ISBN

9798290917313

Author

Bernal, Henry

Publication Date

2025-08-15

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Travel Behavior and Transportation Policy

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Henry Bernal

Dissertation Committee:
Professor David Brownstone, Chair
Professor Jan Brueckner
Professoor Matthew Freedman

2025

DEDICATION

Dedicated to my parents,
Henry and Mildred Bernal,

my older brother,
Kendrick,

my uncle who is resting with God,
Ronnie Steven Castro,

and my grandmother who mourns her son,
Agustina Castro.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank each of my committee members: my co-chairs, David Brownstone and Jan Brueckner, as well as Matt Freedman, for their invaluable contributions to my work but also for guiding me through so many aspects of graduate life.

I also wanted to give a special thanks to faculty members over at the Institute for Transportation Studies: they are Professors Jean-Daniel Saphores, Michael Hyland, Stephen Ritchie, and Elisa Borowski. I express my greatest gratitude to Victoria Valentine Deguzman, who facilitated many seminars that opened my eyes to the possibilities of transportation research, and who I know well since my time at USC.

I thank my fellow graduate students for providing encouragement and moral support, especially during my most difficult moments. They are Peter Li, Zhongying Gan (now at Roger Williams University), Idil Tanrisever, Noah Kouchekinia, Santiago Campos-Rodríguez, Siwei Hu, and Llorenç Miquel i Sole, and Jooneui Hong.

I thank my family, my parents, and my older brother for taking the time to listen to me in my darkest hour. I also thank my former running coach, Jane Springston, for encouraging me to find my love with running while graduate school became overwhelming.

I want to also thank the Graduate Division, School of Social Sciences, and Department of Economics for funding support. I also ackowledge the support provided by the Statewide Transportation Research Program, offered through the University of California Institute for Transportation Studies.

Chapter 3 of this dissertation is a reprint of the material as it appears in "The Impacts of Bus Use on COVID-19 Dispersion. In: Loukaitou-Sideris, A., Bayen, A.M., Circella, G., Jayakrishnan, R. (eds) Pandemic in the Metropolis. Springer Tracts on Transportation and Traffic, vol 20. Springer, Cham," used with permission from Springer Nature. The co-author listed in this publication is David Brownstone, who directed and supervised research which forms the basis for the dissertation.

VITA

Henry Bernal

EDUCATION

Doctor of Philosophy in Economics **2025**
University of California, Irvine *Irvine, California*

Master of Arts in Economics **2019**
University of Southern California *Los Angeles, California*

Bachelor of Science in Economics/Mathematics **2017**
University of Southern California *Los Angeles, California*

EMPLOYMENT

Consultant **2025–present**
Unison Consulting, Inc. *Laguna Hills, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2021–2025**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2019–2021**
University of California, Irvine *Irvine, California*

AWARDS

Ken Small Award for Best Student Paper in Transportation Economics **2023**
Department of Economics, UCI

ABSTRACT OF THE DISSERTATION

Essays in Travel Behavior and Transportation Policy

By

Henry Bernal

Doctor of Philosophy in Economics

University of California, Irvine, 2025

Professor David Brownstone, Chair

Chapter 1 fills a gap in the existing urban and transportation economics, which has discussed the sale impacts of proximity to transit and those of congestion pricing as seemingly unrelated phenomena. Using Singapore's cordon toll model, this paper tests whether a congestion rate hike that went into effect November 1, 2010 can model how an exogenous shock would affect the "premium" paid for a property closer to transit outside the affected cordon areas. Using a differences-in-differences approach with data on residential real estate transactions in 2010 and 2011, there is evidence of a modest increase in the transit premium outside the cordon following the November 1, 2010 toll increase.

Chapter 2 reviews the impacts of LA Metro's GoPass program, which provides free transit to community college students. Using a discrete choice experiment (DCE), students chose between transit and alternative modes across trip scenarios of varying lengths. Analyses from discrete choice models revealed that GoPass participation improved the probability of bus use in the experiment, but not train use, while having a driver's license decreased the probability of choosing transit. Values of travel time savings (VTTS) ranged from \$27 and \$54 per hour, while respondents valued WiFi availability between \$6 and \$7 per trip. Policy simulations demonstrated that modest reductions in travel time on transit yielded considerably larger impacts on mode choice and consumer surplus than zero-fare transit. Nonetheless, a GoPass-

style policy intervention was found to be a cost-effective intervention while delivering modest welfare gains of 19 to 60 cents per person per trip. Congestion charges, while they promoted mode shifts away from cars, were found to be welfare-reducing. These findings underscore the role that transit reliability and reductions in travel times play in improving mobility among community college students.

Chapter 3 examines how bus use impacts the transmission of the COVID-19 virus in urban areas, focusing on the evolution of the COVID-19 pandemic in Los Angeles County. Using data from the Los Angeles County Metropolitan Transportation Authority on station-level ridership in October 2019, April 2020, and October 2020, station-level ridership was imputed for other months in the study period and then mapped to 231 Countywide Statistical Areas (CSAs) in Los Angeles County, which were used by the Los Angeles Department of Public Health to report community COVID-19 transmission. We obtain CSA-specific COVID-19 case counts between March 16, 2020 and January 31, 2021 to create a monthly panel of bus ridership and COVID-19 cases. After using a dynamic panel regression, the findings provide no evidence that increased ridership levels or trip lengths are associated with higher incidence of COVID-19 at the CSA level in Los Angeles County in the period between June 2020 and January 2021.

Chapter 1

Does congestion pricing affect the "transit premium"? Evidence from Singapore

1.1 Introduction

Road congestion pricing is commonly believed to be a solution to traffic congestion in many cities, whose advocates date as far back as Pigou (1920) and Vickrey (1963). The intuition behind it is to regulate demand during times of high traffic volume on roads that experience the highest congestion levels. Generally, it has less deleterious welfare impacts than other ways of dealing with congestion, such as transit subsidies or fuel taxes (Parry and Bento, 2002). Furthermore, congestion pricing has the benefit of imposing costs on consumers who otherwise would not take into account congestion-related externalities, such as inefficient sprawl, spread-out cities, and high commute times (Brueckner, 2014). In practice, the implementation of congestion pricing can take several forms, ranging from single-lane freeway tolls to cordon tolls in central cities as seen in London, Stockholm, and Singapore. This work intends to focus on the cordon toll model used in Singapore.

Under a cordon tolling regime, a commuter pays a single toll whenever he chooses to drive into a zone that surrounds the city's central business district (CBD). As such, payment of the toll is triggered whenever the boundary of the zone is crossed, thus creating a "cordon". Singapore's cordon has been in place since 1975, when it first created the Area Licensing Scheme (ALS), in which commuters into the CBD, demarcated as the "Restricted Zone" (RZ) during morning peak hours, were required to display a pre-purchased daily or monthly windshield permit, with heavy fines issued to violators. In 1998, this scheme was changed to the current Electronic Road Pricing (ERP) system, where the toll is charged electronically by cars that have a transponder fitted with a fare card inserted. Currently, Singapore has two different types of tolls: a cordon toll to regulate entry into the CBD and central areas of the city, and congestion tolls on major expressways to regulate road demand during rush-hour traffic. Moreover, the amount of the toll levied is reviewed and changed on a quarterly basis.

Several works have studied the impacts of the cordon pricing scheme in Singapore. An early study by Wilson (1988) suggests there was an aggregate welfare loss resulting from the toll, although bus ridership benefited following the implementation of the cordon toll. Olszewski and Xie (2005) find evidence of the Singapore cordon being an effective tool for curbing congestion thanks in part to private vehicles exhibiting the highest elasticity of traffic demand relative to other vehicle types. Furthermore, Agarwal et al. (2015) find that a 2010 toll hike led to a decline in retail real estate prices within the cordon relative to prices outside the cordon, while private office and residential real estate prices saw no noticeable effects.

While informative, this body of literature focuses primarily on direct impacts and externalities that are created from the cordon toll. In other words, the available literature focuses on questions such as whether imposing the cordon toll directly led to increases in real estate prices, whether being in the cordoned area implied any change to property values, or how effective the toll was at changing commuters' travel behaviors into the central business district. This paper aims to address some of the more latent impacts of the cordon, by seeing how it impacts the "transit premium" that is paid for properties that have better access to transit, which is a well-documented phenomenon in the urban economics literature (Mulley et al., 2016; Diao et al., 2017) and derivative from the canonical standard urban model based on the works of Alonso (1964), Muth (1969), and Mills (1967). Fesselmeyer and Liu (2018) deal with effects on the transit premium in Singapore directly, but study how transit network expansion impacts the transit premium.

More generally, this paper attempts to unite two seemingly unrelated phenomena in urban economics that both impact property values: on one hand, it is well-established that congestion pricing can affect real estate prices while on the other, it can be observed that access to public transit infrastructure also has an impact on property values. To unite these perspectives, the present paper collects data from real estate transactions in Singapore in 2010 and 2011 to trace the impacts of a November 1st, 2010 toll increase on the willingness to pay for accessibility to transit and assess how these impacts varied throughout the city following the toll hike. Using a differences-in-differences (DD) approach similar to that found in Fesselmeyer and Liu (2018), the present paper finds that the toll increase was linked with a 0.6 percentage point increase in the transit premium in properties located outside the cordons.

This paper is structured as follows: Section 1.2 includes a brief overview of Singapore's housing market, Singapore's congestion pricing scheme and the 2010 toll increase, and Singapore's mass transit system (known as Mass Rapid Transit, or MRT). Section 1.3 presents the data sources used to conduct the empirical analysis and sets up an empirical framework to test whether transit premia were affected by the toll increase. Section 1.4 discusses the results and additional tests, while Section 1.5 discusses the implications of this paper's findings. Section 1.6 is the conclusion.

1.2 Institutional Background

1.2.1 Housing in Singapore

Singapore is an island city-state that spans 716 square kilometers and had a population of approximately 5.1 million in its 2010 census, of which 3.8 million were either citizens or permanent residents (Singapore Department of Statistics, 2010). Singapore has a unique housing market because of its two-tiered nature; a dominant public housing market which houses over 80 percent of residents (Housing and Development Board, nd) coexists with a smaller private housing market that operates like a more traditional laissez-faire market. The public housing stock mostly consists of high-rise buildings developed by Singapore's Housing Development Board (HDB), an agency established in 1960 with the goal of building housing in an effort to resettle squatter settlements. By 1964, the agency pivoted to to being a pathway to homeownership for the vast majority of Singaporeans, which enabled residents to use their compulsory savings and pension accounts to service down payments and monthly mortgage installments.

In Singapore, land is allocated through two main sources: the first is freehold land held by the private sector, which was held by private owners in perpetuity. The second source is land that is allocated through the Urban Redevelopment Authority (URA), which made new land available to acquisition and land reclamation projects, serving in effect as a public "land bank". Land for non-institutional uses would be leased to the highest bidder at public auctions, and private developers would be required to submit development plans for the parcels sold. In essence, the government tightly controlled both the quantity of land released to the private sector and its use and development. Land purchased from the government, either as an HDB unit or for private development, were typically acquired on 99-year leaseholds.

Traditionally, private housing catered exclusively to the highest income earners who were not eligible to purchase for subsidized housing from the government. This, along with high restrictions on land leased from the government to private developers, implied that private units sold at much higher premiums than comparable public housing units. By the mid- to late-1980s, as Singaporeans became wealthier, many of those who desired to own private properties could not do so due to supply constraints in the private housing market, creating an affordability crisis coupled with long waiting lists for HDB housing. As the government relaxed some HDB restrictions and expanded the land sold to private developers in the 1990s, existing HDB units could be resold in the private market at a premium. New developments in Singapore increasingly fell within the purview of the private market because the Singaporean government expanded 99-year land leases to private developers to build high rise communities that produced a higher number of affordable units for middle-class residents (Lum, 2002). To date, Singapore has some of the highest property values in the world, which are highly influenced by price movements in the public sector.

1.2.2 Urban Road Pricing in Singapore

Throughout the 2000s, the population of motor vehicles has steadily risen per Land Transport Authority (LTA) statistics, which signaled increased levels of congestion on Singapore's roads. Private cars increased from a total of 400,000 around 2005 to over 540,000 in 2013 (shown in Figure 1.1). This comes despite Singapore having strict quotas and high taxes on private vehicles, in an effort to drive citizens to taking public transportation (Ho et al., 2018).[1] Additionally, because Singapore is connected to Malaysia through a bridge and causeway, Malaysian-registered vehicles pay a daily vehicle entry permit (VEP) fee (in 2010,

[1]Since 1990, Singapore limits new vehicle registrations by issuing Certificates of Entitlement (COEs) and limiting the number of COEs issued, which the government auctions twice a month. Furthermore, the government imposes a tax known as an Additional Registration Fee (ARF), which was 100 percent of the wholesale import price of the vehicle in 2018.

the fee was S$20; in 2022, the fee was S$35) to enter Singapore from Malaysia.[2]

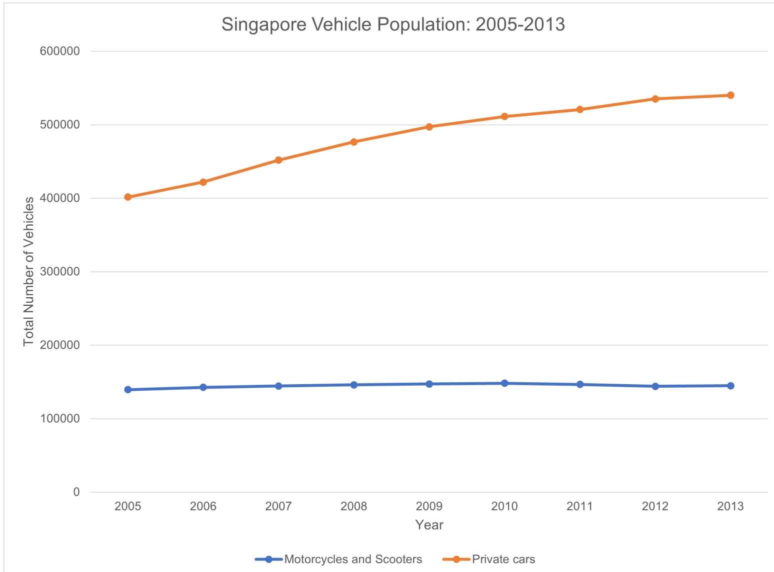


Figure 1.1: The total population of private cars, motorcycles and scooters in Singapore between 2005 and 2013.

Singapore introduced its road pricing system in 1975, via the Area Licensing Scheme (ALS), which was a manually operated system where vehicles were checked for a permit to enter the CBD. By 1998, this system was replaced in favor of the current electronic road pricing (ERP), which then consisted of a network of gantries, or overhead signs, which required commuters to have a transponder fitted onto the dashboard that contained a card with enough balance for the toll, which would be charged electronically. Initially, these gantries were only surrounding the CBD (also known as the Bugis-Marina Cordon) and the Orchard Road shopping district, but more gantries have been placed on major expressways. In effect, the expansion of the ERP gantries has led Singapore to having two congestion pricing systems in its road network: the first being the "zonal" gantries which cordon off access to the CBD and the Orchard Road shopping district and "radial" gantries erected along expressways which do not pass through the CBD. The expressway tolls are only levied when entering the expressway and drivers can avoid these by using smaller, more congested roads.

[2]These fees are waived for Malaysian-registered vehicles for the first 10 calendar days of the year in which the fees would need to be paid, and VEP fees go into effect thereafter.

ERP rates are set based on traffic speeds of priced roads, and rates are adjusted when speeds fall below an optimal range defined by the LTA for arterial roads and expressways. The LTA subsequently conducts a quarterly review and releases new rates; any increases or decreases in ERP rates are public information.

Because the expressway tolls do not clearly demarcate an area of the city, identifying the effects of this system of congestion pricing is difficult given that these expressways are connected to arterial roads that provide access to different areas of the city. Hence, studying an event where expressway tolls are impacted is not practical for purposes of the present paper. Instead, changes only to the cordon tolls will be used to identify how toll increases can be capitalized in property values *outside* the cordoned area.[3] Since the LTA tightly guards information related to ERP increases before its announcement and does so independently of property values, real estate markets do not have information regarding toll increases and thus, any increase in a cordon toll can be used as a policy shock variable.

This paper will study the same event as Agarwal et al. (2015), which looks at an increase in two of Singapore's cordon tolls (Orchard Road and Bugis-Marina Centre), which took effect November 1st, 2010. Drivers entering the cordons at 23 different ERP gantries were charged an additional S$1 per car, which represents a 50% increase in ERP rates at the Bugis-Marina Cordon, from $2 to $3 per car and a 100% increase in ERP rates at the Orchard Cordon, from $1 to $2 per car. In an October 2010 article in The Straits Times, local transportation journalist Christopher Tan noted the increase was "one of the most widespread in recent memory" and that "usually, ERP rate adjustments... affect anywhere from three to a dozen gantries."

[3]This is the same reasoning offered by Agarwal et al. (2015), modified to satisfy the empirical design of the present paper.

1.2.3 Mass Transit in Singapore

Due to land supply constraints for road construction, the Singaporean government undertook several feasibility studies in the late 1960s on the construction of a new rail system to alleviate congestion on its roads. The Parliament eventually approved a massive project, known as Mass Rapid Transit, known to be MRT in 1982, with construction of two lines beginning in 1983 at an initial cost of over S$5 billion (in 1982 Singapore dollars). The initial phase was to consist of the North-South Line (NSL) and the East-West Line (EWL), which were opened in phases beginning in 1987 until both lines were fully completed in 1990.

Since then, the MRT has been expanded to incorporate two more lines: the Northeast Line (NEL), which opened in 2003, and the Circle Line (CCL), which opened in phases between 2009 and 2012. In 2013, the Singaporean government has since announced intentions to expand the MRT network to 360km by 2030, aiming to connect 8 in 10 Singapore residents to the nearest MRT station within a 10 minute walk. This plan involves extending the Circle Line and building two new lines: the Cross Island Line (CRL) and the Jurong Region Line (JRL). At the time the plan was unveiled, two lines were already in the construction or planning stages: the Downtown Line (DTL), which opened in 2013, and the Thomson-East Coast Line (TEL) opened in 2020. As of 2023, six lines and 134 total stations comprise the MRT network, which served 2.8 million daily passenger trips in 2022 per LTA data.[4]

[4] Available at `https://www.lta.gov.sg/content/dam/ltagov/who_we_are/statistics_and_publications/statistics/pdf/PT_Ridership_Yearly_2015-2022.pdf`

1.3 Empirical Framework

1.3.1 Data

To test the effects on housing prices, the present paper uses private residential real estate transactions for calendar years 2010 and 2011[5]. These data were obtained through the Real Estate Information System (REALIS) of Singapore's Urban Redevelopment Agency (URA). This dataset includes data on each property's address, size (floor space), property type, sale type (new sale, resale, or sub-sale[6]), purchaser type (private or public), dwelling age, and sale price. The address also contains the floor number of each property, which was extracted and added to the dataset. Finally, this dataset includes a HDB indicator, which is a demographic variable on whether the purchaser of each property was living in an HDB unit at the time of the purchase.

To determine whether properties were outside the affected cordon boundaries, the coordinates of all transacted properties were mapped against the a shapefile of the cordon boundaries using geographic information software (namely QGIS). Properties that were inside the affected boundaries were excluded from the regression sample. To compute access to the nearest MRT station, it was necessary to geocode (using QGIS) each station in Singapore's MRT to compute the linear distance from each property. Because a handful of MRT expansions occurred during this study period, the process was repeated each time a new station opened, with the distances recomputed. In addition to the distance from each property to the nearest MRT station, the distance from each property sold outside the cordon to the cordon's boundary was computed, and properties outside the cordon were further grouped based on their distance from the cordon Table 1.1 below lists the key regression variables

[5]These data were kindly supplied by Tien Foo Sing; no public housing sales transactions are included in this dataset.

[6]In Singapore, a sub-sale is defined as a transaction where the original buyer of a new property sells the property to another buyer before construction is complete.

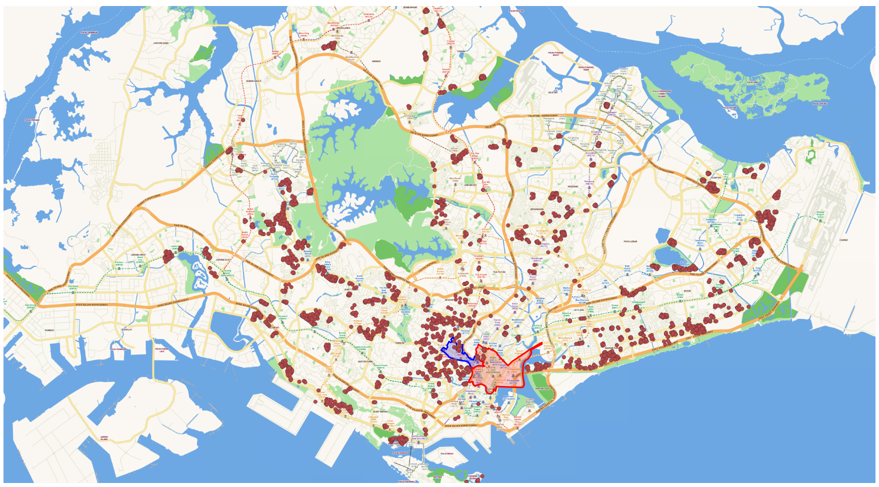
included in the dataset. Properties within 0.5 kilometers of the nearest MRT station were placed in a "treatment" group while properties whose nearest MRT station was farther away were placed in a "control" group.

