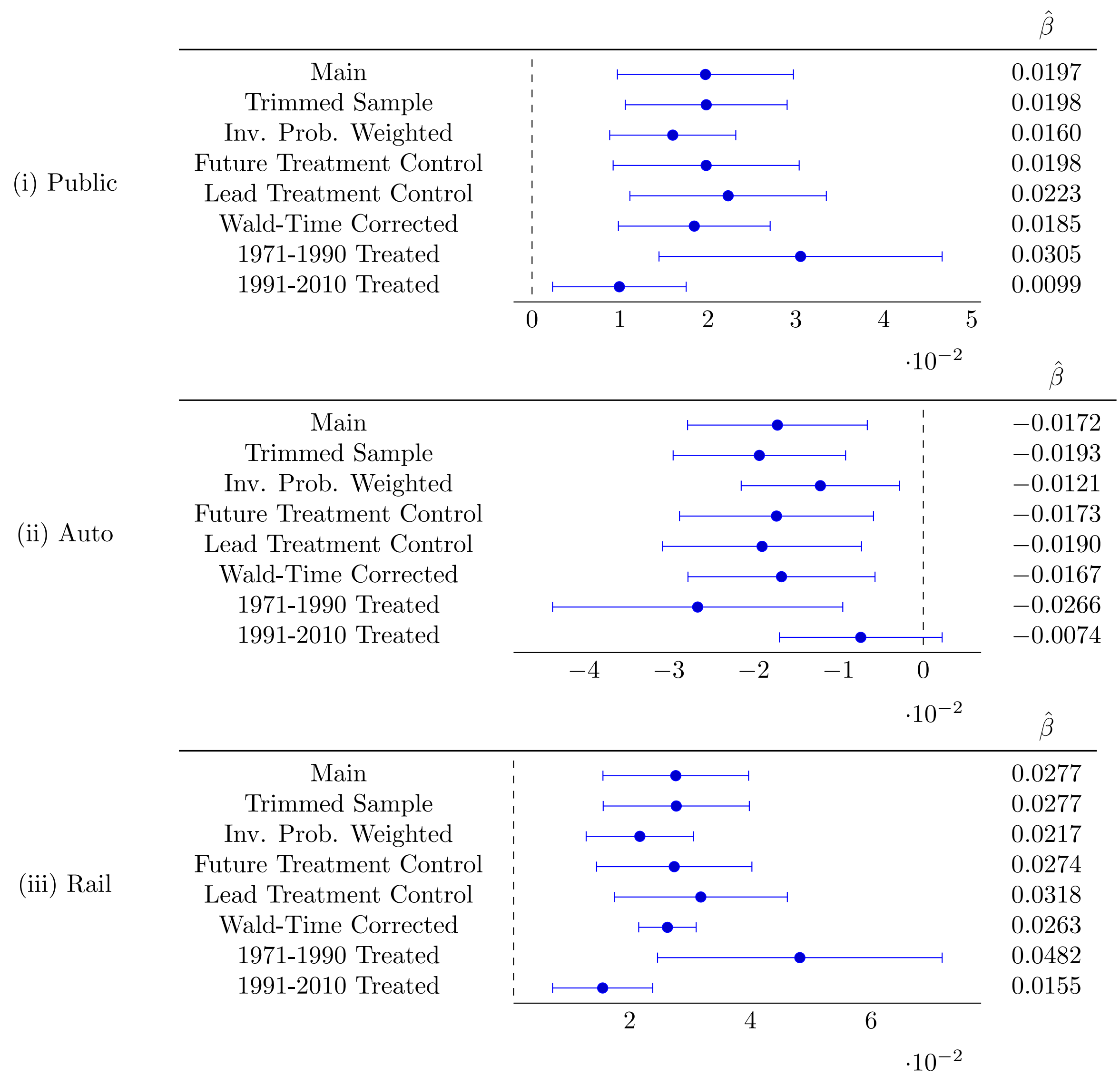



Figure A.2: Robustness Specifications - $\leq$ Median Poverty (continued on next page)