Table 1.1: List of variables and descriptions.

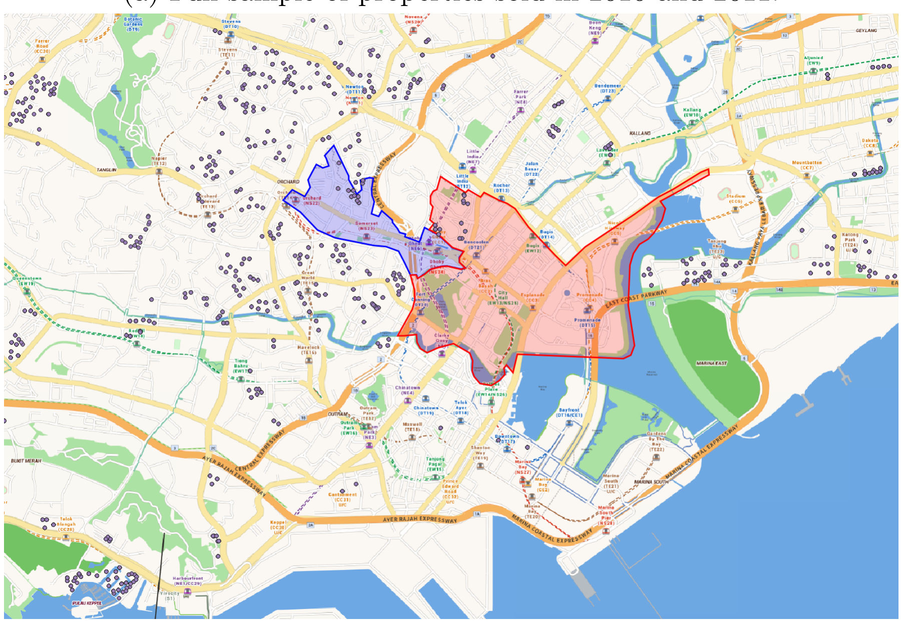
Variable	**Description**
$\log(Price)$	Natural logarithm of transacted price per square meter
Post	An indicator for properties sold after November 1st, 2010
MRT	An indicator for properties within 0.5 km of the nearest MRT station
D_{resale}	An indicator for properties that were resold
$D_{sub-sale}$	An indicator for transactions that were sub-sales
D_{HDB}	A dummy variable equal to 1 if the property's purchaser was living in an HDB property at the time of the purchase
Age	The age of the building, defined as the year in which the property was sold minus the year the property was built
Area	Floor space of the property (in square meters)
Floor	Floor number of the sold property
$PctLeaseRem$	The number of years remaining on the lease at the time of the sale divided by the total length of the lease. Freeholds were automatically assigned a value of 100%.

After excluding transactions of properties inside the cordon boundaries, and omitting observations with unknown construction dates, the dataset contains 37,007 transactions. Of these, 17,592 take place before the toll increases and 19,415 take place after the toll increase.

Figure 1.1 gives a visualization of properties' locations and the boundaries of the affected cordons. Panel (a) maps the universe of private transactions in Singapore in 2010 and 2011, and panel (b) shows a closer visualization of the cordon boundaries where the tolls were increased in 2010. The blue and red regions are the affected cordons by the toll increase that went into effect on November 1st, 2010. The blue area is the Orchard Road cordon, and the red area is the Bugis-Marina Centre cordon.



(a) Full sample of properties sold in 2010 and 2011.



(b) Properties sold near the Central Area and cordon boundaries.

Figure 1.2: Map of Singapore with locations where sale transactions took place in 2010 and 2011.

Table 1.2: Summary statistics of key variables in the dataset.

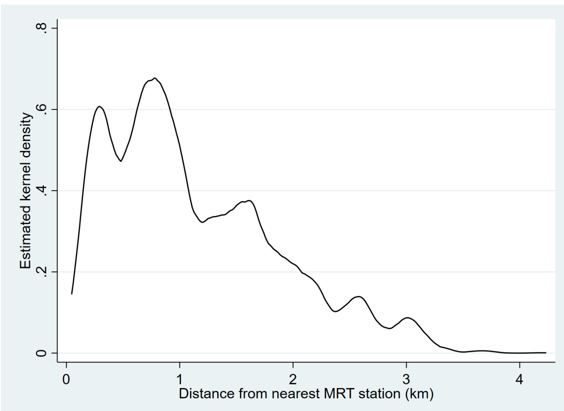
Variable	N	Mean	Std. Dev.	Min.	Max.
Panel A: Full sample					
Unit price (per m^2)	37,007	S$11,968.48	S$4,748.95	S$3,274	S$71,576
Distance to MRT (km)	37,007	1.135	0.750	0.047	4.234
Area (m^2)	37,007	120.933	123.649	33	11,036
Building age (years)	37,007	4.586	6.789	-4	33
Floor number	37,007	8.706	7.181	1	56
New sale	37,007	0.477	0.499	0	1
Sub sale	37,007	0.092	0.289	0	1
Resale	37,007	0.431	0.495	0	1
HDB purchaser	37,007	0.377	0.485	0	1
Panel B: Before toll increase					
Unit price (per m^2)	17,592	S$11,611.49	S$4,807.90	S$3,274	S$49,517
Distance to MRT (km)	17,592	1.165	0.792	0.047	4.234
Area (m^2)	17,592	124.576	65.377	33	11,036
Building age (years)	17,592	5.081	6.922	-4	32
Floor number	17,592	9.351	7.548	1	56
New sale	17,592	0.419	0.493	0	1
Sub sale	17,592	0.101	0.301	0	1
Resale	17,592	0.480	0.500	0	1
HDB purchaser	17,592	0.352	0.478	0	1
Panel C: After toll increase					
Unit price (per m^2)	19,415	S$12,291.95	S$4,671.52	S$4,672	S$71,576
Distance to MRT (km)	19,415	1.107	0.707	0.047	4.234
Area (m^2)	19,415	117.633	158.895	33	11,036
Building age (years)	19,415	4.137	6.635	-4	33

Floor number	19,415	8.122	6.779	1	56
New sale	19,415	0.530	0.499	0	1
Sub sale	19,415	0.083	0.277	0	1
Resale	19,415	0.386	0.487	0	1
HDB purchaser	19,415	0.399	0.490	0	1

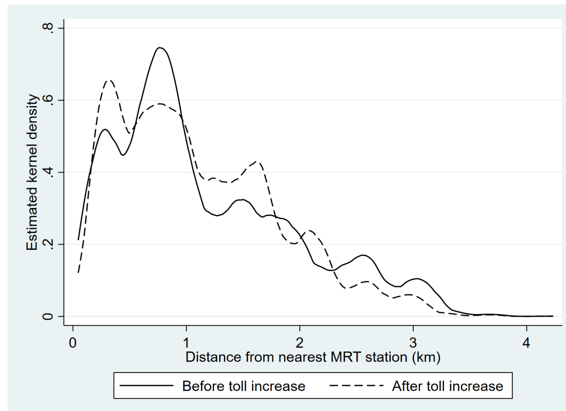
Table 1.2 reports the descriptive statistics for the regression sample in this paper. The average unit price (in Singapore dollars) was S$11,968.48 per square meter in the full sample, with the average unit price being S$11,611.49 pre-increase and S$12,291.95 post-increase. Distance to the nearest MRT station averaged 1.135 kilometers, with little difference pre- compared to post-increase (1.165 pre-increase versus 1.107 post-increase.) Of the properties sold, roughly 48 percent were new sales, 9 percent were sub-sales, the remaining 43 percent being resales. The average area of the properties in the full sample was 120.933 square meters, with the average being 124.576 square meters before the toll increase and 117.633 after the increase. Properties sold had an average dwelling age of about 4.6 years, with the average being 5.1 years pre-increase and 4.1 post. The mean floor number was lower post-increase, with the mean floor number being 9.351 pre-increase compared to 8.122 post-increase and 8.706 for the full sample. New sales made up a larger share of transactions post-increase compared to pre-increase, with new sales making up 53 percent of sales after the increase compared to 41.9 percent before the increase. The shares of sub sales and resales declined from the pre-increase subsample to the post-increase subsample, declining from 10.1 percent and 48 percent down to 8.3 percent and 38.6 percent for sub sales and resales, respectively. The share of purchasers with an HDB address was 37.7 percent overall, increasing from 35.2 percent pre-increase to 39.9 percent post-increase.

Since one of the key variables is the distance of each transacted property from its nearest

MRT station, it is relevant to look at this variable more closely. This is because the toll increase may also have shifted the distribution of the distance from transacted properties to the nearest MRT station. Figure 1.2 shows the estimated kernel density plots for the distribution of the distance to the nearest MRT station. Panel (a) shows the plot for the full sample, while panel (b) shows the plots for the pre- and post-increase subsamples.



(a) Estimated kernel density of distance to MRT for full regression sample.



(b) Estimated kernel density of distance to MRT for pre- and post-increase subsamples.

Figure 1.3: Estimated kernel densities of distances to MRT stations.

Panel (a) shows that because of Singapore's dense urban environment, there is a high concentration of transacted properties within 1 kilometer of the nearest MRT station. However, panel (b) shows there are some subtle differences in the density plots before the increase compared to after: first, the pre-increase subsample has fatter tails, meaning a more properties transacted before the toll increase were 3 kilometers away or farther from the nearest MRT station compared to after the increase. Second, the density plot shows an earlier peak for the post-increase subsample, meaning that more properties sold were closer to MRT after the toll increase compared to before. While not definitive, this shift could suggest a stronger preference for properties closer to MRT after the increase, even if this is partly due to MRT expansion making the average property closer to its nearest MRT station.

1.3.2 Empirical Model

To empirically assess whether the toll increase appreciably changed the transit premium, a differences-in-differences estimator is employed on transactions assigned to the treatment group using both a MRT dummy, which is equal to 1 if the building is located within 0.5 kilometers of the nearest MRT station and 0 otherwise, and a "post" dummy, which is equal to 1 for transactions that took place after November 1st, 2010 and 0 otherwise. The baseline specification has the log-real estate price, log(Price) as the dependent variable and includes MRT, $POST$, and $POST \times MRT$ to test the "diff-in-diff" effects on the transit premium before and after a cordon toll increase on properties outside the cordon. Letting i denote the property sold in building b during month t, the baseline estimating equation is therefore

$$\log(Price)_{ibt} = \alpha + \beta_1 MRT_{bt} + \beta_2(POST_t \times MRT_{bt}) + \sum_j \eta_j T_{ijt} + \gamma_b + \epsilon_{ibt} \tag{1.1}$$

where α represents the intercept, β_1 represents the level of the transit premium, and β_2, the coefficient of interest, represents the differential premium for MRT access after the toll increase. T_{ijt} is a set of monthly indicators that is equal to one if the transaction was made in month j and 0 otherwise (in other words, the only T_{ijt} indicator that is set to 1 is the one for which $j = t$) and η_j is the respective coefficient of each monthly indicator. γ_b represents unobserved time-invariant building fixed effects based on the 6-digit postal code of the building, and ϵ_{ibt} is the error term.

If the toll increase is significant enough to the point that buyers' preferences shift to locations closer to MRT stations, then β_2 should be positive. However, the sign of β_1 is not obvious from this model: while the first-order MRT term is included in the estimating equation to avoid omitted variable bias in estimating β_2, reported values of $\hat{\beta}_1$ will not be especially reliable for identifying the *level* of the transit premium. This is because the estimating equation only draws from within-building variation, making identification of a MRT coefficient weaker and its estimate difficult to interpret since a coefficient would only be identifiable from buildings which initially were not within 0.5km of their nearest MRT station, but through construction of new MRT stations, did fall within 0.5km of their nearest station at the end of the study period in the dataset. However, the main objective of this paper is to identify the *change* in the transit premium post-increase, rather than finding the level of transit premium.

This model can be modified to also control for hedonic attributes, sale type, and HDB purchaser status from the transactions data. The specification defined in (1.2) is a modification of the estimating equation defined in (1.1) which now includes $\mathbf{X_{ibt}}$, a vector of sale type indicators, hedonic attributes, and the HDB variable. The estimating equation now changes to

$$\log(Price)_{ibt} = \alpha+\beta_1 MRT_{bt}+\beta_2(POST_t \times MRT_{bt})+\mathbf{X'_{ibt}}\theta+\sum_k \delta_k F_{ikt}+\sum_j \eta_j T_{ijt}+\gamma_b+\epsilon_{ibt} \tag{1.2}$$

where θ is the vector of coefficients on the variables included in $\mathbf{X_{ibt}}$ (see Table 1.1). F_{ikt} is a collection of indicator variables for the floor number of property i, which is equal to 1 if the floor number of the property sold is equal to k and δ_k represents the premium for a property on the k-th floor.

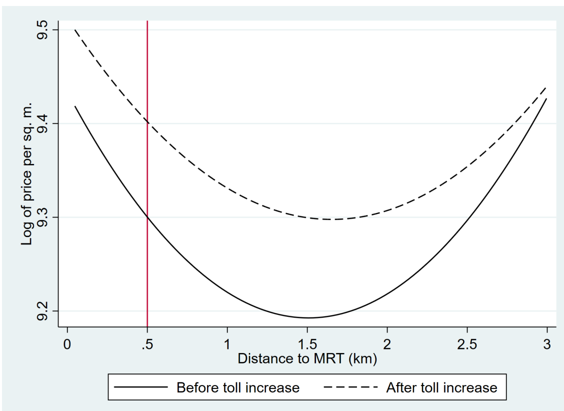
It is necessary to be careful when controlling for the floor number in the estimating equation. This is because units closer to the city center are more likely to be taller buildings, and will therefore, have higher floors, thus creating potential endogeneity. However, the floor in which a property is sold remains highly relevant in its transacted price because of unobserved amenities such as reduced exposure to insects and noise, and more privacy compared to lower floors even when only looking at within-building changes in property values. Additionally, it is likely to be the case that the premium for purchasing a property one floor higher will also be dependent on the floor levels being compared: for example, the premium for an fifteenth floor property compared to a fourteenth floor property is likely to be smaller than the premium for a second floor property compared to a first floor property. For this reason, the floor number is controlled for by using a set of floor indicator variables instead of a linear floor variable.

Finally, it is crucial to justify the use of building fixed effects instead of coarser geographic fixed effects, like the postal district or sector. Identification of β_2 in equations (1.1) and (1.2) is not precluded by using building fixed effects, but failing to use building fixed effects may create endogeneity, since the floor number could be correlated with the distance from the cordon, which is excluded from the estimating equation. Using coarser geographic fixed effects would facilitate identification of a first-order transit premium at the risk of failing to identify changes in the transit premium following the toll increase from the exclusion of unobservable property characteristics. Additionally, because of how the leases are issued by the government to developers, the variation in the amount of time remaining on leases is diminished significantly when only using within-building variation.

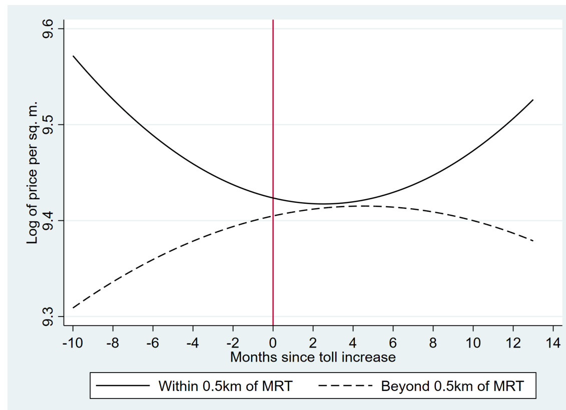
1.4 Results

1.4.1 Graphical Evidence

The analysis of this paper begins by plotting prices with respect to distance to the nearest MRT station and with respect to the months from the beginning in the sample in 2010. To plot the price against distance to the MRT, sale prices are averaged across postal codes and weighted by the number of units sold, and then are plotted using a quadratic fit and separated by transactions before the toll increase and transactions after the increase. To plot price against time, sale prices are averaged by month and weighted by number of units sold per transaction. Again, a polynomial fit of the average log price per square foot is plotted by month and this is done separately for properties within 0.5km of the nearest MRT station and beyond 0.5km of the nearest MRT station. Figure 1.3 shows these plots, with panel (a) showing the price plotted against distance to MRT, and panel (b) showing the price plotted against time.



(a) Polynomial fit of average log price per m^2 to distance to MRT, averaged by postal code.



(b) Polynomial fit of average log price per m^2 over months since January 2010, averaged by MRT treatment group.

Figure 1.4: Two-way graphs of log price per m^2 with respect to distance to MRT and months from January 2010.

While the plots in Figure 1.3 do not include controls, they do show some indication of changes to the transit premium following the toll increase. Panel (a) shows this rather clearly, with average prices being consistently higher in the post-increase sample until a distance of roughly 2.75 kilometers from the MRT station, but that the overall difference seems to be highest for those buildings that were sold within 0.5km to 1.5km of the nearest MRT station. At 0.5 kilometers from the nearest MRT station, prices per square meter averaged roughly S$12,088 (log price of approximately 9.4) per m^2 after the increase, compared to S$10,938 before the increase (log price of roughly 9.3). This difference, while not explained entirely by the toll increase (due to controls and monthly trends), merits further exploration. Next, panel (b) descriptively shows trends in the transit premium over time. In January 2010, the premium is rather large with properties within 0.5km of MRT far exceeding the price of properties beyond 0.5km of MRT. However, this price premium shrinks considerably, until it once again begins to increase in early 2011. While not as clear cut as with prices plotted against distance to MRT, the idea of the price premium trends changing in the months following the toll increase (while not fully explained by the toll increase based on the graph) is worthy of further exploration and provides visual support to the idea of the transit premium changing after the toll increase.

1.4.2 Main Results

The regression specifications defined in equations (1.1) and (1.2) are estimated using standard errors clustered by postal code. In both regression specifications, the default case is a property sold as a new sale to a non-HDB purchaser before November 1st, 2010 and whose location is farther than 0.5km from its nearest MRT station. Table 1.3 shows the results from both the baseline specification and the specification including sale attributes and the HDB purchaser indicator variable. Column (1) shows the ordinary least squares (OLS) regression results without including any building fixed effects, sale attributes, or demographic attributes. Column (2) shows the OLS results with sale attributes and the HDB purchaser variable added to the specification from column (1). Column (3) includes building fixed effects but no sale or demographic attributes, while column (4) shows the regression results with building fixed effects, sale attributes, and the HDB purchaser indicator variable included.

Table 1.3: Results of main regression specification.[1]

Variable	Dependent variable: Log price per m^2			
	(1)	(2)	(3)	(4)
MRT $\leq$ 0.5km	0.071**	0.051***	-0.004*	-0.004*
	(0.030)	(0.021)	(0.002)	(0.002)
MRT $\leq$ 0.5km $\times$ Post	-0.038	-0.030	0.006*	0.006**
	(0.030)	(0.023)	(0.003)	(0.003)
Area (logarithm)		0.003		-0.159***
		(0.021)		(0.008)
Age		-0.005***		0.004
		(0.001)		(0.005)
Resale		-0.032		-0.012
		(0.022)		(0.010)
Sub sale		0.045*		0.011
		(0.027)		(0.007)
HDB purchaser		-0.175***		-0.002**
		(0.007)		(0.001)
Percent of lease remaining		0.012***		0.002
		(0.001)		(0.010)
Floor number indicator variables	No	Yes	No	Yes
Building fixed effects	No	No	Yes	Yes
Monthly indicator variables	Yes	Yes	Yes	Yes
Sample size	37,007	37,007	37,007	37,007
$\bar{R}^2$	0.023	0.292	0.943	0.955

[1] Standard errors are in parentheses. Significance is denoted by * at the 10% level, ** at the 5% level, and *** at the 1% level.

Columns (1) and (2) are included in Table 1.3 to show that when variation is not restricted to within-building variation, the estimated values of $\hat{\beta}_1$ *do* represent a premium paid for properties close to transit and that these results are intuitive. Column (1) shows a transit premium of 7.1 percent without controlling for any sale attributes or the HDB indicator, while column (2) shows a transit premium of 5.1 percent once these attributes are controlled for. There is no significant change in the premium after the toll increase reported in columns (1) or (2). In column (2), area does not appear to have a significant effect on prices, while a one year increase in building age appears to decrease sale price by 0.5 percent. Resales do not appear to have a significant effect on sale prices in column (2) relative to new sales (the default case), but sub sales increased sale price by roughly 4.5 percent, though this significance is weak. Properties bought by a purchaser living in an HDB property sold for a 17.5 percent discount, while a one percentage point increase in the years remaining on the property lease increased sale price by 1.2 percent.

Once building fixed effects are included in columns (3) and (4), there is evidence of a significant and expected positive change in the transit premium following the toll increase, increasing by approximately 0.6 percentage points. While modest in size, this effect does show some evidence of transit becoming more desirable following a toll increase, which validates this paper's hypothesis. Additionally, lack of significance in columns (1) and (2) may be a product of functional misspecification from excluding the building fixed effects. Some of the coefficients of the sale attributes incorporated in the specification reported in column (4) are statistically significant: the elasticity to area is now negative, with the per unit sale price declining by 1.5 percent for a 10 percent larger unit in area. Age has no significant effect on the sale price, since only within-building variation is considered, and the percentage of the lease remaining has virtually no effect. Resales and sub sales are both insignificant in column (4), while properties sold to purchasers in HDB housing see a per square meter discount of only 0.2 percent per column (4). This is because although HDB purchasers will be more price sensitive in the private market, most of that behavior will be seen *across* buildings as they choose more inexpensive private developments. For the sake of brevity, individual floor effect coefficients are not reported, but behave as expected, with high-floor units commanding higher premiums over first floor units.

1.4.3 Heterogeneous Effects by Sale Category

It is helpful to consider how the effects of the toll increase on the transit premium vary across property subsamples in the data. Using the available data, it is possible to explore how the effect on the transit premium does differ across characteristics that are allowed to vary across properties within buildings, such as sale type, HDB purchaser status, and floor number. The present subsection will explore how these characteristics influence the change in the transit premium following the toll increase.

To empirically assess the different changes by subsample, a triple differences-in-differences (DDD) approach is employed. With the exception of the floor number, the indicator variables for resale, sub-sale, and HDB are each interacted with the $POST$, MRT, and $MRT \times POST$ variables and then these terms are added to the estimating equation specified in Equation (1.2). To deal with floor number, a new binary variable is created to indicate whether a property is above the seventh floor, which is the median floor number. As with the resale, sub-sale, and HDB indicators, this new indicator is now interacted with $POST$, MRT, and $MRT \times POST$ to create the triple difference-in-differences setup.

Table 1.4 reports the regression estimates of the triple difference-in-differences model when resale, sub sale, HDB, and high floor variables are interacted with MRT, $POST$, and $MRT \times POST$. Column (1) shows the regression results of the triple difference-in-differences model with the resale interaction terms included, column (2) shows the results with the sub sale interaction terms included, column (3) shows the results with the HDB interaction terms included, and column (4) shows the results with the high floor number interaction terms included. To avoid multicollinearity with the high floor indicator variable, only the created high floor indicator variable and its interaction terms were included in the regression in column (4) while individual floor number indicator variables were excluded.

Table 1.4: Results of test for heterogeneous effects.[1]

Variable	Dependent variable: Log price per m^2			
	(1)	(2)	(3)	(4)
Interacting variable:	Resale	Sub sale	HDB	High floor
MRT $\leq$ 0.5km	0.005	-0.004	-0.005*	-0.004*
	(0.003)	(0.002)	(0.003)	(0.003)
MRT $\leq$ 0.5km $\times$ Post	-0.004	0.006**	0.005	0.006
	(0.004)	(0.003)	(0.004)	(0.004)
Resale	-0.011	-0.015*	-0.012	-0.023**
	(0.010)	(0.009)	(0.010)	(0.010)
Resale $\times$ MRT $\leq$ 0.5km	-0.016***			
	(0.004)			
Resale $\times$ Post	0.018***			
	(0.005)			
Resale $\times$ MRT $\leq$ 0.5km $\times$ Post	0.018***			
	(0.005)			
Sub sale	0.016**	0.004	0.010	0.002
	(0.007)	(0.007)	(0.007)	(0.007)
Sub sale $\times$ MRT $\leq$ 0.5km		0.002		
		(0.006)		
Sub sale $\times$ Post		0.011*		
		(0.006)		
Sub sale $\times$ MRT $\leq$ 0.5km $\times$ Post		-0.004		

		(0.009)		
HDB	-0.002***	-0.002***	-0.007***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
HDB $\times$ MRT $\leq$ 0.5km			0.003	
			(0.003)	
HDB $\times$ Post			0.007***	
			(0.002)	
HDB $\times$ MRT $\leq$ 0.5km $\times$ Post			0.001	
			(0.004)	
Above 7th floor				0.058***
				(0.002)
Above 7th floor $\times$ MRT $\leq$ 0.5km				-0.000
				(0.003)
Above 7th floor $\times$ Post				-0.007**
				(0.003)
Above 7th floor $\times$ MRT $\leq$ 0.5km $\times$ Post				0.003
				(0.004)
Area (logarithm)	-0.159***	-0.159***	-0.159***	-0.156***
	(0.005)	(0.008)	(0.008)	(0.007)
Age	-0.002	0.004	0.004	0.002
	(0.005)	(0.005)	(0.005)	(0.005)
Percent of lease remaining	0.001	0.002	0.003	0.002
	(0.008)	(0.009)	(0.009)	(0.009)
Floor number indicator variables	Yes	Yes	Yes	n/a
Building fixed effects	Yes	Yes	Yes	Yes
Monthly indicator variables	Yes	Yes	Yes	Yes
Sample size	37,007	37,007	37,007	37,007

$\bar{R}^2$	0.963	0.963	0.943	0.955

[1] Standard errors are in parentheses. Significance is denoted by * at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.4 shows that transaction or building-level characteristics largely did not influence the effect the toll increase had on the transit premium, outside of whether the property was a resale. The coefficients of the triple interaction terms reflect percentage point adjustments to the change in the transit premium estimated in column (4) of Table Table 1.3, when the change to the transit premium was estimated for all properties in the regression sample. For resales, prices increased by 1.8 percent after the toll increase and the transit premium increased by 1.8 percentage points more compared to new sales. This result is intuitive since buying in the secondhand private market is a standard entry point into the private housing market from the public housing market; therefore, the argument that resale buyers may be more sensitive to toll increases than buyers who buy in other types of transactions would be reasonable, though not directly implied through these results.[7] For sub sales and properties with an HDB buyer, these were found to be a 0.4 percentage point decrease and a 0.1 percentage point increase, respectively, to the transit premium although neither are statistically significant. A similarly small change to the transit premium is observed among properties selling at or above the seventh floor, with properties on the seventh floor seeing a statistically insignificant 0.3 percentage point increase to the transit premium.

1.4.4 Sensitivity to Alternate MRT Distance Thresholds

To verify the robustness of the results found in column (4) of Table 1.3, a sensitivity analysis is performed by modifying the distance threshold of MRT treatment dummy used in the

[7] This argument, however, would only hold if supply of private housing near transit stations remained constant or did not increase at the same rate as demand. This paper does not study the supply side of the private market closely, but could serve as a point to explore further.

regression specifications. This threshold starts at 0.05 kilometers from the nearest MRT station and is gradually increased, in increments of 0.01 kilometer, up to 1 kilometer to test the sensitivity of the estimated post-increase premium change. Exploring this effect is necessary because it is not obvious how large a catchment area would be required for the toll increase to be capitalized in transaction prices, even though the present paper uses 0.5 kilometers as a distance threshold for the MRT catchment in the regressions. The results of this sensitivity analysis are displayed graphically in Figure 1.5. [8]

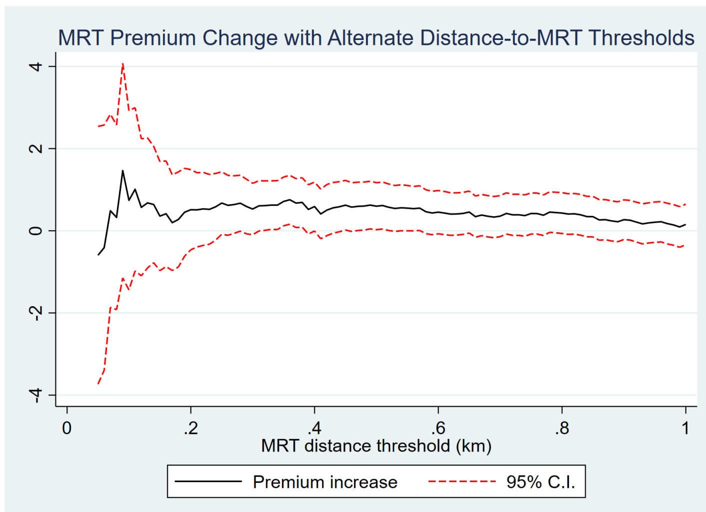


Figure 1.5: Estimated transit premium increase at different MRT treatment distance thresholds.

When the MRT treatment distance threshold is set very low (i.e. closest to 0 kilometers), the estimated premium change from the toll increase is very noisy, and the point estimate is insignificant from zero. The estimated premium change stabilizes when the threshold distance is set to 0.4 to 0.6 kilometers, to a premium change of close to 0.6 percentage points, which is the value reported in 1.3. At distances greater than 0.6 kilometers, the estimated post-increase premium change begins to decline gradually as the threshold distance increases, reaching zero the MRT treatment distance threshold is set to 1 kilometer.

[8]To elaborate further, this sensitivity analysis is done by replacing the MRT treatment variable and "switching" it on at a distance as small as 50m (0.05km) from the nearest MRT station, re-estimating the regression results in column (4) of Table 1.3, and then repeating this procedure by increasing the catchment area in increments of 10m (0.01km).

1.4.5 Announcement Effects

The present paper also explores whether incumbent residents are "pricing in" the toll increase prior to its taking effect. This possibility arises because toll increases are typically announced within a roughly 90-day period prior to the increase taking effect. To test for announcement effects empirically, a simple falsification test is used: first, transactions that occurred after the toll increase (i.e. transactions after November 1, 2010) are discarded since these would capture effects from the true toll increase. Then, the $POST$ variable is redefined so that it set to "switch" on before the *actual* toll increase, which is repeated to set the "fake" toll increase 1, 2, 3, and 4 months before the real toll increase took effect, and this new $POST$ variable is interacted with the MRT treatment variable to determine if the transit premium was already changing before the toll increase went into effect. Finally, Equation (1.2) is estimated using the new $POST$ and $POST \times MRT$ variables. Table 1.5 reports the regression results when incorporating the "falsified" $POST$ variable. Column (1) shows the results when the event date of the toll increase is set to July 1, column (2) shows results when the event date is set to August 1, column (3) shows results when the event date is set to September 1, and column (4) shows results when the event date is set to October 1.

Table 1.5: Results of testing for anticipation effects.[1]

Variable	Dependent variable: Log price per m^2			
	(1)	(2)	(3)	(4)
Event date:	Jul. 1	Aug. 1	Sep. 1	Oct. 1
MRT $\leq$ 0.5km	-0.001	-0.002	-0.001	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)
MRT $\leq$ 0.5km $\times$ Post	-0.002	-0.003	-0.006	-0.012**
	(0.004)	(0.004)	(0.004)	(0.006)
Area (logarithm)	-0.152***	-0.152***	-0.152***	-0.152***
	(0.008)	(0.008)	(0.008)	(0.008)
Age	-0.003	-0.003	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Resale	-0.014	-0.014	-0.014	-0.014
	(0.013)	(0.013)	(0.013)	(0.010)
Sub sale	0.014	0.014	0.014	0.014
	(0.008)	(0.008)	(0.008)	(0.008)
HDB purchaser	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Percent of lease remaining	-0.147***	-0.147***	-0.147***	-0.147***
	(0.010)	(0.010)	(0.010)	(0.010)
Floor number indicator variables	Yes	Yes	Yes	Yes
Building fixed effects	Yes	Yes	Yes	Yes
Monthly indicator variables	Yes	Yes	Yes	Yes
Sample size	17,592	17,592	17,592	17,592
$\bar{R}^2$	0.972	0.972	0.972	0.972

[1] Standard errors are in parentheses. Significance is denoted by * at the 10% level, ** at the 5% level, and *** at the 1% level.

Because the only change made between each regression column is the falsified event date, the coefficients of area, age, resale, sub sale, HDB purchaser, and percent of lease remaining variables are identical in all four regressions. Area remains significant and negative, reflecting a 1.5 percent decrease in the per square meter price per 10 percent increase in area. Age, resale, subsale, and HDB purchaser status are all insignificant, but now there is a 15 percent *discount* shown for a percentage point increase in the amount of the lease remaining. Most importantly, none of the coefficients of the MRT and $POST$ interaction terms are positive and significant, with the coefficient even suggesting a 1.2 percentage point *decrease* in the premium when the falsified event date of the toll increase is set to October 1st, 2010.

1.5 Discussion

In this paper, a differences-in-differences (DD) framework is employed to test the effects of the November 1st, 2010 increase in ERP tolls on the transit premium paid for properties within a 0.5km radius from MRT, using transactions data in 2010 and 2011. Overall, the toll increase led to a significant, but modest, increase in the transit premium of about 0.6 percentage points. This is an expected result, as cars in Singapore are already highly taxed and regulated, and there is some evidence that those who can afford to drive private vehicles will likely have continued to do so because of a perceived sunk cost (Ho et al., 2018). Therefore, even a substantial increase in the cordon toll, like the one studied, would likely exhibit a much smaller impact on the transit premium than further increases in taxes or more restrictive quotas on new private vehicle registrations.

Because of the use of within-building variation to estimate the difference-in-differences model, it was not feasible to compute a transit premium level: rather, the MRT treatment coefficient identified the short-term price change in properties that had MRT stations built within 0.5 kilometers. The reason these coefficients generally turned out to be insignificant, or in some cases negative, is because two opposing forces are at play: the first is the property would become more desirable because it would then be connected by MRT while the second is that because the property is now being connected with transit, it would increase the supply of housing within 0.5 kilometers of MRT, thereby saturating that market. When the building fixed effects were removed, then the coefficient on the MRT treatment variable represented a pre-increase transit premium level, which was estimated to be 5.1 percent using a simple OLS model with controls.

Interestingly, the transit premium increase observed in this paper is driven by the transit premium increase that was observed among resale properties in the data. With an understanding of Singapore's housing market, this is also somewhat intuitive: higher income buyers are more able to afford new sales and sub sales compared to middle class buyers in the private market and hence the segmentation by resale, sub sale, or new sale in the dataset is likely to be correlated with the buyer's income. The main resale effect is not highly identifiable due to the use of within-building variation; the ages of units within a building exhibit little variation, and therefore, an older building is likely to contain exclusively resales, while a newer building will likely only register new sales or sub sales in the dataset and therefore, a main resale effect will be more apparent in across-building comparisons. However, the use of building fixed effects did not preclude identifying interaction effects to show how different segments of the private market reacted to the toll increase.

Furthermore, because of the arbitrary choice to use 0.5 kilometers as an MRT catchment area for which the transit premium would be capitalized, it was necessary to vary the MRT distance threshold to understand where the toll increase would be capitalized most in the transit premium. Admittedly, using Euclidean distance is a potentially contentious decision, but one that facilitates the empirical analysis of the present paper – although analyses that incorporate spatial models such as that used in Diao et al. (2017) do use network distances. Despite such controversy, the estimated change to the transit premium following the toll increase is observed when the MRT catchment area threshold is set to a Euclidean distance between 0.4 and 0.6 kilometers. This is roughly consistent with survey evidence of MRT users' willingness to walk to stations, which suggests an average walking distance of 608m (Olszewski and Wibowo, 2005).

Crucially, testing for announcement effects failed to produce any evidence of effects in the transit premium being capitalized before the toll increase. When using only pre-toll increase transactions, there were no differences found in the transit premium across properties sold after July 1st, August 1st, and September 1st. Furthermore, the transit premium turned out to *decline* when the falsified toll increase date was set to October 2010 – which clearly fails to prove announcement effects happening in the months leading up to the toll increase on November 1st.

While some aspects of the analysis in this paper seemed to be limited by the use of building fixed effects in the empirical model, these were necessary to avoid endogeneity issues with some unobserved building characteristics without compromising the ability to identify the changes to the transit premium following the toll increase, which was the main goal of this paper. Acknowledging the potential endogeneity that incorporating floor number could cause was the primary justification: buildings closer to the CBD would likely be taller, and necessarily make sales of units on higher floors possible, but proximity to the CBD may have also influenced the change to the transit premium effected by the toll increase. By controlling for building fixed effects, selling a unit on a higher floor would still be capitalized in the sale price, but not subject to comparison with units in buildings with fewer floors or farther away from the CBD. The use of building fixed effects also eliminated any heterogeneous effects to the transit premium observed at different distances from the CBD, and it can be argued that understanding these heterogeneous effects could enrich the analysis presented in this paper.

1.6 Conclusion

This work aims to measure the impact to property values with access to transit that can result from changes in congestion pricing. Intuitively, if the standard urban model suggests that commuters value lower transport costs in the price paid for housing, then any mechanism which increases auto commuting costs would shift housing location preferences closer to alternative modes of transport. This shift would imply an increase in the unit price of housing with improved access to other forms of transport. This paper exploits an increase in a cordon congestion toll to confirm an increase in the premium paid for housing close to transit following the increase.

However, external validity of this paper's main findings may be limited. First, the cordon toll model is a rarely implemented form of congestion pricing; Singapore is only joined by London, Stockholm, and Gothenburg, Sweden in implementing cordon pricing. While there are plans already underway to implement a cordon entering Lower Manhattan in New York City, it has not come without being politically controversial, with prominent New Jersey politicians voicing fierce opposition to the toll. Second, Singapore has a quite unique housing market, in which prices from a two-tiered system dominated by the public housing market can filter into the private market. Finally, Singapore is also unique in that it is much more monocentric than other major cities, having a single central business district, despite its numerous HDB towns with their own retail and commercial areas, health care, and recreational facilities. Still, this paper comes away with a modest effect of a toll increase on the transit premium, raising the transit premium by 0.6 percentage points, which represents an additional appreciation of roughly S$70 per square meter for comparable units within the same building.

Chapter 2

Time or Cost? An experimental analysis of travel behavior in LA Metro's GoPass program

2.1 Introduction

Following the COVID-19 pandemic, transit agencies introduced new programs that eliminated fares for students and youth. These programs come during a time in which younger cohorts are delaying getting a driver's license (Gibson, 2023) on top of exhibiting higher transit-dependence than other groups (Nash and Mitra, 2019). Throughout Southern California, numerous countywide transit agencies have created new programs targeting youth, such as Orange County's Youth Ride Free program, Riverside County's Go-Pass and U-Pass programs, and Los Angeles County's GoPass program, which is the focus of the present paper. Programs that subsidize or eliminate student and youth fares are not novel (Arranz et al., 2019; Lachapelle et al., 2022; Brown et al., 2003; Boyd et al., 2003), but the scale of these new policy interventions warrant further investigation for their role in influencing students' travel behavior, which changed considerably due to the COVID-19 pandemic (Haseeb and Mitra, 2024).

Travel behavior among post-secondary students is understudied in the travel behavior literature. This literature gap is not without merit; students are often undersampled in wide-scale travel surveys such as the National Household Travel Survey (NHTS) and are more prone to time constraints and survey fatigue than the general population. Due to these barriers in reaching students, there is growing concern that changes in travel behavior are due to nonresponse biases instead of actual behavioral responses (Bradley et al., 2018). Despite these challenges, studying student travel behavior is critical because colleges are major trip generators in urban areas, therefore contributing to congestion-related externalities in urban environments.

Recent literature has shown a flurry interest in the travel behavior of community college students thanks to several advances. First, the COVID-19 pandemic uniquely altered travel patterns and changed preferences due to isolation protocols, increased telecommuting, and increased car ownership (Aaditya and Rahul, 2023; Anwari et al., 2021). Second, advances in data collection and online recruitment methods mitigate the low response rates that are commonly encountered when contacting hard-to-reach groups (Tabasi et al., 2024; Lubitow et al., 2021). Finally, the StudentMoveTO project surveyed students across several universities and community colleges in the Greater Toronto Area, which has produced numerous travel behavior studies speficially examining community college students.

Despite these advances, scant studies address the effects that wide-scale free student transit programs cause on students' travel behavior. As a result, the lack of literature in this space creates an opening to ask a significant research question, which is whether fare-free student transit programs effect changes to travel behavior that support the costs from the forgone revenue. To address this research gap, this study focuses on students at community colleges in the Los Angeles Community College District (LACCD), which encompasses nine college campuses across Greater Los Angeles. I employ a discrete choice experiment (DCE) to recreate trips of varying lengths where students can pick between transit and other frequently used modes. I then use discrete-choice models to find that these students value savings in travel time at rates ranging from $27 to $54/hour. Students who participated in the GoPass program were more likely to select bus options but did not display such behavior for train options. A policy simulation suggests that a modest reduction in travel times on transit is broadly more effective than eliminating fares at improving transit mode choice and consumer surplus. A recreation of the GoPass program increased consumer welfare between 19 and 60 cents per trip.

This paper is structured as follows: Section 2.2 provides the institutional background of the GoPass program and LACCD. Section 2.3 presents the survey instrument and the design of the discrete choice experiment used in this study. Section 2.4 presents the discrete choice models used to model the decisions that respondents made in the DCE tasks. Section 2.5 presents the sample statistics from the survey, along with regression results from the discrete choice models. Section 2.6 reviews the results of the policy analyses that follow from the models presented in this study. Section 2.7 discusses the implications of this paper's findings, and Section 2.8 is the conclusion.

2.2 Institutional Background

2.2.1 GoPass Program

In September 2020, the Los Angeles County Metropolitan Transportation Authority (or LA Metro) launched the Fareless Systems Initiative, which proposed to study potential impacts of eliminating fares on Metro buses and trains. Initially, the initiative was meant to be incorporated in phases, with Phase 1 being a pilot program that eliminated fares for K–12 and community college students (Los Angeles Metropolitan County Transportation Authority, nd). On September 23, 2021; the LA Metro Board approved a funding plan for a two-year pilot program aimed at K–12 and community college students, which later became known as GoPass. GoPass was introduced on October 1, 2021, and then renewed for an additional year in early 2023 near the end of its initial two-year pilot. GoPass then became a permanent program in April 2024.