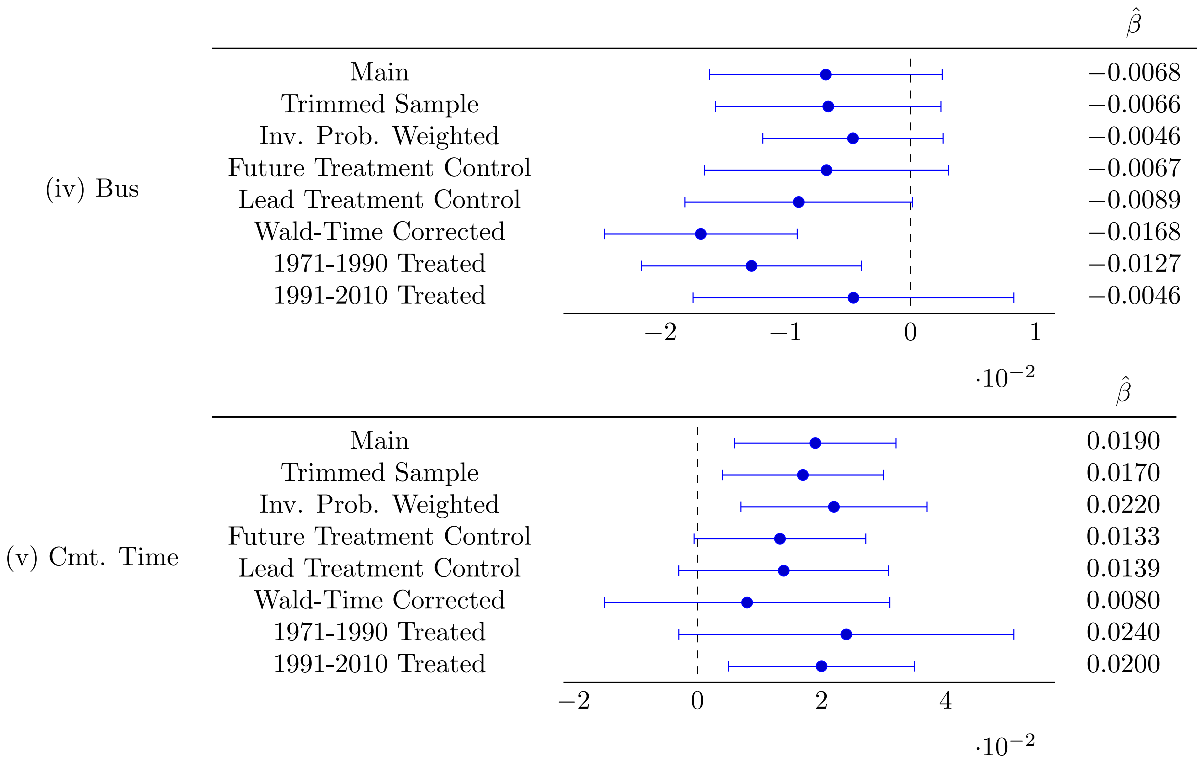



Figure A.2: Robustness Specifications - $\leq$ Median Poverty

Estimated coefficient and 95% confidence interval for Equation 1.1 and each robustness specification described in Section 1.5.