The GoPass program offers free fares to eligible students onboard most Metro and municipal transit services, except some services like Metro Micro.[1] Because Los Angeles County is served through many school districts and independent transit operators[2], the GoPass program is the result of an agreement between multiple school districts and transit agencies throughout the county. To be considered eligible for GoPass, students at any eligible community college must be actively enrolled in classes for the current academic term. Eligible students may opt into GoPass by requesting a special reloadable fare card (typically referred to as a TAP card) from their college, which they activate with LA Metro for use on transit.

What is notable about the GoPass program is its scale. As of March 2024, GoPass covers 17 transit agencies and 113 K-12 public school districts, charter school networks, and independent charter and private schools. Over 400,000 unique GoPass participants have benefited from the program and GoPass boardings have grown consistently every year, from at 5.4 million boardings in 2021–22, to 17.8 million in 2022–23, and then to 19 million in 2023–24 (Los Angeles County Metropolitan Transportation Authority, 2024a).

[1]Metro Micro is a ridesharing microtransit service that operates within ten designated zones in Los Angeles County, and charges a base fare of $2.50 per ride.

[2]LA Metro is far and away the largest transit operator in L. A. County, but different regions of the county are known to have transit services operated by municipal agencies, such as Long Beach Transit or the Santa Monica Big Blue Bus.

2.2.2 Los Angeles Community College District

The Los Angeles Community College District (LACCD) is a district that contains nine community colleges serving Los Angeles and 40 incorporated cities in Los Angeles. It is the largest community college district in the United States with 194,000 students enrolled in courses during the 2023–24 academic year. LACCD colleges awarded 35,000 degrees and distributed $261 million in financial aid awards to students in the 2022–23 academic year.

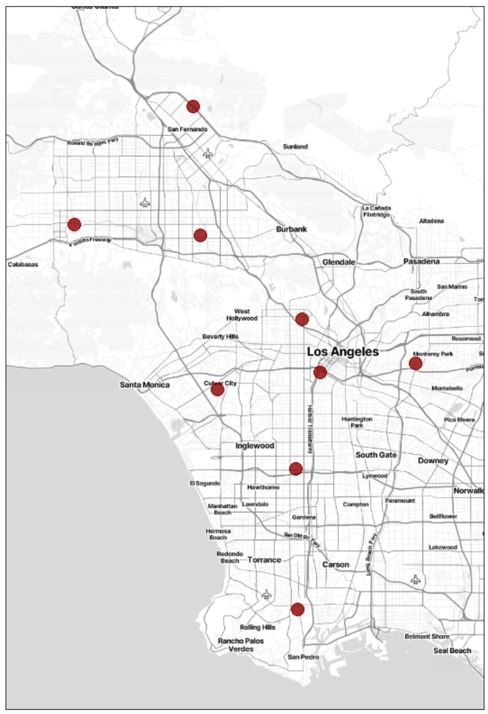


Figure 2.1: Map of LACCD campuses within the Los Angeles area.

In its demographics, LACCD's student body is diverse, low-income, and faces high housing insecurity relative to the general population in Los Angeles County. 45 percent of LACCD students are first-generation students and 48 percent of students later transfer to a four-year institution, most of whom transfer to a California State University or University of California campus. 57 percent of the student body is Hispanic, 17 percent is White, and 9 percent is Black. The LACCD student body also skews female, which accounts for 58 percent of all students. The LACCD student body is also older than the typical four-year university age demographic; only 27 percent of students are under the age of 20, 26 percent are between 20 and 24, 25 percent are between 25 and 34, and 22 percent are over 35. 55 percent of students are housing insecure and 63 percent of students are reported to be food insecure. Most students attending LACCD colleges either part-time students or enrolled in non-credit courses, making up 73 percent and 8 percent of the student body, respectively. Figure 2.1 shows the map of LACCD college campuses across the Los Angeles area (Los Angeles Community College District, nd).

2.3 Data Collection

2.3.1 Survey Instrument

LACCD students were recruited for an online survey in April and May 2024. Participants were recruited via email invitations that provided personalized links produced from an email list of 115,000 actively enrolled students for the Spring 2024 term. 2655 responses were received, and all participants were entered in a lottery for a $50 Amazon gift card, out of which roughly 1 in 15 participants were selected as winners. Eligibility for the lottery did not require participants to answer all questions and selection for a prize was not awarded based on survey completion. At the end of the survey, participants were allowed to complete an additional task which entered them into an additional lottery of up to $19.

The survey instrument contained several parts, which most students completed in roughly 15 minutes. The information collected from each part of the survey is listed below:

1. **Screening Questions:** Students were asked about basic household and enrollment information, their familiarity with GoPass, and whether they participated in GoPass. They were also asked questions about household vehicle use, including questions on

whether they had a driver's license, the number of vehicles in their household, and whether they depended on relatives to meet their transportation needs.

2. **Trip-Related Questions:** Students were asked about their viable options for trips to and from their college campus.[3] They were then asked to recall which mode they relied on to reach their college campus and to provide details from their most recent trip to and from their campus. They were asked questions regarding their employment status and work trips (when applicable), followed by questions about their mode choice for leisure trips.

3. **Discrete Choice Experiment:** Here, students were presented with three hypothetical trip scenarios: one to a destination 12 miles away, one to a destination 7 miles away, and then one to a destination 2 miles away. For each trip, they were presented with three alternatives and were told the mode, travel time, waiting time, cost, and whether Wi-Fi would be available for their trip. The modes for these hypothetical trips were bus, train, car, self-driving car, e-bike or scooter rental, or Metro Micro. At the end of the short trip scenario, they were offered an "opt-out" option, which asked whether they would prefer to walk over their preferred alternative. Students in this experiment were assigned to two groups: one in which students picked their preferred alternative in each choice scenario and another where students ranked each alternative from best to worst.

4. **Attitudinal Questions:** These questions first focused on generic concerns about affordability, safety, impacts of climate change, and access to job opportunities. They then shifted to probe respondents on their concerns specific to transit, namely safety, reliability, navigability, and miscellaneous concerns. Students who used GoPass were then asked Likert-scale questions indicating how GoPass impacted their lives, while

[3] Some students are enrolled in courses at multiple campuses, but are assigned to a "home" campus. For these questions, we asked students to focus on their home campus.

students who did not use GoPass were asked Likert-scale questions focused on concerns that detracted them from participating.

5. **Demographic Questions:** Students were asked demographic questions. All students were asked about their age, type of housing, living arrangement, household income, gender, Hispanic origin, and whether they receive financial aid or public assistance.

6. **Delay Discounting Task:** At the end of the survey, students were asked about their time preferences, where they had to choose between a $10 immediate payoff and a delayed payoff, varying between $11 and $19. The delayed payoff started at $12 and moved in increments of $2 until the delayed payoff was selected, after which a follow-up question would be asked and the students would be presented with the delayed payoff reduced by $1 to identify where the students were indifferent between the immediate and delayed payoff.

The analysis presented in this study will focus on the discrete choice experiment, making use of the responses from participants who were randomly assigned to complete the traditional DCE task.

2.3.2 Discrete Choice Experiment

Since a key goal of the present paper is to assess the demand for different modes of travel while incorporating some novel modes, the the best approach is to specify a discrete choice experiment (DCE) (Louviere and Hensher, 1982; Louviere and Woodworth, 1983). Respondents were shown three different scenarios, each representing a hypothetical trip with varying distance, starting with a 12-mile trip, then a 7-mile trip, and finally, a 2-mile trip. The attributes shown to respondents were: mode of travel, total travel time, travel cost, wait time, and whether Wi-Fi was available. Attribute values for the SP tasks were set using the NGENE software package (ChoiceMetrics, 2024). A D-efficient design for the multinomial logit model was used, and 72 tasks were created, divided into 24 blocks to create the DCE. Blocks were then transcribed from NGENE into a survey software, Qualtrics, using an approach consistent Weber (2021).

The chosen modes included in the design were available in the Los Angeles area, but included some that were not widely available at the time, such as Metro Micro and a self-driving car service. Metro Micro is a shared ride service operated by the Los Angeles Metro transportation agency and is designed primarily for short neighborhood trips. Google's Waymo "robo-taxi" service began serving Los Angeles just before the survey was fielded, but its availability is currently limited. For the short trip scenario, walking is included as a fourth "opt-out" alternative. The travel time for the walk alternative is assumed to be 40 minutes, which is the time it would take a person to walk 2 miles at a 20-minute per-mile pace.

In the back-end, modes were coded, as were the attribute values from a set of pre-specified values, which are shown in Table 2.1. The modes (coded 0–4) were car, train, self-driving car, bus, and e-bicycle rental in the first two scenarios, while in the final scenario, the possible modes respondents could be shown were car, Metro Micro, self-driving car, bus, and e-bicycle rental. Figure 2.2 shows a sample screenshot containing a DCE task that was shown to a respondent. For the prompts used in the DCE and the full design, see Appendices A.1.1 and A.1.1.

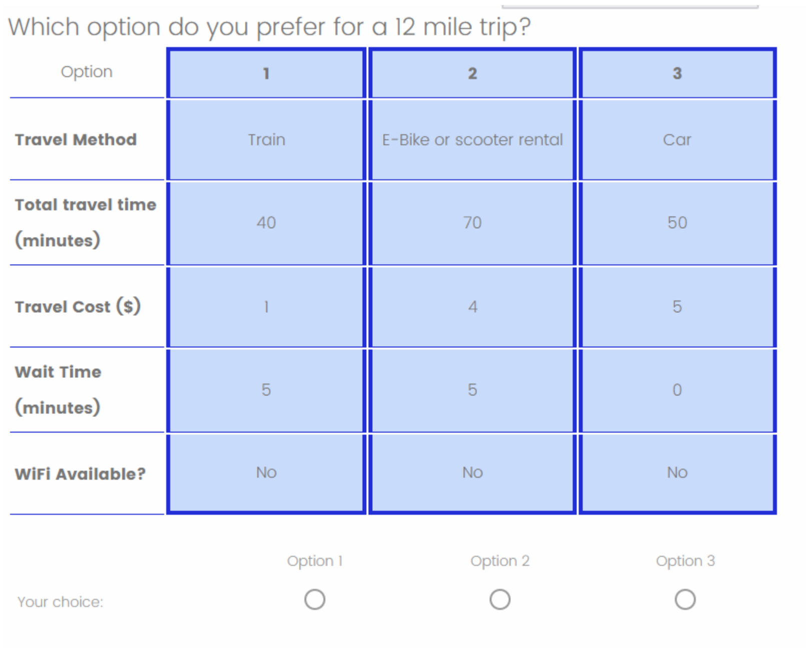
Table 2.1: Possible values for trip attributes across all stated preference scenarios. [1]

Attribute	Option 1	Option 2	Option 3
Scenario 1 (12 mile trip)			
Mode of travel	Train Bus	Car Self-driving car	E-bicycle/ Scooter rental
Total travel time (minutes)	40, 55, 70	30, 40, 50	60, 70, 80
Travel cost ($)	Free, 1, 2	5, 10, 15	3, 4, 5
Waiting time (minutes)	5, 10, 15	0 (car) 5, 10, 15 (self-driving car)	0, 5, 10
Wi-Fi available	Yes, no	No (car) Yes, no (self-driving car)	No
Scenario 2 (7 mile trip)			
Mode of travel	Train Bus	Car Self-driving car	E-bicycle/ scooter rental
Total travel time (minutes)	30, 40, 50	25, 35, 45	30, 40, 50
Travel cost ($)	Free, 1, 2	4, 7, 10	3, 4, 5
Waiting time (minutes)	5, 10, 15	0 (car) 5, 10, 15 (self-driving car)	0, 5, 10
Wi-Fi available	Yes, no	No (car) Yes, no (self-driving car)	No
Scenario 3 (2 mile trip)			
Mode of travel	Metro Micro	Car	E-bicycle/

	Bus	Self-driving car	scooter rental
Total travel time (minutes)	10, 15, 25	7, 12, 17	10, 20, 30
Travel cost ($)	Free, 1, 2	4, 7, 10	2, 3, 4
		0 (car)	
Waiting time (minutes)	5, 10, 15	5, 10, 15	0, 5, 10
		(self-driving car)	
		No (car)	
Wi-Fi available	Yes, no	Yes, no	No
		(self-driving car)	

[1] Travel modes were a design attribute, and therefore, were not constrained to be shown in any specific order. The ordering of modes shown is for demonstration purposes only.

The DCE was formatted in such a way so that each task could be viewed on a smartphone, for ease of completion. One key challenge of DCE design is to balance between task complexity and the amount of information provided for model estimates to be informative. Because the mode of travel was incorporated into the design, modes were not constrained to a specific option order in each DCE (such as the first or second option). Moreover, the DCE had an efficient design to minimize D-error, which prevented dominant (or dominated) alternatives from being present in the choice set.


Which option do you prefer for a 12 mile trip?

Option	1	2	3
Travel Method	Train	E-Bike or scooter rental	Car
Total travel time (minutes)	40	70	50
Travel Cost ($)	1	4	5
Wait Time (minutes)	5	5	0
WiFi Available?	No	No	No

	Option 1	Option 2	Option 3
Your choice:	○	○	○

Figure 2.2: Screenshot of a stated preference task shown to respondents.

2.4 Choice Modeling Framework

2.4.1 Multinomial Logit (MNL) Model

In the discrete choice preference experiment, participants within a scenario k faces a choice set C_n that is drawn from a complete set of J alternatives. The choice model framework is developed from the canonical random utility model where participant n chooses alternative $i \in \{1, ..., J\}$. Utility of individual n in scenario k is expressed by the following:

$$U_{nik} = V_{nik} + \varepsilon_{nik}. \tag{2.1}$$

Equation (2.1) decomposes utility into a deterministic component, V_{nik}, and a stochastic component, ε_{nik}. This deterministic component is defined by $V_{nik} = \beta' X_{nik} + \delta_i' Z_n$, where that X_{nik} is a vector of attribute values (namely travel time, waiting time, travel cost, and

Wi-Fi availability), Z_n is a set of variables that characterize individual n, and the parameters β and δ_i refer to the coefficient vector for attribute values and the coefficient vector for alternative i on individual variables, respectively.

Each individual n in scenario k faces a choice set, $C_{nk} \subset \Omega_k$, such that C_{nk} is selected from a space of possible modes Ω_k, which is different in each scenario k. For ease of estimation, there are two identifying assumptions for this model: first, the error terms, ε_{nik}, of the utility function are independently and individually distributed across individuals and scenarios and follow a type I extreme value distribution, which is commonly known as the independence of irrelevant alternatives (IIA) assumption. Second, as shown in Train (2009) and McFadden (1978), any choice set has an equal probability of being selected conditional on any of its members being selected into the choice set. This condition provides the conditional probability of choosing an alternative i on the selection of choice set C_{nk}, which is represented in Equation (2.2).

$$P_{nk}(i|C_{nk}) = \frac{\exp\left(V_{nik}\right)}{\sum_{j\in C_{nk}} \exp\left(V_{njk}\right)} \tag{2.2}$$

Now that the choice probability is defined, the binary variable is defined as $y_{nik} = 1(U_{nik} > \max\{U_{njk}\})\forall j \neq i \in C_{nk}$, where $1(\cdot)$ is the indicator function. Namely, y_{nik} is equal to 1 if individual n chooses alternative i in scenario k, and 0 otherwise, since the utility derived from choosing alternative i will dominate those from choosing any other alternative under this paradigm. Individual n's contribution from her choice in scenario k to the conditional log-likelihood function is then the product of y_{nik} and the choice probability from Equation 2.6, summed across all alternatives $i \in C_{nk}$. After substituting in the deterministic utility $V_{nik} = \beta' X_{nik} + \delta_i' Z_n$, the conditional log-likelihood function to estimate parameters β and δ_i can now be expressed as

$$CLL = \sum_{n} \sum_{k} \sum_{i \in C_{nk}} y_{nik} \cdot \ln \frac{\exp(\beta' X_{nik} + \delta_i' Z_n)}{\sum_{j \in C_{nk}} \exp(\beta' X_{njk} + \delta_j' Z_n)}. \tag{2.3}$$

The conditional log-likelihood in (2.3) is then maximized to estimate the model parameters, which gives the multinomial logit model as initially introduced by McFadden (1972), allowing for joint modeling across the trip scenarios in the DCE.

2.4.2 Mixed Multinomial Logit (MMNL) Model

The mixed multinomial logit (MMNL) model is a generalization of the MNL model, where the IIA assumption is relaxed and β follows a mixing distribution $f(\beta)$. This modification is commonly introduced to reflect heterogeneity in tastes (Revelt and Train, 1998), to reflect more realistic substitution patterns between alternatives (Brownstone and Train, 1998), or to jointly model revealed-preference and stated-preference choices (Brownstone et al., 2000). It is necessary to specify parameters θ that describe the density of the mixing distribution, $f(\beta)$, so that the mixing distribution is denoted by the conditional distribution $f(\beta|\theta)$. For some variables, the mixing distribution is degenerate and some parameters are fixed, so the random utility model from Equation (2.1) is modified to reflect the following change:

$$U_{nik} = \beta_n' X_{nik} + \gamma' W_{nik} + \delta_i' Z_n + \varepsilon_{nik}. \tag{2.4}$$

X_{nik} now reflects the vector of variables for which coefficients β_n are random, travel times interacted with the trip scenario, and they follow the mixing distribution $f(\beta|\theta)$. γ is a vector of fixed parameters and W_{nik} is the corresponding vector of attribute values whose coefficients will not be random. δ_i remains a vector of alternative-specific coefficients for individual variables Z_n. The most commonly used distribution of β_n is a normal (or lognormal) distribution (Hensher and Greene, 2003), where $\beta_n \sim N(\bar{\beta}, \Sigma)$ and $\bar{\beta}$ and Σ are parameters that must be estimated, representing the means and standard deviations of β, respectively.

This modification of allowing β_n to represent random parameters also modifies the choice probability, because β_n is not known directly. However, conditional on β_n, the choice probability is known. Therefore, it is first necessary to express this conditional choice probability:

$$L_{nik}(\beta_n) = \frac{\exp\left(\beta_n' X_{nik} + \gamma' W_{nik} + \delta_i' Z_n\right)}{\sum_{j \in C_{nk}} (\beta_n' X_{njk} + \gamma' W_{njk} + \delta_j' Z_n)}. \tag{2.5}$$

To obtain the choice probability P_{nik}, it is necessary to integrate P_{nik} over all possible values β_n following the mixing distribution of β. Using the conditional probability from Equation (2.5), the choice probability can now be expressed as

$$P_{nik} = \int \frac{\exp\left(\beta_n' X_{nik} + \gamma' W_{nik} + \delta_i' Z_n\right)}{\sum_{j \in C_{nk}} (\beta_n' X_{njk} + \gamma' W_{njk} + \delta_j' Z_n)} \cdot f(\beta|\theta) \mathrm{d}\beta. \tag{2.6}$$

The integral from Equation (2.6) does not lead to a closed-form log-likelihood function that can be found analytically. The choice probabilities must simulated, using drawn values of β_n from the mixing distribution. Each vector of drawn β_n values is expressed by β^r, where r represents the draw number $r \in \{1, ..., R\}$, and R is the total number of draws. The mean choice probability is computed across all β^r. As with the MNL model, a dummy y_{nik} is defined such that $y_{nik} = 1$ if individual n chooses alternative i in scenario k and 0 otherwise. The simulated choice probability $\hat{P}_{nik}$ is given by

$$\hat{P}_{nik} = \frac{1}{R} \sum_{r=1}^{R} L_{nik}(\beta^r). \tag{2.7}$$

Now that the simulated choice probabilities are computed, the mixed multinomial logit is estimated by maximizing the simulated log-likelihood, which is given by

$$SLL = \sum_{n} \sum_{k} \sum_{i \in C_{nk}} y_{nik} \cdot \ln \frac{1}{R} \sum_{r=1}^{R} L_{nik}(\beta^r). \tag{2.8}$$

Equation (2.8) is similar to that from the conditional log-likelihood in the MNL model, with the exception that the log-likelihood is no longer derived from exact choice probabilities.[4]

2.4.3 Value of Travel Time Savings (VTTS)

Using these choice models, it is straightforward to compute the amounts that subjects would need to be compensated for (a) increases in travel time, (b) increases in waiting time, (c) removal of Wi-Fi, and (d) switching from preferred modes to less preferred modes. The primary valuation of interest, however, is the value of travel time savings (VTTS).

Valuations of tradeoffs can be identified when utilities are kept constant, which is equivalent to setting marginal utility equal to zero. Individuals make time and cost tradeoffs, such that marginal utility of their travel time is given by $MU_{TT} = \partial U_{nik}/\partial TT_{nik}$ and the marginal utility of their travel cost by $MU_{TC} = \partial U_{nik}/\partial TC_{nik}$. In equilibrium, the relative "prices" of travel time and travel cost would therefore need to be equal to the marginal rate of substitution between the two, which is given by $MRS_{TT,TC} = MU_{TT}/MU_{TC}$. Thanks to the linearity of the random utility model with respect to travel time and travel cost, VTTS is computed by taking the ratio of the travel time and travel cost coefficients. However, in the MMNL model, values of β_n are not fixed, which requires the use of the expected value of β_n for travel time.

[4]Another potential issue with MMNL models is that the simulated log-likelihood function is not guaranteed to be globally concave, which is a feature of MNL models. This also extends to other discrete choice models, such as nested logit (NL).

$$VTTS = \frac{\partial U_{nik}/\partial TT_{nik}}{\partial U_{nik}/\partial TC_{nik}} = \frac{E[\beta_{n,TT}]}{\beta_{TC}} \tag{2.9}$$

In the case of the MNL model, the distribution of β_n is degenerate, and only the coefficient β_{TT} enters Equation (2.9).

This measure is consistent with Bates (1987) and Truong and Hensher (1985). Because of how VTTS is constructed in (2.9), it is expected to remain stable over travel time and cost since utility is linear in β_{TT} and β_{TT} is fixed in expectation. However, this is unlikely to be empirically true, as VTTS is often found to vary based on distance and trip purpose (Axhausen et al., 2008).

2.5 Results

2.5.1 Sample Summary

From the sample, roughly 1,000 students who were asked to complete the standard discrete choice experiment provided complete information on their travel behaviors and demographics[5]. Table 2.2 presents descriptive summary statistics from the sample that participated in the DCE, which were weighted by age distribution, gender, student enrollment status, and Hispanic origin to better represent the LACCD student body.

The first section of Table 2.2 summarizes the students' familiarity with GoPass and whether they participated in the program, alongside household driving and vehicle use information. Roughly 40 percent were unaware that their college offered the GoPass program, while 26

[5]Another 1,000 students completed a modified DCE with ranked preferences and provided demographic and trip information, but are excluded from this analysis. I acknowledge that joint modeling of the two DCE subgroups is an exercise of interest

percent knew GoPass was offered but did not use GoPass, and the remaining 34 percent of students used GoPass. A majority, 60 percent, of the students in the sample had a driver's license and the mean number of household vehicles was 2.1.

21 percent of the students were full-time students and the rest were either part-time or enrolled in non-credit courses. Students were also asked about their employment status: 20 percent of students reported they were employed full-time, 35 percent reported being employed part-time, 15 percent had no other occupation beyond their full-time student status, and 31 percent reported they were unemployed.

Table 2.2: Weighted summary statistics of sample demographics, GoPass use, and vehicle information.

Variable	N = 1,054
GoPass Awareness and Use	
Unaware of GoPass	420 (39.8%)
Aware, but does not have a GoPass	274 (26.0%)
Registered and used GoPass	360 (34.2%)
Have a Driver's License	629 (59.7%)
Number of Household Vehicles	2.11 (1.21)
Enrollment Status	
Enrolled full-time	219 (20.8%)
Enrolled part-time or not for credit	835 (79.2%)
Employment Status	
Employed full-time	209 (19.8%)
Employed part-time	365 (34.7%)
Full time student	150 (14.3%)
Unemployed	328 (31.2%)
Household Income	

Less than \$15,000	215 (20.5%)
\$15,000 to \$44,999	282 (26.9%)
\$45,000 to \$74,999	171 (16.3%)
\$75,000 or more	112 (10.7%)
Prefer not to answer	267 (25.5%)
Living Arrangement	
Owner	73 (6.92%)
Renter	346 (32.9%)
Dependent, living with parents or relatives	325 (31.0%)
Independent, in shared housing	236 (22.4%)
Temporary housing or alternative housing	71 (6.72%)
Household Size	3.99 (2.77)
Receiving Financial Aid	634 (61.0%)
Receiving Government Assistance	556 (53.5%)
Age	
18 or 19	261 (24.7%)
20-24	309 (29.3%)
25-34	268 (25.5%)
35 or older	216 (20.5%)
Gender	
Male	401 (38.1%)
Female	636 (60.3%)
Non-binary	9 (0.82%)
Prefer not to answer	8 (0.76%)
Hispanic Origin	622 (59.0%)

The sample was low-income and faced high housing insecurity. 21 percent reported having an annual household income below $15,000, 27 percent reported a household income between $15,000 and $45,000, 16 percent reported a household income between $45,000 and $75,000, while 11 percent reported a household income above $75,000. About 25 percent of the sample did not know or declined to report their household income. Moreover, few (6 percent) are homeowners, while the great majority were either renters (33 percent), dependents (31 percent), residents in shared housing arrangements (22 percent), or residents in temporary or alternative housing (7 percent). They lived with many people in their household; the mean household size was roughly 4 people per household. 61 percent received financial aid to attend classes, either through Pell grants, student loans, or fee waivers; and 54 percent received some form of public assistance.

25 percent of the students were 18 or 19 years old [6], 29 percent were between ages 20 and 24, 26 percent were between ages 25 and 34, and the remaining 21 percent were over the age of 35. 38 percent of this sample was male, and 60 percent of this sample was female, while the remainder was non-binary or had an unknown gender. 59 percent of the sample was of Hispanic origin.

2.5.2 Discrete Choice Experiment Summary

In the DCE, three scenarios were shown to each participant: one asking about a twelve-mile trip, one asking about a seven-mile trip, and a third scenario asking about a two-mile trip. At the end of the two-mile trip scenario, respondents were asked if they preferred to walk over using their chosen alternative, which created a partial ranking of alternatives in the data. Table 2.3 shows the distribution of offered and chosen modes shown in experiment, and also shows the distributions disaggregated by trip scenario.

[6] Some minors are allowed to enroll in LACCD courses, but they were excluded from the sample.

Overall, the self-driving car service alternative made up 24 percent of all offered alternatives across the scenarios, while making up 27 percent of all chosen alternatives. Buses comprised 23 percent of the displayed alternatives and comprised 25 percent of the chosen alternatives. E-bicycles and scooter rentals were rarely selected, representing 16 percent of the displayed alternatives but only 4 percent of selected alternatives. Metro Micro had a 9 percent share of offered alternatives, while it only had a 4 percent share of selected alternatives. Trains made up roughly equal shares of offered and chosen alternatives, comprising roughly 15 percent of both offered and chosen alternatives. Finally, cars made up 5 percent of offered alternatives, but they made up 9 percent of all the chosen alternatives. The walk alternative was chosen by 486 participants of the 988 times it was displayed after the short scenario.

A disaggregation of the offered and chosen alternatives by scenario suggests that buses were selected most often in the medium-length trip scenario, where buses were 26 percent of the displayed alternatives but 35 percent of the chosen alternatives. Cars were primarily selected for the long trip scenario, where they were 10 percent of the alternatives but 20 percent of the selected alternatives. Self-driving cars were disproportionately chosen in the long and medium-length trip scenarios. The e-bicycle/scooter rental option was selected much less often than offered in every scenario, where it made up a much smaller share of the chosen alternatives than it did of the offered alternatives. The relative selection rate of trains held constant across the long and medium-length trip scenarios, comprising similar shares of the offered and chosen alternatives.

Table 2.3: Mode choice counts and percentages by scenario and overall.

Mode	Total (All Scenarios)		Scenario 1		Scenario 2		Scenario 3	
	Total (%)	Chosen (%)	Total (%)	Chosen (%)	Total (%)	Chosen (%)	Total (%)	Chosen (%)
Bus	2,235 (22.6%)	765 (25.9%)	705 (23.8%)	239 (24.2%)	760 (25.6%)	340 (35.0%)	770 (19.5%)	186 (19.0%)
Car	454 (4.6%)	204 (9.0%)	289 (9.8%)	206 (20.6%)	83 (2.8%)	41 (4.2%)	82 (2.1%)	22 (2.2%)
E-Bike or scooter	1,532 (15.5%)	107 (3.6%)	510 (17.2%)	18 (1.8%)	603 (20.4%)	58 (5.9%)	388 (10.6%)	23 (3.2%)
Metro Micro	877 (8.9%)	105 (3.6%)	—	—	—	—	877 (22.1%)	89 (10.7%)
Self-driving Car	2,314 (23.4%)	779 (26.4%)	674 (22.7%)	289 (29.3%)	824 (27.8%)	340 (34.6%)	816 (20.7%)	150 (15.3%)
Train	1,480 (15.0%)	442 (15.0%)	786 (26.5%)	239 (24.2%)	694 (23.4%)	203 (20.7%)	—	—
Walk	988 (10.0%)	379 (16.5%)	—	—	—	—	980 (25.0%)	486 (49.6%)
Total	9,880	2,287	2,964	988	2,964	982	3,952	980

Table 2.4: Summary statistics of attribute values among chosen alternatives

Scenario	**Variable**	**Obs**	**Mean**	**S. D.**	**Min**	**Max**
Long trip	Travel cost ($)	988	5.14	5.01	0	15
	Travel time (minutes)	988	46.51	12.73	30	70
	Waiting time (minutes)	988	7.48	5.55	0	15
	WiFi availability	988	0.42	0.49	0	1
Medium trip	Travel cost ($)	982	3.33	3.05	0	10
	Travel time (minutes)	982	36.55	9.08	25	50
	Waiting time (minutes)	982	9.08	4.44	0	15
	WiFi availability	982	0.47	0.50	0	1
Short trip (excluding walking)	Travel cost ($)	978	3.05	3.45	0	10
	Travel time (minutes)	978	15.19	7.13	7	30
	Waiting time (minutes)	978	6.33	3.65	0	15
	WiFi availability	978	0.51	0.50	0	1

Finally, Table 2.4 summarizes the trip attributes of the chosen modes from each scenario. For the long trip, respondents chose modes with a mean cost of $5, a mean travel time of 46 minutes, and a mean waiting time of about 7.5 minutes. 42 percent of selected modes in this scenario offered WiFi connectivity. In the medium-length trip scenario, the mean travel cost of selected alternatives was $3.33, while mean travel time and waiting time were 37 minutes and 9 minutes, respecitvely. 51 percent of selected modes offered WiFi connectivity. In the short trip, mean travel cost was close to $3, mean travel time was 15 minutes, and mean waiting time was 6 minutes. 51 percent of chosen alternatives offered WiFi connectivity in the two-mile trip scenario.

2.5.3 Estimation Results

Regression Results

Table 2.5 displays the regression results for the DCE using the MNL and MMNL specifications presented in Section 4. For the walk alternative at the end of the last trip scenario, travel time is assumed to be 40 minutes with no associated travel cost or waiting time. For ease of identification, these models incorporate a small number of individual-specific variables and the alternative-specific coefficients are relative to a car base alternative. In addition, travel time coefficients were specified in the model as three separate parameters, an approach consistent with Standen et al. (2019) to allow for travel time utility weights to be heterogeneous across the trip scenarios. For the MMNL, the reported travel time coefficients are the means of the random parameters, and the standard deviations on the travel time parameters were constrained to be equal across all scenarios.

In the MNL model, travel time has a negative and significant coefficient for all trip length scenarios, which is an expected result. The coefficient on waiting time is also negative and significant, suggesting that the respondents penalized waiting time more than travel time[7]. The coefficient on travel cost is predictably and significantly negative. Finally, the coefficient of WiFi connectivity is positive and significant, which shows respondents received positive marginal utility from choosing modes with WiFi.

The alternative-specific constants are negative and statistically significant for all non-car modes, except walking. This means that car alternatives dominated most alternatives using other modes, even after holding all trip attributes constant, which is not a surprising result. However, in the shorter trip scenarios, buses and e-bicycle rentals were less dominated by cars, since the scenario-specific constants were positive and significant. The self-driving car service is far less dominated by cars in the short-trip scenario as the scenario-specific constant is also positive and significant. Having a GoPass increases the probability of choosing buses relative to cars when holding all trip attributes constant, but does not have that same effect for trains. This result may suggest that GoPass holders are more familiar or comfortable with buses than trains. Moreover, having a GoPass also decreased the probability of selecting the self-driving car service relative to cars. Having a driver's license decreased the probability of selecting buses and trains relative to cars, indicating that those with driver's licenses exhibited a stronger dislike of transit options when holding all attributes constant.

[7] In the experiment displayed to respondents, travel time was indicated by "Total travel time", which leads to the waiting time coefficient identifying a penalty of travel time over waiting time.

Table 2.5: Regression results for the multinomial logit and mixed multinomial logit models

Variable	MNL		MMNL	
	β	SE	β	SE
Travel Attributes				
Travel time (long trip)	-0.018	(0.004)	-0.024	(0.005)
Travel time (medium trip)	-0.034	(0.005)	-0.037	(0.005)
Travel time (short trip)	-0.021	(0.006)	-0.017	(0.006)
Travel cost	-0.040	(0.012)	-0.041	(0.013)
Wait time	-0.012	(0.006)	-0.013	(0.006)
WiFi availability	0.255	(0.046)	0.271	(0.048)
$\sigma_{\text{TravelTime}}$	—	—	0.035	(0.009)
Individual Variables and ASCs (Base = Car)				
Bus				
Alternative-specific constant	-1.301	(0.296)	-1.359	(0.270)
Medium trip	0.562	(0.265)	0.607	(0.296)
Short trip	1.009	(0.316)	0.971	(0.331)
Has GoPass	0.509	(0.245)	0.537	(0.236)
Driver's license	-1.023	(0.269)	-1.088	(0.233)
E-bike or scooter rental				
Alternative-specific constant	-3.621	(0.369)	-3.716	(0.378)
Medium trip	1.212	(0.391)	1.275	(0.406)
Short trip	2.287	(0.435)	2.289	(0.442)
Has GoPass	-0.190	(0.296)	-0.182	(0.296)
Driver's license	-0.278	(0.287)	-0.323	(0.288)
Metro Micro				
Alternative-specific constant	-1.059	(0.236)	-1.058	(0.425)

Short trip	-0.125	(0.461)	-0.247	(0.457)
Has GoPass	0.157	(0.296)	0.169	(0.277)
Driver's license	-0.246	(0.282)	-0.274	(0.272)
Self-driving Car				
Alternative-specific constant	-1.140	(0.216)	-1.190	(0.425)
Medium trip	-0.139	(0.266)	-0.247	(0.295)
Short trip	0.658	(0.293)	0.632	(0.318)
Has GoPass	-0.522	(0.236)	-0.528	(0.235)
Driver's license	0.202	(0.223)	0.212	(0.227)
Train				
Alternative-specific constant	-1.428	(0.274)	-1.487	(0.236)
Medium trip	0.109	(0.275)	0.128	(0.298)
Has GoPass	0.347	(0.258)	0.370	(0.243)
Driver's license	-0.799	(0.251)	-0.861	(0.240)
Walk				
Alternative-specific constant	-0.039	(0.354)	-0.137	(0.372)
Has GoPass	-0.194	(0.270)	-0.214	(0.276)
Driver's license	-0.074	(0.259)	-0.084	(0.269)
Number of individuals (N):	988		988	
Number of choice situations:	3,434		3,434	
Number of total observations:	11,282		11,282	
Log pseudolikelihood (Simulated for MMNL):	-3297.365		-3294.535	

The MMNL model estimates paint a similar picture as those of the MNL model, with slight differences in the magnitudes of the attribute coefficients. The fixed parameters on waiting time, WiFi availability, and travel cost all have the same signs and statistical significance shown in the MNL estimates. What is notable is that the standard deviation of travel time parameters is statistically significant and suggests that the effects of travel times on utilities is heterogeneous across the sample. This result also indicates that specifying mixing of the travel time parameters is appropriate[8].

Valuations

From the estimation results, the relative valuations of trip attributes were computed. Because travel time coefficients were allowed to be estimated for each trip scenario, it is possible to detect heterogeneity in the value of travel time savings across trip scenarios. These results are reported in Table 2.6.

Table 2.6: Valuations of trip attributes (in 2024 $) of travel time and WiFi availability.

Attribute	**MNL**	**MMNL**
Travel time savings (long trip, per hour)	27.27	35.46
Travel time savings (medium trip, per hour)	51.59	53.92
Travel time savings (short trip, per hour)	30.34	25.48
Marginal waiting time savings (per hour)	18.00	19.21
WiFi availability	6.37	6.65

In the first column, valuations are shown from the estimates produced by the MNL model. The value of travel time starts at $27/hour in the long trip scenario, increasing to $52/hour in the medium-length trip scenario, and then decreasing to $30/hour in the short trip scenario. Notably, the difference in VTTS between the long and medium trip scenarios is over $24/hour, which is considerable. In the MMNL model, respondents were found to display a

[8]McFadden and Train (2000) lay out a simple procedure using a Lagrange multiplier test from estimating the MNL model, which is carried out for robustness. See Appendix A.2.1 for the detailed procedure.

higher mean VTTS in the long and medium-length trip scenarios: \$41/hour in the long trip and \$56/hour in the medium trip. However, the mean VTTS from the short trip is lower than that from the MNL model, at \$25/hour. The additional penalty on waiting time is similar between the two models, estimated to be \$18/hour in the MNL model and \$19/hour in the MMNL model. Willingness to pay for WiFi availability is \$6.37 and \$6.65 from the MNL and MMNL model estimates, respectively.

2.5.4 Tests for Respondent Biases

In DCEs, it is possible that respondents behave in ways to strategically complete the task without revealing their true preferences. This strategic method of completing the tasks increases the risk of lexicographic biases being present in the choice models, which threaten the validity of valuation estimates (Sælensminde, 2006; Meginnis et al., 2021). For robustness of the results, I test the MNL model results for two important ways in which respondents could strategically complete the DCE tasks without revealing their true behavior in their preferences. First, I test whether the order in which alternatives were shown to respondents in the DCE is a relevant attribute. If the null hypothesis is rejected, then the DCE exhibits an *ordering bias*. Second, I test for the stability of the coefficients of the MNL model by eliminating respondents who repeatedly choose options listed first, second, or third, which is a practice known as *straight-lining*[9].

Tests for Ordering Bias

The first test verifies whether alternatives are more likely to be chosen if they are shown to respondents in a specific order in the choice tasks. The procedure for this test is straight-

[9]These two tests are not the same, the first test imposes is a test on the attribute coefficients, while the second is a test on the individual level characteristics (i.e. whether the individual chose the same order option).

forward: two dummy variables are added to the MNL specification, one for alternatives displayed in the second column, and another for alternatives displayed in the third column. The MNL is then re-estimated with these dummies, and a likelihood ratio test is conducted to verify the joint significance of the two dummy coefficients.

Table 2.7: Likelihood ratio test for the joint significance of coefficients on dummies for alternatives listed second or third.

Variable	**No Dummies**		**With Dummies**	
	β	SE	β	SE
Travel Attributes				
Travel time (long trip)	-0.018	(0.004)	-0.018	(0.004)
Travel time (medium trip)	-0.034	(0.005)	-0.034	(0.004)
Travel time (short trip)	-0.021	(0.007)	-0.020	(0.006)
Travel cost	-0.040	(0.012)	-0.040	(0.012)
Wait time	-0.012	(0.006)	-0.012	(0.006)
WiFi availability	0.255	(0.046)	0.255	(0.046)
Option listed second	–	–	–0.004	(0.051)
Option listed third	–	–	–0.006	(0.051)
Likelihood-ratio χ^2 (d. f.)			0.02	(2)

Table 2.7 displays the MNL estimates when dummies are introduced for options listed in the second or third columns. For brevity, only the coefficients for alternative-specific attributes are shown and it is evident that they are identical across both specifications. It is also clear that the dummy coefficient estimates are both insignificant, meaning there is no evidence of systematic bias leading the sample to gravitate towards alternatives listed first compared to others listed second or third in each task after controlling for trip attributes and individual characteristics. The likelihood ratio test for joint significance resoundingly fails to reject the

null hypothesis that the dummy coefficients are zero.

Test for Straight-Lining Bias

The second test is to determine the bias caused by respondents who "straight-line", or pick any alternative that is placed in a particular column of the DCE they are shown. To perform the test, a dummy variable is used to identify individuals who pick the same alternative order in each scenario. The MNL model is refit on the non-straight lining observations and the straight-lining observations separately and then a Chow test is performed to test the hypothesis that the parameter estimates are equal across the two groups of observation. The procedure is to perform a likelihood ratio test using the log-likelihood of the estimation on the pooled sample and the combined log-likelihoods from the individual regressions.

Table 2.8: Likelihood ratio test for the joint significance of coefficients on dummies for alternatives listed second or third.