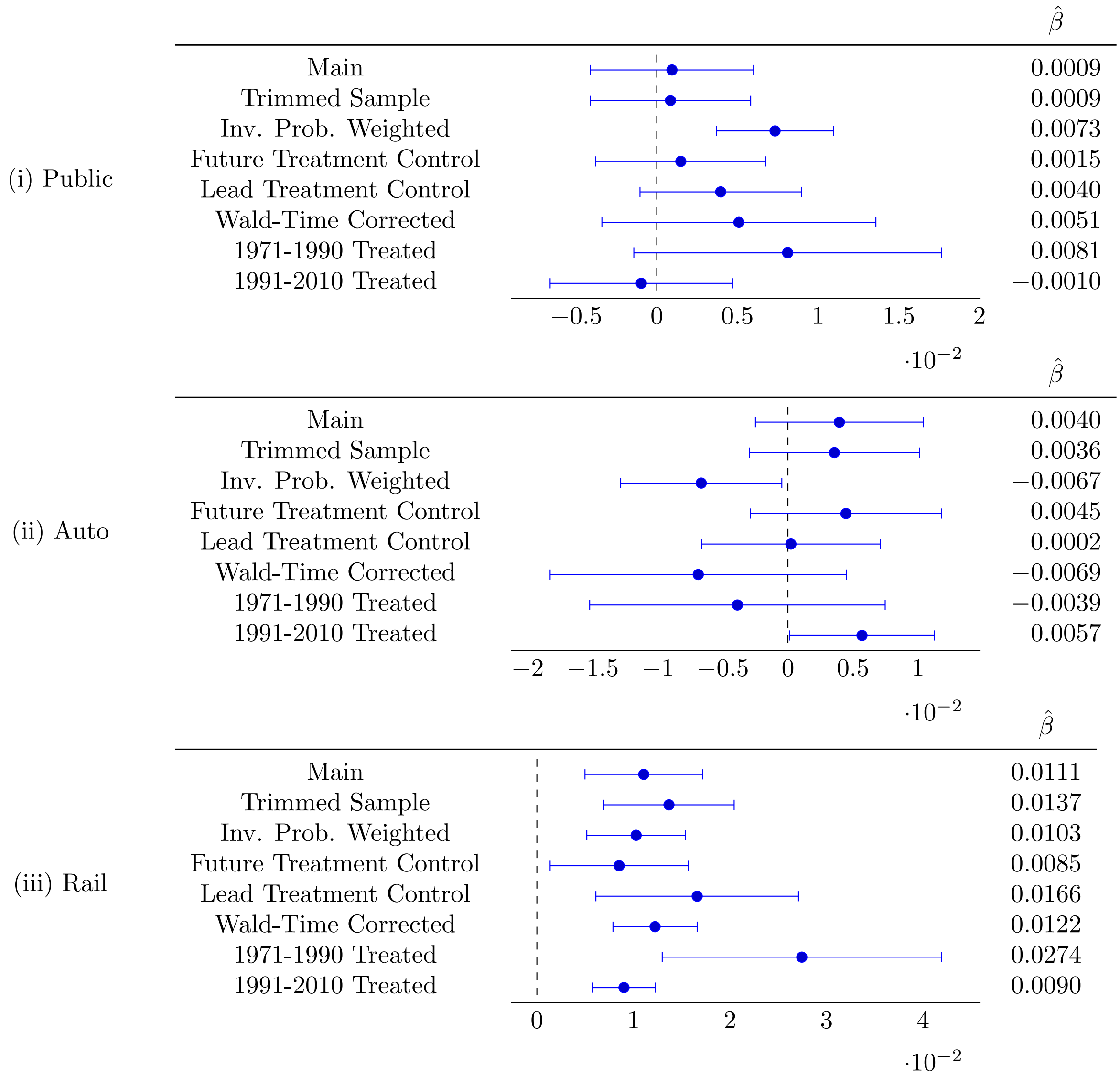



Figure A.3: Robustness Specifications - > Median Poverty (continued on next page)

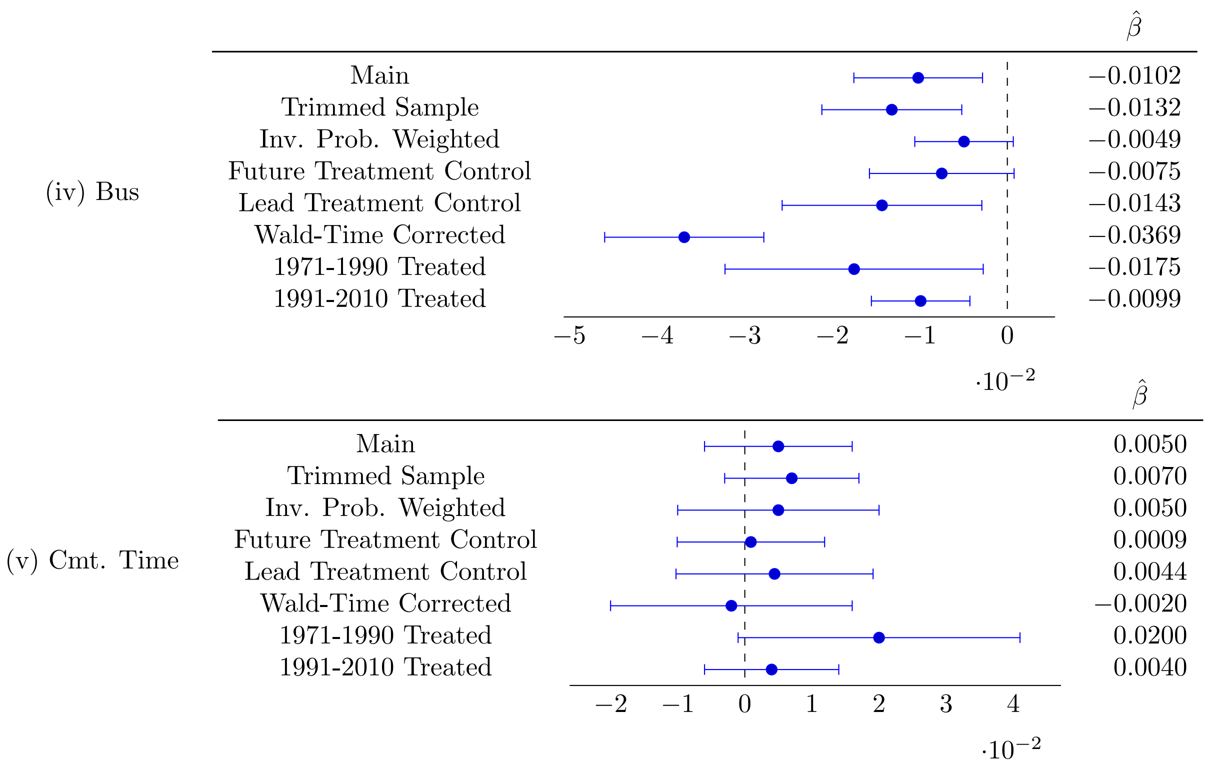



Figure A.3: Robustness Specifications - > Median Poverty

Estimated coefficient and 95% confidence interval for Equation 1.1 and each robustness specification described in Section 1.5.

Table A.1: Urban Area Aggregate - Proportions

| | (1)<br>Prop. Public | (2)<br>Prop. Auto | (3)<br>Prop. Rail | (4)<br>Prop. Bus | (5)<br>Commute Time |
|---|---|---|---|---|---|
| Panel A: # Stations | | | | | |
| # Stations | -0.000274** | -5.66e-07 | 0.000134 | -0.000409** | -0.00144*** |
| | (0.000128) | (0.000143) | (0.000151) | (0.000167) | (0.000479) |
| Panel B: Log # Stations | | | | | |
| 1(0 Stations) | -0.0186*** | 0.00901 | 0.00251 | -0.0224*** | -0.0425 |
| | (0.00482) | (0.00634) | (0.00492) | (0.00572) | (0.0410) |
| Log # Stations | -0.00673*** | 0.00103 | 0.00146 | -0.00799** | -0.0361** |
| | (0.00234) | (0.00301) | (0.00200) | (0.00306) | (0.0153) |
| Panel C: Stations Per Capita (Per 100,000 Residents) | | | | | |
| Stations Per Capita | 0.000194 | -0.000281 | 6.35e-06 | 3.20e-05 | 0.000310 |
| | (0.000152) | (0.000190) | (9.55e-05) | (3.45e-05) | (0.000551) |
| Observations | 381 | 381 | 304 | 304 | 312 |
| Mean Outcome | 0.05 | 0.89 | 0.01 | 0.03 | 3.12 |

U.S. Census defined 2010 Urban Area aggregate estimates of Equation 1.4. Outcomes are the proportion of commuters by mode. Clustered standard errors (at Urban Area level) in parentheses. Sample includes Urban Areas that contained at least 1 rail station by 2010.