Variable	**Pooled Sample**		**No Straightlining**		**Straightlining**	
	β	SE	β	SE	β	SE
Travel Attributes						
Travel time (long trip)	-0.018	(0.004)	-0.017	(0.004)	-0.037	(0.013)
Travel time (medium trip)	-0.034	(0.005)	-0.039	(0.005)	(wrong sign)	–
Travel time (short trip)	-0.021	(0.007)	-0.021	(0.006)	-0.039	(0.022)
Travel cost	-0.040	(0.012)	-0.046	(0.012)	(wrong sign)	–
Wait time	-0.012	(0.006)	-0.010	(0.006)	-0.043	(0.021)
WiFi availability	0.255	(0.046)	0.245	(0.049)	0.272	(0.157)
Number of individuals (N):	988		874		114	
Number of choice situations:	3,434		3,092		342	
Number of total observations:	11,282		10,158		1,124	
Log pseudolikelihood	-3297.365		-2559.690		-310.937	
Likelihood-ratio χ^2 (d. f.)	53.47*	(35)				

Table 2.8 shows the results from the Chow test for lexicographic bias from straight-lining in the DCE task. When only respondents who straight-lined all choice tasks are included, the parameter estimates are not indicative of rational behavior because the travel cost and medium-length trip travel time coefficients are unexpectedly positive. Furthermore, the χ^2 test statistic produced from the Chow test is statistically significant at the 5% level ($p = 0.024$), indicating that there is evidence that a small bias is introduced thanks to participants who straight-lined across the DCE tasks. However, these respondents make up a small proportion of all individuals in the sample – approximately 1 in 9 respondents and 1 in 10 choice scenarios. Despite the statistical significance of the bias, the differences in the magnitudes of the coefficients between the pooled sample and the sample without straight-

lining are not meaningful enough to compromise the valuations presented in Table 2.6 or the qualitative interpretation of the results in Table 2.5[10]

2.6 Policy Scenarios and Evaluation

Using the estimation results, it is now possible to consider the impacts of policy interventions and predict the effects of each intervention. Results are taken from the MMNL model to test three policy interventions from the DCE results. The first policy intervention is a reduction in transit fares, with an emphasis on the zero-fare case, which recreates the GoPass program. The second policy intervention I test is a reduction in travel times among transit alternatives. The third and final policy intervention of the analysis is the introduction of a congestion charge on all car alternatives. In each policy intervention, two outcomes of interest are predicted: changes to the choice probabilities of each mode and changes to consumer surplus.

2.6.1 Choice Probabilities

To evaluate changes to the choice probabilities, I consider each policy change and compute the predicted choice probabilities at varying intensities of the policy intervention, which are then adjusted by the sampling weights of the individual observation. Since the universe of possible modes varies across each trip scenario, it is sensible to assess the changes to mode choice in individual scenarios instead of making a joint evaluation across all scenarios. For simplicity, modes are grouped into three categories: transit (buses and trains), cars, and all other modes. The first policy intervention studied is a reduction in fares, carried out in 10% increments until fares are eliminated, which emulates the GoPass program. Figure 2.3 shows the predicted impacts of fare reductions on the choice probabilities.

[10]See Appendix A.2 for the MMNL model estimates after excluding the straight-lining observations from the sample.

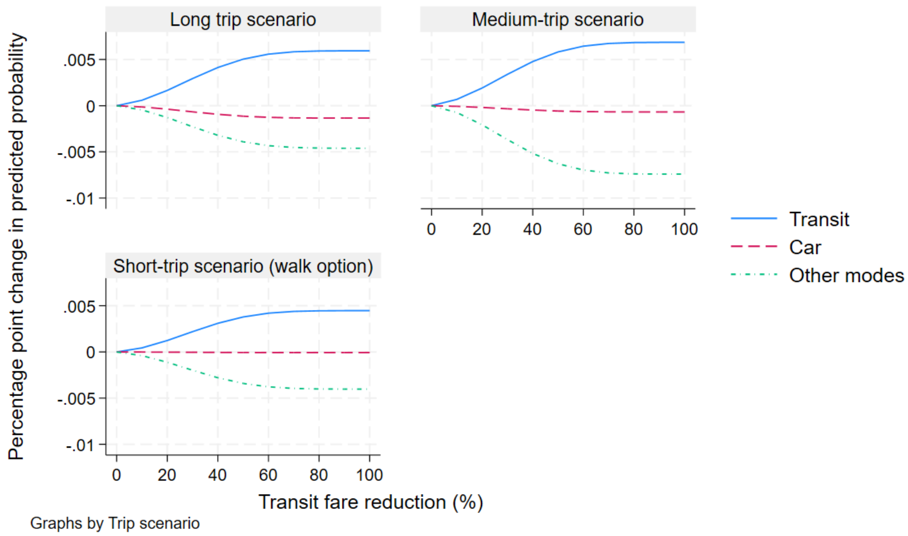


Figure 2.3: Changes to choice probabilities of mode types across fare reduction intensities.

In the long trip scenario, a decrease in transit fares increases the choice probability of transit options, but this change comes at the expense of other non-car modes instead of cars, which only sees a 0.1 percentage point decrease in the predicted choice probability. Notably, the mean transit choice probability hardly increases, seeing about a 0.5 percentage point increase even when fares are eliminated. In the medium-length trip scenario, the mean choice probability of transit modes jumps slightly more than in the long trip scenario, but this increase is substituted almost entirely from the other non-car modes and not cars. In the short trip scenario, the increase in predicted transit choice probabilities is the smallest of the three scenarios, and again, this increase is generated almost exclusively from substitution from other non-car modes. An important caveat is that cars were the least-included mode in the respondents' choice sets, which mechanically forces the choice probabilities of cars to be low. However, in the long trip scenario, where cars were shown most often, substitution from cars to transit was low after transit fares were reduced.

The second intervention studied was a reduction in travel times of all transit alternatives. Realistically, this can be achieved through infrastructure projects such as installing dedicated bus lanes, converting existing bus routes to bus rapid transit, and improving signalization on train routes. As was the case with the fare reductions, travel times on transit were reduced incrementally, in steps of 5% until reaching a 25% reduction in travel times, and predicted choice probabilities were computed and adjusted based on the sampling weights. The impacts to the choice probabilities are shown in Figure 2.4.

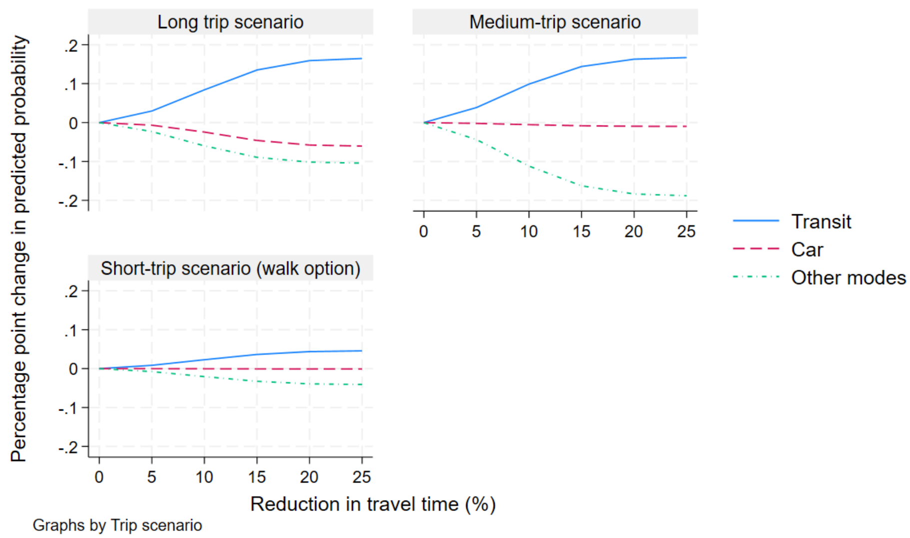


Figure 2.4: Changes to probabilities of mode types across travel time reductions.

Relative to the fare reduction case, reducing travel times on transit alternatives had a much more positive impact on the predicted transit choice probabilities. In the long trip scenario, a 5% reduction in travel times on transit increased the transit choice probability by roughly 3 percentage points, which exceeded the effect of eliminating fares. Gains in the transit choice probability leveled out and reached a roughly 18 percentage point increase as travel times were reduced by 25%. Notably, the reductions in travel times led to some mode switching from cars, but more switching was observed from the other non-car modes. In the medium-length trip scenario, the gains in the transit choice probabilities exhibited a similar pattern to the ones shown in the long trip scenario, with the exception that mode substitution came solely from other non-car modes. In the short trip, the gains in the transit choice probabilities were muted compared to the other two trip scenarios, but the transit choice probability still increased by about 5 percentage points as travel times were reduced by 25%.

The final intervention studied is a congestion charge levied on cars. These are generally implemented by dynamically setting a toll for entering a demarcated zone, which regulates traffic demand within that zone. In the interest of simplicity, I ignored the dynamic nature of real-world congestion charges and applied the congestion charge universally, studying its impacts in $2 increments. Once again, the choice probabilities of each mode group were computed and adjusted using the sampling weights of the survey, and repeated until the congestion charge was $10. The results of this analysis are shown in Figure 2.5.

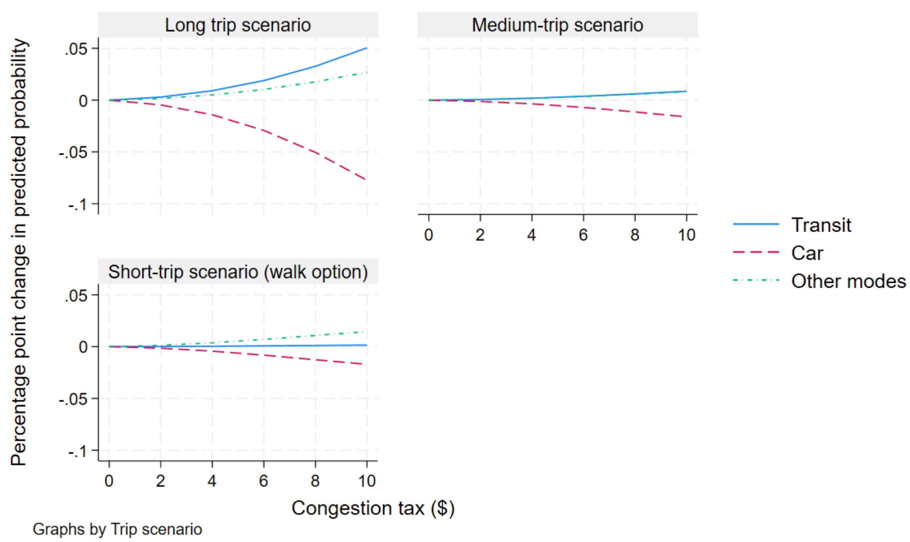


Figure 2.5: Changes to probabilities of mode types across levels of congestion taxes.

The first notable difference with this policy intervention is that the concavity of the choice probabilities seemed different, but this is only because $10 is not a sufficiently large window to show the full trajectory of the mode choice probabilities as the amount of the congestion charge increased[11]. In the long trip scenario, a congestion charge did not substantially

[11]In all cases, the graphical representation between choice probability P_{ni} and trip attribute x_{ni} must exhibit a sigmoid function, since the probabilities conditional on the parameters are a logistic function of the alternative's attributes. The graph's steepness, characterized by $\partial P_{ni}/\partial x_{ni}$, is positively correlated with the respective β. However, two of the three involve decreasing individual attributes x_{ni} for which $\beta < 0$, which explains the positive sigmoid shape of the transit shares. See Figure A.1 in Appendix A.2

decrease car demand until the congestion charge exceeded \$5. When the congestion charge was increased to \$10, the car choice probability decreased by about 8 percentage points. Substitution from cars to transit was evident, with the transit choice probability increasing by 5 percentage points when a \$10 congestion charge was implemented. Substitution to other non-car modes was more modest, as their choice probabilities only increased by 3 percentage points. In the medium-length and short trip scenarios, the decrease in car choice probabilities is much smaller – but this is because car alternatives were rarely shown to respondents in these two scenarios, so even a \$10 charge only decreased the choice probability of cars by 1-2 percentage points. Mode substitution in the medium-length trip scenario was split evenly between transit and other non-car modes, while mode substitution in the short trip scenario led to an increase in the choice probability of the other non-car modes instead of transit.

2.6.2 Consumer Surplus

In this analysis, another outcome of interest is to compute the consumer surplus changes that result from each specific intervention. In practice, the "logsum" approach is used to compute changes to consumer surplus in discrete choice analysis (De Jong et al., 2007; Odeck et al., 2003). For simplicity in notation, changes to consumer surplus will be analyzed within trip scenarios, which results in dropping the notation k for each scenario of the DCE. Using the formula presented in Train (2009), the logsum formula using the MNL model starts as

$$E[CS_n] = \frac{1}{\alpha_n} E[\max_j (V_{nj} + \varepsilon_{nj}) \forall j] = \frac{1}{\alpha_n} \log \left(\sum_{j \in C_n}^{J} \exp (V_{nj}) \right) + K. \tag{2.10}$$

Here, α_n is individual n's marginal utility of income and K is a constant which, without loss of generality, can be set equal to zero. However, in the MMNL model (see Section 4.2), V_{nj} contains the random parameters β_n, which are drawn from the distribution $f(\beta|\theta)$. Recall that the term $\log \left(\sum_{j \in C_n}^{J} \exp (V_{nj}) \right)$ is simply the denominator from the expression for $L_{ni}(\beta_n)$ (see Equation (2.5)). Because V_{nj} contains the expression β_n and β_n is not observed, it is necessary to integrate over all possible values in the distribution of β, which modifies Equation (2.10) to

$$E[CS_n] = \frac{1}{\alpha_n} \log \left(\sum_{j \in C_n}^{J} \int \exp\left(V_{nj}\right) f(\beta|\theta) \, \mathrm{d}\beta \right).$$

Once again, it is not feasible to compute a closed-form solution of the integral in the above expression, and the values of β_n are drawn from the distribution $f(\beta|\theta)$ a total of R times. Each drawn vector of individual parameters β_n takes β^r, and the integral is approximated by $\frac{1}{R}\sum_{r=1}^{R} \exp\left[V_{nj}(\beta^r)\right]$. The final component needed to evaluate consumer surplus is α_n. However, α_n is not directly observed from the DCE. By construction, however, travel cost enters utility by subtracting from income, which means that α_n is exactly the negative of the cost coefficient. The MMNL model specification presented in this paper contains a singular travel cost coefficient, γ_{TC}, which allows for a simple final expression for computing consumer surplus:

$$E[CS_n] = -\frac{1}{\gamma_{TC}} \log \left(\sum_{j \in C_n} \frac{1}{R} \sum_{r=1}^{R} \exp\left[V_{nj}(\beta^r)\right] \right). \tag{2.11}$$

From a policy perspective, reporting raw consumer surplus levels is uninformative. Instead, it is more relevant to study *changes* to consumer surplus. The procedure for studying consumer surplus changes is straightforward: first, a "baseline" consumer surplus level is computed, then a new consumer surplus level is computed after implementing a specific policy intervention, and then the change is merely the difference between the two consumer surplus levels. Utilities are denoted by V^0 and V^1, to represent utilities at the baseline and utilities after the policy change respectively. Using this convention, the change in consumer surplus can be expressed as

$$\Delta E[CS_n] = -\frac{1}{\gamma_{TC}} \left\{ \log\left(\sum_{j \in C_n} \frac{1}{R} \sum_{r=1}^{R} \exp\left[V_{nj}^{1}(\beta^r)\right] \right) - \log\left(\sum_{j \in C_n} \frac{1}{R} \sum_{r=1}^{R} \exp\left[V_{nj}^{0}(\beta^r)\right] \right) \right\}. \tag{2.12}$$

Changes to consumer surplus were analyzed in the context of three specific policy interventions discussed in Section 6.1. First, the impact of eliminating transit fares, similar to the GoPass program, was examined. Second, a 5% reduction in travel times for all transit alternatives was considered. Third, a $5 congestion charge was applied to all car alternatives. Consumer surplus changes were first computed for each individual and then adjusted using each individual's sampling weight. These measures of consumer surplus were aggregated in two ways:

- First, the mean change in consumer surplus was calculated across all individuals within each trip scenario.
- Second, the weighted sum of consumer surplus changes was computed and scaled to the total population of the LACCD student body. This provides aggregate consumer surplus changes under the assumption of one trip per student.

Importantly, consumer surplus changes were not aggregated across different scenarios because each scenario represented distinct travel conditions, which affected consumer preferences and welfare outcomes in unique ways. Therefore, cross-scenario aggregation of consumer surplus changes was inappropriate. The resulting changes in consumer surplus are presented in Table 2.9.

Table 2.9: Comparison of aggregate and per person-trip consumer surplus changes for different policy interventions, by trip scenario. (95% confidence intervals for mean per person-trip consumer surplus changes are shown in parentheses.)

Policy	**Scenario**		
	Long	Medium	Short
Panel A: Aggregate consumer surplus changes ($, in '000s)			
GoPass ($0 fares on transit)	112.8	122.5	38.6
5% reduction in travel times	501.1	549.1	39.1
$5 congestion charge	–297.0	–84.6	–84.6
Panel B: Mean consumer surplus change ($, per trip)			
GoPass ($0 fares on transit)	0.56	0.60	0.19
	(0.53, 0.58)	(0.58, 0.63)	(0.18, 0.21)
5% reduction in travel times	2.47	2.71	0.19
	(2.41, 2.53)	(2.64, 2.77)	(0.18, 0.20)
$5 congestion charge	–1.47	–0.42	–0.42
	(–1.61, –1.32)	(–0.50, –0.33)	(–0.50, –0.33)

When the GoPass program was implemented in the experiment, the aggregate increase in consumer surplus was approximately $113,000 for one trip per student in the long-trip scenario. In the medium-length trip scenario, the aggregate surplus gain increased to $122,500, while in the short trip scenario, the aggregate change in consumer surplus was around $39,000[12]. This suggests that, individually, the GoPass intervention delivers an average of approximately 56 cents of additional consumer surplus per trip in the long-trip scenario, 60 cents per trip in the medium-length scenario, and 19 cents per trip in the short-trip

[12]Aggregate changes to consumer surplus were also disaggregated by campus in Appendix A.2.

scenario.

The second policy tested was a 5% reduction in transit travel times. Across all LACCD students, a 5% reduction in travel times on transit resulted in an aggregate consumer surplus increase of $501,000 in the long-trip scenario, which corresponds to a mean consumer surplus gain of $2.47 per trip. In the medium-length trip scenario, the aggregate surplus increase was $549,000, leading to a mean consumer surplus gain of $2.71 per trip. In contrast, the short-trip scenario showed a smaller aggregate gain of $39,000, with a mean consumer surplus gain of only 19 cents per person per trip. This demonstrates that the consumer surplus gains from improving travel times on transit are significantly higher compared to eliminating fares.

Lastly, a $5 congestion charge was tested. Across all students in the long-trip scenario, the congestion charge resulted in a consumer surplus loss of $297,000, equating to a mean individual consumer surplus loss of $1.47 per trip. In the medium-trip scenario, the consumer surplus loss was much smaller, at around $85,000 in the aggregate, or 42 cents per trip. The welfare losses in the medium and short trip scenarios were almost identical, both showing a loss of about 42 cents per person per trip.

2.7 Discussion

In this paper, I use a discrete choice experiment (DCE) to assess mode choice and valuations of trip characteristics among community college students in Los Angeles. The analysis included varying trip lengths and was complemented with policy simulations to assess the impacts of LA Metro's GoPass program and other policy interventions on mode choice and welfare. Overall, the results reveal that cars were the preferred mode of transportation across all scenarios, except for walking in the two-mile trip scenario. Key individual characteristics that influenced mode choice included possession of a driver's license and GoPass participation. Specifically, GoPass participation was associated with a higher probability of selecting bus alternatives in the DCE but not trains, suggesting a higher preference for buses among students participating in GoPass. Interestingly, GoPass holders were less inclined to select the self-driving car service. Additionally, having a driver's license predictably decreased the probability of selecting transit options, capturing the effect of car access on travel decisions.

Valuations of travel time savings varied based on the trip length and the model specification used. In the simple MNL model, VTTS ranged between \$27 and \$52 per hour, while the MMNL model estimated that VTTS ranged between \$25 and \$54 per hour. These valuations do not diverge from prior estimated values of VTTS in the literature, such as the \$57 per hour from Steimetz and Brownstone (2005); Brownstone et al. (2003), and the \$45/hour VTTS from Lam and Small (2001)[13]. Additionally, the results find that WiFi availability is valued at approximately \$6 to \$7 per trip. However, interpreting these valuations requires great caution, because they are derived from taking the ratio of two estimates, which are random variables and are therefore not guaranteed to have finite moments[14]. For that reason, these interpretations are strictly *qualitative*. For greater interpretability, future research can use reparametrization methods like those in Train and Weeks (2005) that estimate willingness-to-pay directly, but these require strong parametric assumptions of the WTP distribution.

Concerns about lexicographic biases in the DCE, such as systematic preferences for a specific option order or straight-lining (Hess et al., 2010), were taken into account through design choices or post-estimation tests. The analysis found no significant evidence of any particular option order being a relevant attribute, after controlling for the trip attributes in the design and individual characteristics. Straight-lining behaviors, where respondents could make their choices based on specific patterns, were also mitigated. Two types of straight-lining were accounted for in this analysis: the first was cases in which respondents could blindly select

[13] Adjusted to 2024 dollars using CPI values

[14] If the parameter estimates for travel time and travel cost are normally distributed, then VTTS follows a Cauchy distribution, which is known not to have defined moments.

a specific mode for all scenarios, which was restricted by modifying the choice set that individuals saw across every scenario. The second was cases in which respondents always selected alternatives in a specific option order, which was accounted for through a post-estimation test. While 10% of all respondents displayed straight-lining behaviors by selecting alternatives in the same option order and the hypothesis test was statistically significant, these cases did not materially impact the qualitative interpretations or valuations derived from the models.

From a policy standpoint, the findings highlight that the willingness to pay for reduced travel times is high. Simulations showed that even a 5% reduction in travel times on transit had a more pronounced impact on mode choice than eliminating fares on transit. While free transit increased consumer surplus, its effects on mode substitution were limited compared to improvements in travel times. Congestion charges were welfare-damaging in this analysis, but they created the highest degree of mode substitution away from cars. However, car alternatives were underrepresented in the DCE relative to their dominance in students' actual choice sets, which is a potential limitation in translating these findings to broader contexts.