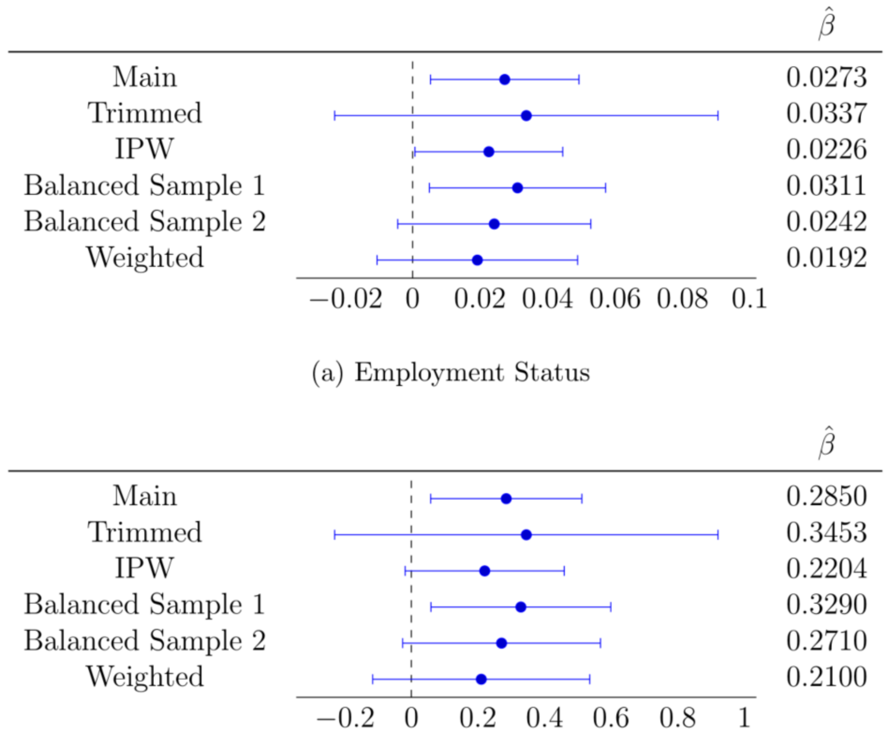
Table A.2: Variable Definitions

| Variable | Definition |
|---|---|
| Population | Population |
| Pop. Density | Persons per square mile |
| Dist. From CBD | Distance in miles from metropolitan central business district |
| Prop. Public Transit | Proportion of public transit commuters |
| Prop. Rail | Proportion of commuter rail commuters |
| Prop. Bus | Proportion of bus commuters |
| Prop. Private Auto | Proportion of private automobile commuters |
| Prop. Other Commute | Proportion of other means commuters (primarily walking, cycling, and taxi) |
| Med. Commute Time | Median commute time |
| Prop. Male | Proportion of male individuals |
| Prop. Black | Proportion of black individuals |
| Prop. Native | Proportion of native individuals |
| Prop. Same House | Proportion living in same house five years prior |
| Prop. Less HS | Proportion of adults with less than a high school diploma |
| Prop. HS | Proportion of adults with a high school diploma as highest attainment |
| Prop. Some College | Proportion of adults with some college/2 year degree as highest attainment |
| Prop. Bach./Grad. | Proportion of adults with bachelor degree as highest attainment |
| Unemp. Rate | Proportion of individuals in labor market currently unemployed |
| Prop. Poverty | Proportion of individuals with family income below federal poverty line |
| Family Income | In 2010 dollars |
| Prop. 0-17 | Proportion 0-17 years old |
| Prop. 18-24 | Proportion 18-24 years old |
| Prop. 25-29 | Proportion 25-29 years old |
| Prop. 30-34 | Proportion 30-34 years old |
| Prop. 35-44 | Proportion 35-44 years old |
| Prop. 45-64 | Proportion 45-65 years old |
| Prop. 65+ | Proportion 65 years or older |
| Num. Households | Number of households |
| Prop. Non-Family HHs | Proportion of households not a family |
| Housing Units | Number of housing units |
| Median Rent | In 2010 dollars |
| Prop. Occupied | Proportion of occupied housing unites |
| Prop. Renter Occupied | Proportion of occupied housing units that are rented |

Commuters defined employed individuals that do not work from home. Commute type defined as the primary means by which employed individuals get from home to work. Adults defined as those 25 years and older for education categories.

# Appendix B

# Appendix: Chapter 2




(a) Employment Status

(b) Log Annual Labor Income

Figure B.1: Robustness Specifications

Estimated coefficient and 95% confidence interval for Equation 1.1 and each robustness specification described in Section 1.5.