The trip length variations were included in the DCE to reflect the diverse travel needs of LACCD students and capture distinct contexts in which students would demand travel. Passenger-mile data from LA Metro indicates that the short and medium trip scenarios align with a typical bus trip, while the medium and long trip scenarios better represent rail travel (Los Angeles County Metropolitan Transportation Authority, 2024c). Car trip lengths from the 2022 National Household Travel Survey (NHTS) (Federal Highway Administration, 2022) align with the long trip scenario, which provides further context to the analysis. Across all trip lengths in the DCE, improved welfare gains from reducing travel times on transit persisted, emphasizing that this finding is robust to different trip lengths.

There are also practical considerations to implement the policy implications of this analysis.

Travel time reductions on transit can be achieved through infrastructure improvements such as increasing the number of dedicated bus lanes, implementing strict enforcement of bus lane violations, creating bus rapid transit right-of-ways, and improving signalization on existing train lines. However, if reducing trip generation from car trips is a stated goal of the GoPass program, then eliminating fares is insufficient. Congestion taxes offer a direct way to reduce car usage, but are politically challenging and potentially welfare-damaging without complementary measures. Case studies from Singapore suggest that pairing congestion pricing with transit service improvements can enhance transit ridership (Olszewski and Xie, 2005) in the presence of aggregate welfare losses (Wilson, 1988). Alternatively, an implementation in a narrower context, such as increased parking fees, could serve as a significant incremental step toward achieving broader congestion management goals.

2.8 Conclusion and Future Work

This work contributes to the travel behavior literature by studying community college students and employing an experimental approach to assess the potential benefits of removing fares on transit, modeled after LA Metro's GoPass program. Using a discrete choice experiment, create three trip scenarios of varying trip distances were designed to test the effects of a GoPass-like policy and other travel demand management measures. The findings presented validate the idea that travel time reductions on transit, rather than fare elimination, generate greater welfare gains and improvements in transit mode choice.

Recently, the travel behavior literature has increasingly questioned the reliability of estimates from discrete choice experiments (DCE) and their corresponding policy prescriptions (Brownstone and Small, 2005; Tsoleridis et al., 2022; Tveter, 2023). While these critiques are valid, other literature suggests that D-efficient designs can mitigate the differences stated preference valuation estimates and revealed preference outcomes in some contexts (Fezzi et al., 2014). Additionally, including opt-out options, as implemented in this study, can also help bridge the gap between stated preference and revealed preference valuations (Hensher, 2010). The present study justifies the use of stated preference methods for several reasons: first, logistical hurdles to collecting revealed travel behavior are more pronounced among college students compared to the general public; second, discrete choice experiments remain the only viable method for measuring mode choice preferences for emerging or non-existent transportation modes; and third, given the GoPass program's implementation before this study, replicating its impacts is only feasible through an experimental setting.

Even when the values of travel time savings varied, the policy insight that travel time reductions outweighed fare elimination from a welfare perspective remained consistent. This aligns with the intuitive notion that LA Metro's previous fare discount programs for students positioned transit to be perceived as a relatively low-cost option, even for a low-income cohort like LACCD students. Therefore, if transit is perceived as unreliable rather than costly, then students stand to gain more from modest reductions in travel time on transit (e.g. 5%) than from fare elimination. On the flip side, the GoPass program remains a low-cost initiative to administer, requiring just $20 million in 2023–24 (Los Angeles County Metropolitan Transportation Authority, 2024b)[15]. The program's low costs make it a financially viable intervention compared to the infrastructure improvements that would be needed to significantly reduce travel times on transit.

Future work can build on these findings by making use of additional outcomes that are collected in the survey questionnaire. First, GoPass participation itself is an outcome of interest that warrants further exploration – understanding the factors influencing awareness and adoption could inform strategies to improve program reach. Second, the findings from the stated preference choices could be complemented with revealed preference choices from students' trips to and from their college campuses. This exercise would deepen the under-

[15] LA Metro has an operating budget of $9 billion in 2024–25.

standing of students' real-world tradeoffs and likely validate the DCE-derived valuations. Third, the DCE responses analyzed in this study only include students who were assigned a "traditional" DCE, where students only selected their preferred alternative. The remaining students ranked their preferences in the DCE, and the analysis of their responses remains an area for future exploration. Including ranked DCE data would be beneficial because it would enable a comparison of traditional DCE responses with ranked preferences and test the robustness of valuations and welfare estimates under alternative elicitation methods.

In conclusion, this study highlights the importance of travel time reductions and suggests that fareless transit programs such as GoPass can generate modest improvements to consumer surplus for community college students. However, future work can expand on these findings by examining GoPass participation, leveraging RP data, and analyzing ranked preference data.

Chapter 3

The Impacts of Bus Use on COVID-19 Dispersion

This chapter

3.1 Introduction

This paper investigates the links between COVID transmission and bus use in the region served by the Los Angeles Metro system. Los Angeles bus use declined dramatically at the beginning of the COVID pandemic. Some of this decline was clearly caused by stay-at-home orders designed to slow down the spread of COVID. Additionally, the widespread fear that

[01] This study was made possible through funding received by the University of California Institute of Transportation Studies from the State of California through the Public Transportation Account and the Road Repair and Accountability Act of 2017 (Senate Bill 1). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented. This document is disseminated under the sponsorship of the State of California in the interest of information exchange and does not necessarily reflect the official views or policies of the State of California. We are also grateful for help obtaining and processing data from Cassie Hall, Rollin Baker and Joshua Schank (LA Metro) as well as Craig Rindt (UCI ITS). Rollin Baker and editors of this book also provided excellent comments on earlier drafts. We are responsible for all errors and unwarranted conclusions.

crowded buses could spread and transmit the virus across neighborhoods led to large drops in transit ridership, even before stay-at-home orders were formally announced. However, because Los Angeles has more low-wage "essential" workers that depend on buses to get to their jobs compared to other cities such as New York (Czeisler et al., 2020), decline in bus use in many Los Angeles neighborhoods was less drastic than in other parts of the country.

Ideally, establishing a causal link between bus use and COVID transmission would require random assignment of commute trips to either bus or a safe alternative like a single-occupancy car. We could then track COVID infections across the bus and car groups to measure the impact of bus use. However, this experiment is impossible to carry out in practice. The next best alternative is due to Granger (1969), and this is equivalent to causality if the underlying prediction model is correctly specified. We first fit the best possible predictive model of COVID transmission without using any information about bus use, and then we add bus use information to this model. If adding bus use information improves the predictive ability of the augmented model, then we will say that bus use "causes" COVID transmission. We do not find any evidence that knowing bus use improves our ability to predict COVID transmission in the Los Angeles Metro system area. This is consistent with the claim that mitigation measures (including mandatory mask wearing and improved ventilation) were successful in stopping the spread of COVID on LA Metro buses. Since our data stop before the spread of the Delta variant or the rollout of COVID vaccinations we do not know whether our conclusion still holds.

The next section of this chapter summarizes the literature on COVID transmission and early research done on the role of transportation modes in COVID transmission. Section 3.3 offers a description of the data sources used in our analysis, while Section 3.4 describes the methods used to analyze the relationship between bus ridership and COVID transmission in Los Angeles County. Section 3.5 presents the results of our analyses, and Section 3.6 offers a discussion of possible interpretations and ramifications of our findings.

3.2 Literature Review

The main transmission mechanism of COVID is through transmissions from individuals who either experience an asymptomatic infection or who later experience symptoms (Yu et al., 2020; Bai et al., 2020; Wei and Wehby, 2020). It is also well-documented in the literature that COVID is better spread in indoor settings (Qian et al., 2020). Mitigation measures such as face covering mandates and capacity restrictions were believed to reduce COVID transmission and prevent infections within weeks after their introduction in areas where such mandates were imposed (Li et al., 2020; Sun and Zhai, 2020). Furthermore, COVID transmission in indoor settings can also be mitigated through adequate ventilation where air exchanges are frequent (Blocken et al., 2021), as well as in spaces where thermal destratification (i.e., warmer, contaminated air rises to the ceiling of a ventilated space) is better encouraged (Bhagat et al., 2020).

Some early epidemiological literature studying the 2003 SARS epidemic points to a relationship between transit use and transmission. Wang (2014) provides a rather bleak outlook for transit use in Taipei, finding an immediate ridership decline of 1200 trips for each new SARS case that was announced in its underground rapid transit. Further, Wu et al. (2004) found that taking a bus or riding the subway in Beijing more than once a week was a potential risk factor for contracting SARS. More recently, an investigation of a COVID outbreak following a Buddhist worship event in China found that a large portion of the infected individuals contracted the virus through airborne transmission while riding a bus to the event (Shen et al., 2020). However, Hu et al. (2021) show evidence that transmission is heavily dependent on passenger density, seat spacing, and co-travel time (the time that passengers are aboard with other passengers). Hence, transmission can be reduced if measures are taken to reduce passenger density by increasing spacing between seats and minimizing co-travel time.

The argument that elevated transit use or high population density directly translate to increased rates of community spread of COVID is contentious. For instance, Hamidi and Hamidi (2021) find no evidence of a direct link between subway ridership and community COVID cases in New York City, offering an alternative hypothesis that transmission of COVID is more strongly linked to racial and socioeconomic factors. This does seemingly contrast McLaren (2020), who uses county-level data to find that disparities in COVID deaths among African Americans are correlated with areas where public transportation use is high, although these numbers are based on Census data that pre-date the pandemic. This hypothesis is echoed by Fathi-Kazerooni et al. (2021), who use turnstile data from the New York City subway finding a high correlation between subway ridership and COVID deaths between March and May 2020, detectable across all boroughs. To the best of our knowledge, this is the first study to investigate a dynamic causal link between subway use and COVID transmission in a large metropolitan area. We find no causal link between bus use and COVID transmission in Los Angeles, and this is likely due to differences in the periods examined, mitigation measures, and less crowding compared to New York subways.

Some existing research shows correlations between transit use, COVID, and demographics. Hu and Chen (2021) note that following the onset of the COVID pandemic in Chicago, wealthier, better educated areas saw larger declines in transit ridership than areas with lower income and educational attainment. There is also research and press headlines supporting the hypothesis that neighborhoods and cities with high percentages of people of color were far more likely to see higher death rates and more infections per capita (Chowkwanyun and Reed, 2020; Shah et al., 2020). Moreover, increased susceptibility to COVID among ethnic and racial minorities can be traced to these groups representing a disproportionate share of occupations categorized as essential in the pandemic (Rogers et al., 2020).

3.3 Data

We constructed a panel of monthly COVID cases and average weekly bus ridership within Countywide Statistical Areas (CSAs) in Los Angeles County (we will refer to these areas as "neighborhoods"). We supplement these data with socio-demographic variables like the ones used in McLaren (2020) on race and ethnicity, as well as variables about income, education, commuting, population, and household density.

3.3.1 COVID Data

Our data on COVID cases comes from the Los Angeles County Public Health agency and provides weekly counts of new COVID cases for 236 neighborhoods in LA County from March 2020 through January 2021. We obtained historical data for each CSA for each week between March 16, 2020 to January 31, 2021. We aggregated COVID cases at the CSA level to monthly frequency because of some reporting discrepancies or irregularities. This aggregation removed negative case numbers in the weekly data. These were likely due to data corrections that resulted in a case assigned in a certain CSA and being reassigned to another. Using a larger time unit reduced possible noise from reporting lags or periodic case backlogs.

3.3.2 Transit Ridership Data

Data on transit ridership are from the Los Angeles County Metropolitan Transportation Authority (Metro). First, we used ridership data at each bus stop for the months of October 2019, April 2020, and October 2020. Second, we used ridership on each bus line (not disaggregated by stop), as well as passenger-miles traveled, which were available monthly beginning October 2019 and ending in December 2020.

Because geographic dispersion of bus ridership was only known for three specific months, we imputed the CSA-specific bus ridership for the remaining months. Using the stop-specific data, bus stops were mapped to CSAs, and ridership numbers within CSAs were added to obtain CSA-level ridership in October 2019, April 2020, and October 2020. We then used ridership and passenger-miles numbers by line in the remaining months to generate estimates of the share of ridership that each CSA comprises within each line, as well as estimates of the share of ridership that each line comprises within a specific CSA. These ridership shares were then used to estimate CSA-level ridership in months outside those for which stop-level ridership was available. Additionally, we estimate passenger-miles traveled for all months between October 2019 and December 2020 (see Section 4.1 and the Appendix for a detailed description of bus ridership imputations).

Metro is by far the largest transit provider in Los Angeles County, but there are 66 additional transit operators serving cities other than Los Angeles. LA Metro provides bus service that overlaps the service areas of some of these agencies, so it is likely that our data underestimate bus utilization in some neighborhoods served by multiple transit agencies. Our key results rely on differences in bus utilization and COVID cases, so if bus utilization from other agencies changes proportionally with changes in LA Metro bus utilization our results are still valid. LA Metro also provides some bus service in parts of Orange and Ventura counties, and we exclude these data from our analyses.

3.3.3 Demographic Data

To supplement our data on bus ridership and COVID cases at the CSA level, we use geographic information software (ArcGIS) to impute demographic data. We draw our data from two sources: the first is ESRI's demographic estimates using up-to-date Census estimates, and the second is the 2018 5-year American Community Survey (ACS). We categorize our demographic variables into four main groups: race and ethnicity, income and education, commuting, and population and density.

1. Race and Demographics - We examine four minority groups defined by the Census: ‘Hispanic or Latino’; ‘Black or African American’ (abbreviated as ‘Black’); ‘American Indian or Alaska Native’, (identified as ’First Nations’; and ’Asian’). These data are obtained from Esri’s 2020 estimates and imputed at the CSA level.

2. Income and Education - We examine two key income measures: median household income and households below the poverty line. Data on educational attainment and median household income are obtained from Esri and data on poverty are obtained from the ACS and imputed using geographical information software.

3. Commuting - We obtained commuting information using ACS data for those who drive alone to work and those who rely on public transit.

4. Population and Density - We obtained from ESRI data for total population and total number of households within a CSA. The quotient of these two variables then gives a rudimentary estimate of the average household density within a CSA.

3.3.4 Descriptive Statistics

3.2 shows how these cases are distributed across neighborhoods and across time. There is a clear wave of new cases in June and July 2020 followed by a much larger wave in November, December, and January, and there is a lot of variation in cases across neighborhoods during these peaks. We also collected bus boardings and trip length information from Los Angeles Metro and aggregated these data into the same neighborhoods defined by the Los Angeles County Public Health agency (see detailed explanation later in the chapter). 3.2 shows the distribution of bus boardings and alightings per 100,000 population across these neighborhoods by month. The steep decline in March and April 2020 is followed by a slow recovery to levels well below pre-pandemic usage. 3.3 shows a map with the CSAs included in our analysis.

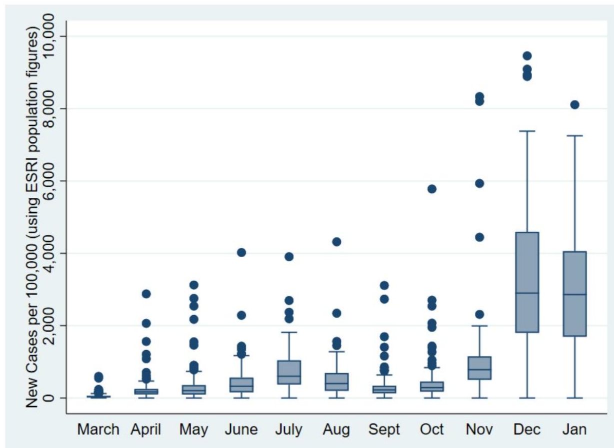


Figure 3.1: Distribution of new COVID cases across neighborhoods by month beginning March 2020

Notes: The shaded rectangles show the area between the 25^{th} and 75^{th} percentile of the distribution across neighborhoods, and the "whiskers" are drawn at the 2.5^{th} and 97.5^{th} percentile. The dots represent "outlier" observations outside of the central 95% of the distribution. The plots exclude 0.29% of the observations that were above 10,000 new monthly cases per 100,000 population.

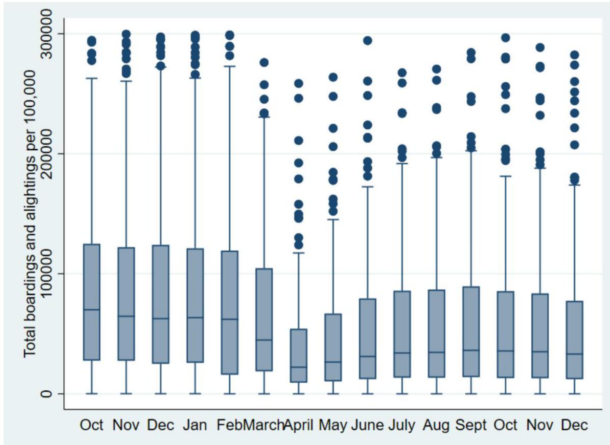


Figure 3.2: Distribution of average weekly bus boardings and alightings across neighborhoods by month beginning October, 2019

Notes: The plots exclude 6% of the observations that were above 300,000 boardings and alightings per 100,000 residents. We also exclude data from stops outside of Los Angeles County.

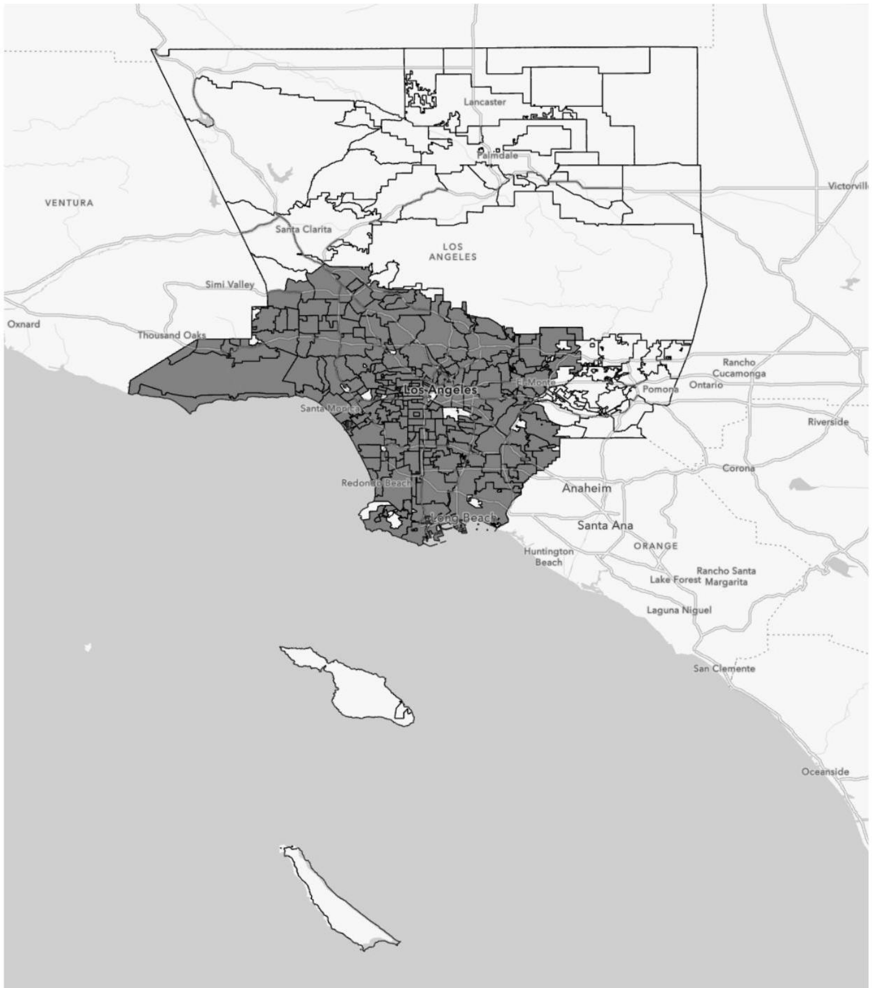


Figure 3.3: Map of included (shaded) and excluded (unshaded) neighborhoods

Table 3.1 shows the distribution of the variables used in our analysis for the full sample and the smaller sample used for estimating the models described in the Results section. The estimation sample excludes all neighborhoods with no Metro bus boardings as well as those with not enough monthly observations to support the lags in our preferred dynamic model. We also excluded all observations with bus boardings per 100,000 population greater than or equal to 400,000 (about 2% of the full sample) because they had undue influence on the precision of estimated parameters; Figure 3 shows the included and excluded neighborhoods. These exclusions account for the large differences in the average bus boardings across the full and estimation sample. Generally, the two samples are quite similar. The neighborhoods in the estimation sample have somewhat higher population, lower median income, and slightly more households below the poverty line.