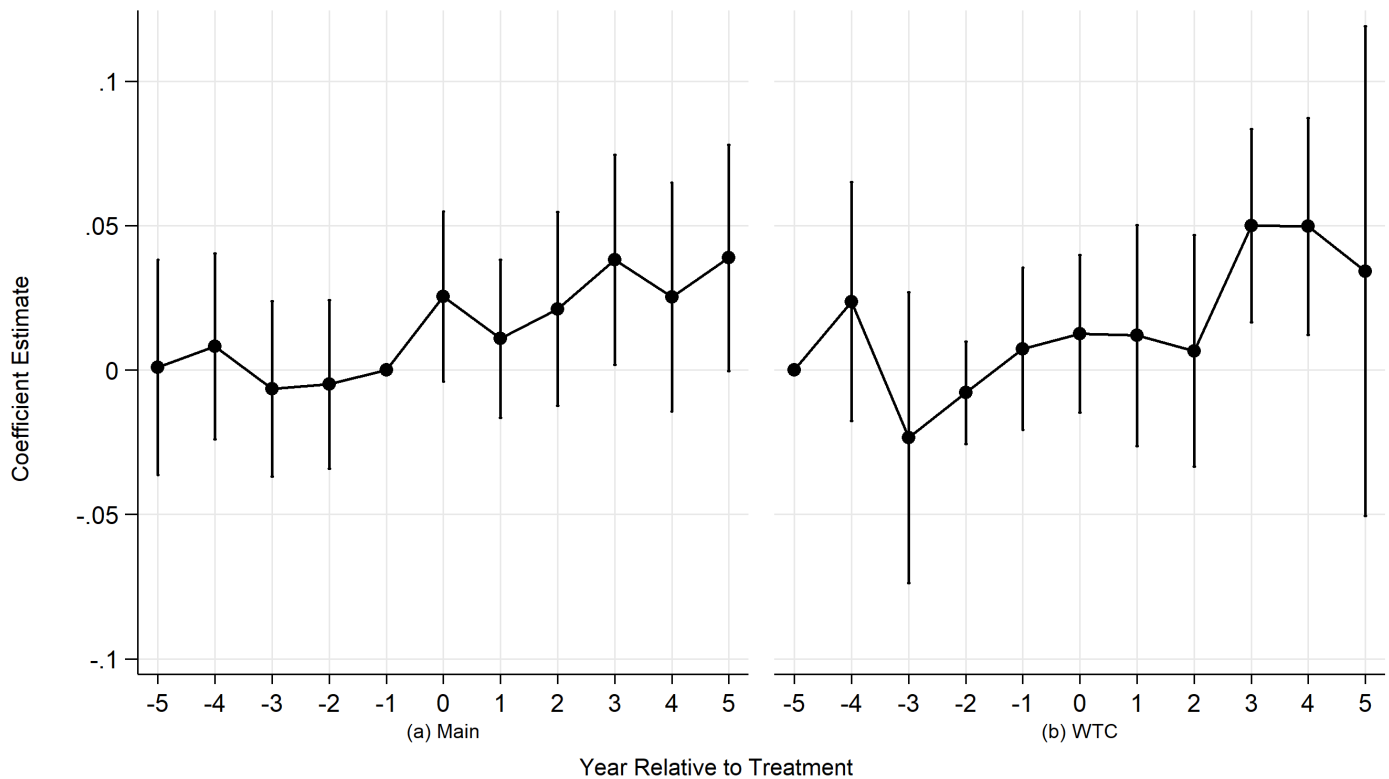



Figure B.2: Wald Time Corrected: Employment Status

Comparison of event study estimates from Section 2.4 with those using Wald Time-Corrected Estimates. The left plot displays the event study estimates from Section 2.4 while the right plot displays those using the Wald Time-Corrected estimator (de Chaisemartin and d'Haultfoeuille (2019)).

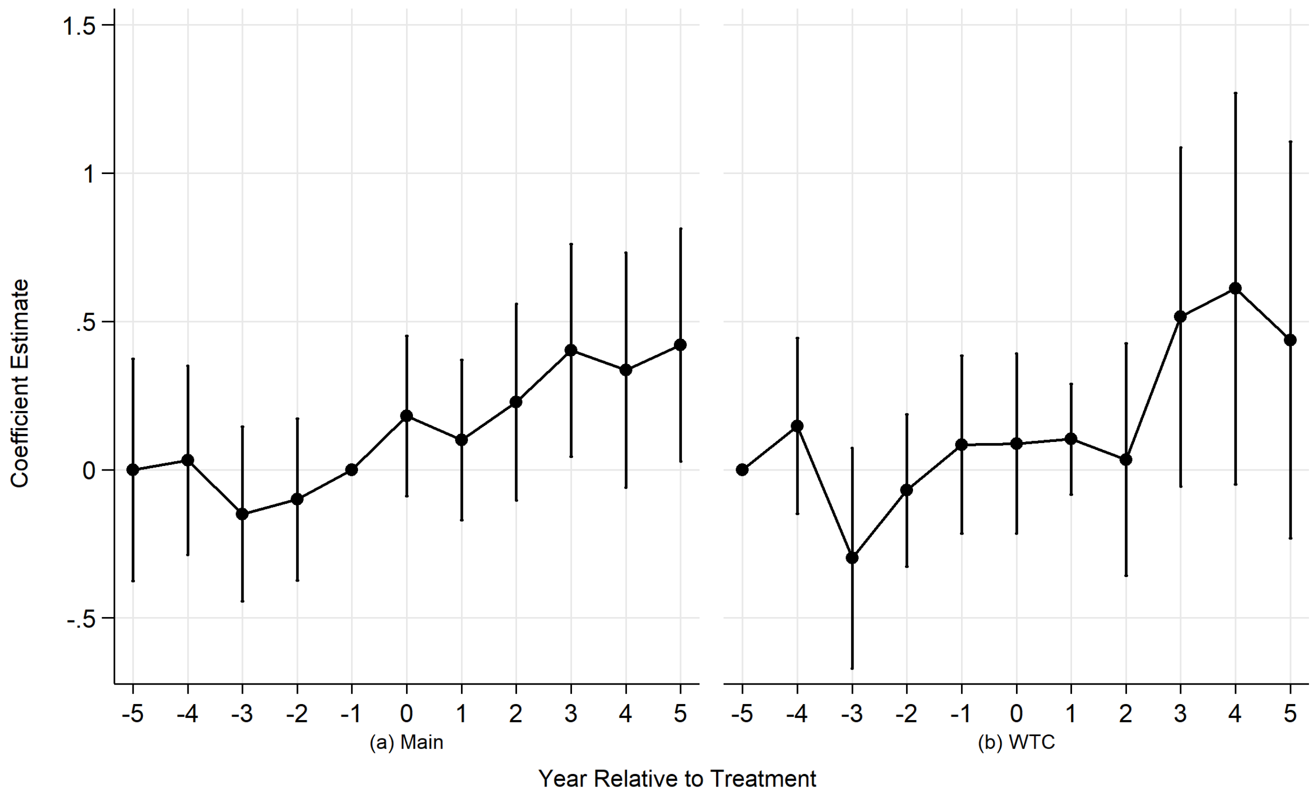



Figure B.3: Wald Time Corrected: Log Annual Labor Income

Comparison of event study estimates from Section 2.4 with those using Wald Time-Corrected Estimates. The left plot displays the event study estimates from Section 2.4 while the right plot displays those using the Wald Time-Corrected estimator (de Chaisemartin and d'Haultfoeuille (2019)).

Table B.1: Variable Definitions

| Variable | Definition |
|---|---|
| Employed | Household head: indicator for having annual labor income greater than 0 |
| Annual Labor Income | Household head: in 2017 dollars |
| Hourly Wage | Household head: in 2017 dollars |
| Annual Hours Worked | Household head |
| Age | Household head: age at time of interview |
| Male | Household head: indicator for being male |
| Black | Household head: indicator for being black |
| Married | Household head: indicator for being married |
| Rental Price | Current household hold rental price if a renter |
| Renter | Indicator for whether housing is rented |
| Moved | Indicator that equals 1 if household head moved within the prior year |
| Owns Car | Indicator for whether the household has access to a private vehicle (owns, leases, etc.) |
| Number Cars | Number of cars available in the household |
| Public Transit | Household head: indicator that equals 1 if primary means of commuting is by public transit, and 0 if by other means. |
| Auto | Household head: indicator that equals 1 if primary means of commuting is by private automobile, and 0 if by other means. |
| Walk | Household head: indicator that equals 1 if primary means of commuting is by walking, and 0 if by other means. |
| Miles | Household head: average number of miles commuted one-way to work. |
| Minutes | Household head: average time in minutes commuted one-way to work. |

Commuters defined employed individuals that work, but not at home. Commute type defined as the primary means by which employed individuals get from home to work.

# Bibliography


**Andersson, Fredrik, John C Haltiwanger, Mark J Kutzbach, Henry O Pollakowski, and Daniel H Weinberg.** 2018. “Job displacement and the duration of joblessness: The role of spatial mismatch.” *Review of Economics and Statistics*, 100(2): 203–218.

**Atkinson-Palombo, Carol.** 2010. “Comparing the capitalisation benefits of light-rail transit and overlay zoning for single-family houses and condos by neighbourhood type in metropolitan Phoenix, Arizona.” *Urban studies*, 47(11): 2409–2426.

**Baum-Snow, Nathaniel.** 2010. “Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960-2000.” *American Economic Review*, 100(2): 378–82.

**Baum-Snow, Nathaniel, and Matthew E Kahn.** 2000. “The effects of new public projects to expand urban rail transit.” *Journal of Public Economics*, 77(2): 241–263.

**Baum-Snow, Nathaniel, Matthew E Kahn, and Richard Voith.** 2005. “Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000 [with Comment].” *Brookings-Wharton papers on urban affairs*, 147–206.

**Bayer, Patrick, Stephen L Ross, and Giorgio Topa.** 2008. “Place of work and place of residence: Informal hiring networks and labor market outcomes.” *Journal of political Economy*, 116(6): 1150–1196.

**Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer.** 2019. “Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice.” National Bureau of Economic Research.

**Billings, Stephen B.** 2011. “Estimating the value of a new transit option.” *Regional Science and Urban Economics*, 41(6): 525–536.

**Brueckner, Jan K, and Yves Zenou.** 2003. “Space and unemployment: The labor-market effects of spatial mismatch.” *Journal of Labor Economics*, 21(1): 242–262.

**Central Puget Sound Regional Transit Authority.** 2006. "Link Light Rail Projects." *Sound Transit Quarterly Report*, Nov 17, 2006. Retrieved from https://web.archive.org/web/20061117033414/http://www.soundtransit.org/x1171.xml.

**Clark, Hugh M.** 2017. "Who rides public transportation." *Washington, DC: American Public Transportation Association.*

**Credit, Kevin.** 2018. "Transit-oriented economic development: The impact of light rail on new business starts in the Phoenix, AZ Region, USA." *Urban Studies*, 55(13): 2838–2862.

**Crump, Richard K, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik.** 2009. "Dealing with limited overlap in estimation of average treatment effects." *Biometrika*, 96(1): 187–199.

**de Chaisemartin, Clément, and Xavier d'Haultfoeuille.** 2019. "Two-way fixed effects estimators with heterogeneous treatment effects." National Bureau of Economic Research.

**Fitzgerald, John, Peter Gottschalk, and Robert Moffitt.** 1998. "An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics." National Bureau of Economic Research.

**Freedman, Matthew.** 2013. "Targeted business incentives and local labor markets." *Journal of Human resources*, 48(2): 311–344.

**Garrett, Thomas.** 2004. "The Costs and Benefits of Light Rail." *The Central Banker, Fall.*

**Gibbons, Stephen, and Stephen Machin.** 2005. "Valuing rail access using transport innovations." *Journal of urban Economics*, 57(1): 148–169.

**Glaeser, Edward L, Matthew E Kahn, and Jordan Rappaport.** 2008. "Why do the poor live in cities? The role of public transportation." *Journal of urban Economics*, 63(1): 1–24.

**Ham, John C, Charles Swenson, Ayşe İmrohoroğlu, and Heonjae Song.** 2011. "Government programs can improve local labor markets: Evidence from state enterprise zones, federal empowerment zones and federal enterprise community." *Journal of Public Economics*, 95(7-8): 779–797.

**Heilmann, Kilian.** 2018. "Transit access and neighborhood segregation. Evidence from the Dallas light rail system." *Regional Science and Urban Economics*, 73: 237–250.

**Hellerstein, Judith K, Melissa McInerney, and David Neumark.** 2011. "Neighbors and coworkers: The importance of residential labor market networks." *Journal of Labor Economics*, 29(4): 659–695.

**Hirano, Keisuke, Guido W Imbens, and Geert Ridder.** 2003. "Efficient estimation of average treatment effects using the estimated propensity score." *Econometrica*, 71(4): 1161–1189.

**Holian, Matthew J, and Matthew E Kahn.** 2015. "Household carbon emissions from driving and center city quality of life." *Ecological Economics*, 116: 362–368.

**Ihlanfeldt, Keith R.** 2003. "Rail transit and neighborhood crime: the case of Atlanta, Georgia." *Southern Economic Journal*, 273–294.

**Ihlanfeldt, Keith R, and David L Sjoquist.** 1998. "The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform." *Housing policy debate*, 9(4): 849–892.

**Kahn, Matthew E.** 2007. "Gentrification trends in new transit-oriented communities: evidence from 14 cities that expanded and built rail transit systems." *Real Estate Economics*, 35(2): 155–182.

**Kain, John F.** 1968. "Housing segregation, negro employment, and metropolitan decentralization." *The quarterly journal of economics*, 82(2): 175–197.

**LeRoy, Stephen F, and Jon Sonstelie.** 1983. "Paradise lost and regained: Transportation innovation, income, and residential location." *Journal of Urban Economics*, 13(1): 67–89.

**McMillen, Daniel P, and John McDonald.** 2004. "Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999." *Real Estate Economics*, 32(3): 463–486.

**McQuilkin, Alec.** 2020. "Public Rail Transit and Commuting: Evidence From 30 U.S. Metropolitan Areas." *Working Paper.*

**Pang, Jindong.** 2017. "Do subways improve labor market outcomes for low-skilled workers." *working paper, Syracuse University.*

**Redding, Stephen J, and Matthew A Turner.** 2015. "Transportation costs and the spatial organization of economic activity." In *Handbook of regional and urban economics*. Vol. 5, 1339–1398. Elsevier.

**Rosenthal, Brian M.** 2017. "The Most Expensive Mile Of Subway Track on Earth: [News; Series]." Name - Metropolitan Transportation Authority-New York City NY; Copyright - Copyright New York Times Company Dec 29, 2017; Last updated - 2018-01-17.

**Schuetz, Jenny.** 2015. "Do rail transit stations encourage neighbourhood retail activity?" *Urban Studies*, 52(14): 2699–2723.

**Severen, Christopher.** 2018. "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification."

**Tsivanidis, Nick.** 2018. "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." *Unpublished manuscript.*

**Turner, Matthew A.** 2019. "Local Transportation Policy and Economic Opportunity." *Policy Proposal*, 3.

**Tyndall, Justin.** 2019. "The Local labour market effects of light rail Transit."

**Voith, Richard.** 1991. "The long-run elasticity of demand for commuter rail transportation." *Journal of Urban Economics*, 30(3): 360–372.

**Voith, Richard.** 1993. "Changing capitalization of CBD-oriented transportation systems: Evidence from Philadelphia, 1970–1988." *Journal of urban economics*, 33(3): 361–376.

**Voith, Richard.** 1997. "Fares, service levels, and demographics: what determines commuter rail ridership in the long run?" *Journal of Urban Economics*, 41(2): 176–197.

**Yen, Maysen.** 2020. "Does Rapid Transit and Light Rail Infrastructure Improve Labor Market Outcomes?" *Available at SSRN 3650709.*