Table 3.1: Descriptive statistics

	Full Sample		Estimation Sample	
Demographics sample sizes	329		231	
Variable	Mean	S. D.	Mean	S. D.
Average COVID cases*	970	833	918	446
Total population	30904	42638	36175	43319
Total number of households	10139	14439	12091	15106
Household density	3.18	0.97	3.08	0.74
Fraction of Hispanic descent	0.43	0.27	0.45	0.28
Fraction non-Hispanic Black	0.090	0.144	0.109	0.164
Fraction Native American	0.002	0.002	0.002	0.001
Fraction Asian	0.135	0.143	0.143	0.150
Median household income	81830	38662	76491	37715
Fraction households below poverty line	0.046	0.028	0.052	0.028
Fraction without high school diploma	0.069	0.055	0.076	0.057
Fraction high school graduates	0.126	0.047	0.121	0.045
Fraction with GED	0.012	0.008	0.011	0.006
Fraction with some college	0.130	0.042	0.122	0.034
Fraction with associate degree	0.049	0.020	0.045	0.016
Fraction with bachelor's degree	0.160	0.093	0.166	0.098
Fraction with graduate degree	0.095	0.082	0.098	0.084
Fraction who drive alone to work	0.343	0.073	0.349	0.062
Fraction who take public transit to work	0.025	0.028	0.032	0.031
Covid Variables sample sizes	3465		2496	

New monthly cases*	964	1732	917	1314
Cumulative cases* (first lag)	2022	2897	1955	2400
Bus variables sample sizes	2548		2496	
Bus boardings/alightings* (first lag)	154229	1357613	66336	80370
Average bus trip miles (first lag)*	3.68	1.59	3.69	1.59

3.4 Methodology

3.4.1 Imputing Bus Ridership

To estimate ridership in each CSA for the months for which we did not directly obtain stop-level ridership, we first compute the share of riders from each line that board and alight in each CSA. To give an example, suppose that a bus line passes through two CSAs – neighborhoods A and B – and totals 1000 average weekly riders in April 2020. We find that 400 riders board and get off in neighborhood A and the remaining 600 do so in neighborhood B. In that case, 40% of the line's ridership comes from neighborhood A and 60% from B. Because we know the ridership of each line for each month, we can then recover the number of riders that use each line within each neighborhood.

Now suppose we know that 900 riders boarded the same bus line in May 2020. To obtain estimates of May's ridership, we assign 40% of that ridership to neighborhood A, or 360 riders, and the remaining 540 riders to neighborhood B. We then update the neighborhood shares in the next month for which stop-level ridership numbers are available, which is October 2020. Finally, we take the sum of riders from each line within each neighborhood to recover CSA-level ridership. The only exception to our updating rule was March 2020, where the neighborhood shares from April 2020 were used instead of the neighborhood shares calculated from October 2019 data. Here, we use forward interpolation because of the shocks to transit ridership that occurred from the beginning of the pandemic and the initial issuance of stay-at-home orders from the county and state.

Next, we impute a measure of average passenger-miles, which we use as a proxy for exposure to COVID while riding transit. Due to data limitations, we are required to make a key assumption: because passenger-miles are only available for each line and are unknown at the stop-level, we assume that the distribution of passenger-miles within each line does not depend on the neighborhood where riders board or alight (i.e., we assume the average miles traveled by passengers do not change regardless of where along the line they get on or off the bus). Because we have inferred the share of riders from each line pertaining to each CSA, we multiply the average passenger miles by each neighborhood's share of riders from each line and then add these numbers for each line passing through each CSA. Doing this gives a monthly estimate of the average passenger-miles within each CSA, since passenger-miles data are available monthly. (See Appendix for more detailed description of the imputation process.)

3.4.2 Empirical Approach

Recall that we need to specify a dynamic model to predict monthly COVID cases for each CSA. We then want to test whether excluding bus utilization variables significantly reduces the predictive accuracy of the model. Our analysis is based on the general dynamic panel data model:

$$y_{it} = \sum_{j=1}^{3} \alpha_j y_{i,t-j} + \mathbf{x_{it}}\beta_1 + \mathbf{w_{it}}\beta_2 + v_i + \epsilon_{it}, i = 1, \ldots, N, t = 1, \ldots, T_i$$

Here, y_{it} is our new cases per 100,000 residents in CSA i and month t, while $y_{i,t-j}$ represents the j-th lag of new cases per 100,000 residents; in our model, we use a maximum of three lags (note that a lag is simply the previous month's new COVID cases). Furthermore, we incorporate monthly dummy variables (with values 0 or 1 for each month in our panel) for the exogenous covariates $\mathbf{x_{it}}$ and bus system usage variables as the endogenous covariates, $\mathbf{w}_{\mathrm{it}}$. We will first fit the model without the bus system variables and then see if adding them improves its predictive power.

This model can be estimated by running a regression on the differences from the group means for each neighborhood i since then the unknown neighborhood-level effects, v_i, drop out. Unfortunately, this "fixed effects" estimation makes it impossible to identify the impact of any time-invariant variables such as sociodemographic information. An additional problem is that taking differences from group means induces correlation across the errors over time, and this also makes the lagged dependent variables correlated with the new error term. This would lead to "bad" estimates of β_2 (our parameter of interest), and our model describing a spurious relationship between transit ridership and COVID cases at the CSA level.

The current technique for dealing with these problems is to use the instrumental variable estimator developed by Arellano and Bond (1991) together with clustered (or robust) parameter covariance estimators to deal with the error correlations. We also looked at fixed effects panel data estimators which require the stronger assumption that the lagged variables are uncorrelated with the time-varying error terms. These estimates had lower standard errors but led to similar empirical conclusions.

3.5 Results

We begin by first examining the correlations between COVID cases (using December 2020 data), demographics, and bus system utilization data. Table 3.2 shows regressions with and without bus system variables. Both regressions were fit using observations from 231 neighborhoods that are also used in the dynamic panel data models described later. Our final sample excluded five observations corresponding to very high values of bus boardings and alightings per 100,000 population (more than 400,000). Including these outlying observations did not change the coefficient estimates but did lead to very high and unstable residual and parameter covariance estimates. We use this same sample throughout this section.

Table 3.2: Regressions on average monthly cases between June 2020 and January, 2021 per 100,000 population

	Without bus variables		With bus Variables	
Variable	β	t	β	t
Mean boardings/alightings per 100,000			0.001	2.8
Mean average bus trip miles			-24	-3.3
Total population	0.003	2.3	0.003	2.4
Total number of households	-0.009	-2.1	-0.008	-2.1
Household density	276	5.8	287	6.2
Fraction of Hispanic descent	437	3.6	436	3.8
Fraction non-Hispanic Black	85	0.9	-52	-0.5
Fraction Native American	-10761	-1.1	-9407	-1.0
Fraction Asian	-594	-6.6	-603	-7.0
Median household income	-0.002	-2.4	-0.001	-1.8
Fraction households below poverty line	1838	2.8	1446	2.2
Fraction without high school diploma	1214	2.3	911	1.7
Fraction high school graduates	2299	4.9	2338	5.2
Fraction with GED	-1155	-0.6	-1596	-0.8
Fraction with some college	348	0.6	553	1.0
Fraction with associate degree	292	0.3	712	0.8
Fraction with bachelor's degree	953	2.8	863	2.6
Fraction with graduate degree	610	1.4	538	1.2
Fraction who drive alone to work	48	0.2	13	0.1
Fraction who take public transit to work	179	0.4	-459	-0.9
Intercept	-663	-2.1	-615	-2.0

Note that although both average bus boardings and alightings and average bus trip miles are statistically significant, their inclusion only increases the R^2 from .90 to .91 . The coefficients of the demographic variables are as expected. Neighborhoods with higher residents per household, lower income and lower education are associated with higher COVID cases. Neighborhoods with a higher proportion of Hispanic residents are associated with higher COVID cases, while those with a higher proportion of Asian residents are associated with lower cases.

The cross-section regressions only show partial correlations, and without strong unrealistic additional assumptions do not imply anything about causality. A better approach is to take advantage of the panel nature of our data to see whether the bus system variables can improve dynamic forecasts of COVID infections. Table 3.3 shows the results of estimating our preferred models using Arellano and Bond (1991) panel data instrumental variable regressions with 1560 observations from 231 neighborhoods (from June, 2020 through January, 2021). These estimates allow for unrestricted correlations between the unknown neighborhood-level effects, v_i, and the included variables. The standard errors are clustered on neighborhoods to allow for the possibility of autocorrelated errors within a neighborhood.

Our best forecasting model without using bus system utilization data estimates is shown in the first 3 columns of Table 3.3. This model requires three lags of monthly COVID cases per 100,000 population (as indicated by the highly significant coefficients on these lagged variables); we also found a significant impact of the lagged value of cumulative COVID cases per 100,000. The lagged values of COVID cases as well as cumulative COVID cases are endogenous. We use further lags and levels of these variables together with other exogenous variables as instrumental variables following Arellano and Bond (1991). We also include a full set of month indicator variables to account for any policies (such as stay at home orders or mask mandates) that impact all neighborhoods. The pattern of the estimated month effects is consistent with a peak in July 2020 and a much larger peak in December 2020 and January 2021, which closely corresponds to the peaks of COVID cases in Los Angeles

County. The implied dynamics of COVID infections show a small positive impact from last month's cases since the sum of the first lag of new and cumulative cases is positive. This and the positive impact of lagged cumulative cases is likely due to direct transmission within the neighborhoods and reporting lags. The negative impact of cases from two and three months earlier is likely due to these people having immunity and thus reducing the pool of those who can get infected.

The remaining three columns of Table 3.3 add lagged values of bus boardings and alightings as well as average trip miles to the previous model. These variables are separately and jointly statistically insignificant at conventional significance levels. More importantly, the coefficients of the predictive model are essentially unchanged. The overall R^2 is .89 for both models, and this shows that there is no evidence that bus use has any impact on COVID transmission for this sample. These findings are robust to additional specifications of the estimating equation and additional lags being incorporated into our dynamic panel regression. We also fit models based on Blundell and Bond (2000) estimators with the bus system variables treated as endogenous. The empirical results are the same as in Table 3.3.

Although about 40 percent of the residual variation in the models in Table 3.3 is due to variation in the unobserved neighborhood-level effects, v_i, we could not find any significant interactions with these effects and the demographic variables described in Table 3.2. We used Monte Carlo simulations to recover the distribution of a regression of the mean residuals within each neighborhood on that neighborhood's demographic characteristics. We also tried fitting a generalized correlated random effects model (Mundlak, 1978), which makes stronger assumptions. This resulted in statistically significant but small negative coefficients for the fraction of Native American residents and the fraction of residents with a postgraduate degree. Once the dynamics of COVID transmission are modeled as in Table 3.3, the cross-section correlations shown in Table 3.2 are negligible.

Table 3.3: Dynamic panel data models for new monthly COVID cases per 100,000 residents.

	Bus variables excluded			Full model		
Variable	Coeff.	S. E	t	Coeff.	S. E.	t
New monthly cases (first lag)	-0.50	0.09	-5.5	-0.53	0.15	-3.4
New monthly cases (second lag)	-0.83	0.15	-5.5	-0.83	0.19	-4.3
New monthly cases (third lag)	-1.55	0.25	-6.3	-1.55	0.48	-3.2
Cumulative cases per 100,000 (first lag)	0.68	0.04	16.5	0.70	0.06	11.5
Bus boardings/alightings (first lag)				-0.002	0.003	-0.9
Average bus trip miles (first lag)				-155	129	-1.2
June (= 1 for June 2020)	-801	157	-5.1	-799	171	-4.7
July	-445	141	-3.1	-445	141	-3.2
August	-784	129	-6.1	-755	129	-5.9
September	-955	148	-6.5	-965	163	-5.9
October	-892	145	-6.2	-902	134	-6.7
November	-1150	151	-7.6	-1156	153	-7.6
December	615	114	5.4	621	153	4.1
Intercept	1218	183	6.6	1898	433	4.4

Notes: All variables except month dummies and bus trip miles are per 100,000 population and standard errors are clustered by neighborhood. January is the excluded category for the time indicator variables. These models are estimated with 1,560 observations across 231 CSA from June, 2020 through January, 2021.

3.6 Conclusions

The results described in this paper are consistent with the hypothesis that bus system utilization in the Los Angeles Metro service area had no impact on COVID transmission during the June 2020 through January 2021 period. Metro mandated wearing masks for all passengers during this period, and our results are consistent with the hypothesis that mask wearing was effective. Of course, we do not know how our results would change if Metro did not require masks for all passengers. Before concluding that bus transit use is not a factor in COVID transmission, it would be useful to carry out careful contract tracing investigations using virus sample genetic sequencing to pinpoint the source of COVID infections. Countries like Taiwan and South Korea have carried out such investigations, and it would certainly be useful to do more in U.S. and other countries.

Approximately 40 percent of the residual variation in our models of COVID transmission is due to the variation in neighborhood factors that do not change between June 2020 and January 2021. We tried several methods to explain these factors using a large set of demographic variables but found essentially no correlations. This implies that the cross-section correlations frequently noted in press reporting (also in Table 3.2) are all due to these correlations not accounting for the dynamics of COVID transmission. Of course, the neighborhoods we are forced to use are quite large, and our aggregate demographic measures may well obscure important effects for some subgroups.

Our analysis could be usefully improved by extending our panel data further over time. We could account for vaccinations by adding them to the 2 and 3 month lagged values of current cases. It would be useful to include other geographically detailed mobility measures in the model to distinguish between cases potentially caused by bus use and other types of mobility. Although we did not find much evidence of heterogeneity related to demographic differences in the unobserved time-invariant panel effects, there may be important heterogeneity in the dynamics of COVID transmission across the neighborhoods.

The Los Angeles Metro bus system lost about one third of its riders from November/December 2019 to November/December 2020. Many of these riders likely left because of fears about catching COVID while riding buses. Our results suggest that these fears might not be justified if Metro's COVID mitigation strategies are followed. While increasing vaccination rates should reduce COVID transmission on buses, the emergence of the Delta variant clearly increases transmission risk. This suggests that it is not prudent to drop mandatory mask policies if COVID is widely circulating in the population served by bus transit.

Bibliography

Aaditya, B. and Rahul, T. (2023). Long-term impacts of COVID-19 pandemic on travel behaviour. *Travel Behaviour and Society*, 30:262–270.

Agarwal, S., Koo, K. M., and Sing, T. F. (2015). Impact of electronic road pricing on real estate prices in singapore. *Journal of Urban Economics*, 90:50–59.

Alonso, W. (1964). *Location and land use toward a general theory of land rent*. Harvard University Press.

Anwari, N., Tawkir Ahmed, M., Rakibul Islam, M., Hadiuzzaman, M., and Amin, S. (2021). Exploring the travel behavior changes caused by the COVID-19 crisis: A case study for a developing country. *Transportation Research Interdisciplinary Perspectives*, 9:100334.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2):277–297.

Arranz, J. M., Burguillo, M., and Rubio, J. (2019). Subsidisation of public transport fares for the young: An impact evaluation analysis for the Madrid Metropolitan Area. *Transport Policy*, 74:84–92.

Axhausen, K. W., Hess, S., König, A., Abay, G., Bates, J. J., and Bierlaire, M. (2008). Income and distance elasticities of values of travel time savings: New swiss results. *Transport Policy*, 15(3):173–185.

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., and Wang, M. (2020). Presumed asymptomatic carrier transmission of covid-19. *JAMA*, 323(14):1406–1407.

Bates, J. J. (1987). Measuring travel time values with a discrete choice model: a note. *The Economic Journal*, 97(386):493–498.

Bhagat, R. K., Wykes, M. D., Dalziel, S. B., and Linden, P. (2020). Effects of ventilation on the indoor spread of covid-19. *Journal of Fluid Mechanics*, 903:F1.

Blocken, B., Van Druenen, T., Ricci, A., Kang, L., Van Hooff, T., Qin, P., Xia, L., Ruiz, C. A., Arts, J., Diepens, J., et al. (2021). Ventilation and air cleaning to limit aerosol particle concentrations in a gym during the covid-19 pandemic. *Building and Environment*, 193:107659.

Blundell, R. and Bond, S. (2000). Gmm estimation with persistent panel data: An application to production functions. *Econometric Reviews*, 19(3):321–340.

Boyd, B., Chow, M., Johnson, R., and Smith, A. (2003). Analysis of effects of fare-free transit program on student commuting mode shares: BruinGo at University of California at Los Angeles. *Transportation Research Record*, 1835(1):101–110.

Bradley, M., Greene, E., Spitz, G., Coogan, M., and McGuckin, N. (2018). The millennial question: Changes in travel behaviour or changes in survey behaviour? *Transportation Research Procedia*, 32:291–300.

Brown, J., Hess, D. B., and Shoup, D. (2003). Fare-free public transit at universities: An evaluation. *Journal of Planning Education and Research*, 23(1):69–82.

Brownstone, D., Bunch, D. S., and Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5):315–338.

Brownstone, D., Ghosh, A., Golob, T. F., Kazimi, C., and Van Amelsfort, D. (2003). Drivers' willingness-to-pay to reduce travel time: evidence from the san diego i-15 congestion pricing project. *Transportation Research Part A: Policy and Practice*, 37(4):373–387.

Brownstone, D. and Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4):279–293.

Brownstone, D. and Train, K. (1998). Forecasting new product penetration with flexible substitution patterns. *Journal of econometrics*, 89(1-2):109–129.

Brueckner, J. K. (2014). Cordon tolling in a city with congested bridges. *Economics of Transportation*, 3(4):235–242.

ChoiceMetrics (2024). *NGENE 1.4 User Manual*. Accessed: 2024-11-10.

Chowkwanyun, M. and Reed, A. L. J. (2020). Racial health disparities and covid-19-caution and context. *New England Journal of Medicine*, 383(3):201–203.

Czeisler, M. E., Tynan, M. A., Howard, M. E., Honeycutt, S., Fulmer, E. B., Kidder, D. P., and et al. (2020). Public attitudes, behaviors, and beliefs related to covid-19, stay-at-home orders, nonessential business closures, and public health guidance—united states, new york city, and los angeles, may 5–12, 2020. *Morbidity and Mortality Weekly Report*, 69(24):751.

De Jong, G., Daly, A., Pieters, M., and Van der Hoorn, T. (2007). The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*, 41(9):874–889.

Diao, M., Leonard, D., and Sing, T. F. (2017). Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values. *Regional Science and Urban Economics*, 67:64–77.

Fathi-Kazerooni, S., Rojas-Cessa, R., Dong, Z., and Umpaichitra, V. (2021). Correlation of subway turnstile entries and covid-19 incidence and deaths in new york city. *Infectious Disease Modelling*, 6:183–194.

Federal Highway Administration (2022). 2022 National Household Travel Survey (NHTS) Data. Accessed: 2024-11-10.

Fesselmeyer, E. and Liu, H. (2018). How much do users value a network expansion? evidence from the public transit system in singapore. *Regional Science and Urban Economics*, 71:46–61.

Fezzi, C., Bateman, I. J., and Ferrini, S. (2014). Using revealed preferences to estimate the value of travel time to recreation sites. *Journal of Environmental Economics and Management*, 67(1):58–70.

Gibson, C. (2023). Why aren't teenagers driving anymore? *The Washington Post*.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Hamidi, S. and Hamidi, I. (2021). Subway ridership, crowding, or population density: Determinants of covid-19 infection rates in new york city. *American Journal of Preventive Medicine*.

Haseeb, A. and Mitra, R. (2024). Travel behaviour changes among post-secondary students after COVID-19 pandemic – A case of Greater Toronto and Hamilton Area, Canada. *Case Studies on Transport Policy*, 17:101245.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *transportation research part B: methodological*, 44(6):735–752.

Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30:133–176.

Hess, S., Rose, J. M., and Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transportation Research Part D: Transport and Environment*, 15(7):405–417.

Ho, T.-H., Png, I. P. L., and Reza, S. (2018). Sunk cost fallacy in driving the world's costliest cars. *Management Science*, 64(4):1761–1778.

Housing and Development Board (n.d.). Public housing – a singapore icon.

Hu, M., Lin, H., Wang, J., Xu, C., Tatem, A. J., Meng, B., Zhang, X., Liu, Y., Wang, P., Wu, G., and Xie, H. (2021). Risk of coronavirus disease 2019 transmission in train passengers: an epidemiological and modeling study. *Clinical Infectious Diseases*, 72(4):604–610.

Hu, S. and Chen, P. (2021). Who left riding transit? examining socioeconomic disparities in the impact of covid-19 on ridership. *Transportation Research Part D: Transport and Environment*, 90:102654.

Lachapelle, U., Manaugh, K., and Hamelin-Pratte, S. (2022). Providing discounted transit passes to younger university students: are there effects on public transit, car and active transportation trips to university? *Case studies on transport policy*, 10(2):811–820.

Lam, T. C. and Small, K. A. (2001). The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review*, 37(2-3):231–251.

Li, Y., Zhang, R., Zhao, J., and Molina, M. J. (2020). Understanding transmission and intervention for the covid-19 pandemic in the united states. *Science of the Total Environment*, 748:141560.

Los Angeles Community College District (n.d.). Fast facts. `https://www.laccd.edu/offices/epie/research/fast-facts`. Accessed: 2024-11-06.

Los Angeles County Metropolitan Transportation Authority (2024a). 2023-0760 - GOPASS PILOT PROGRAM EXTENSION.

Los Angeles County Metropolitan Transportation Authority (2024b). L.A. Metro's GoPass Program Reaches 40 Million Boardings. Accessed: 2024-11-10.

Los Angeles County Metropolitan Transportation Authority (2024c). Metro Ridership Portal. Accessed: 2024-11-10.

Los Angeles Metropolitan County Transportation Authority (n.d.). Fareless System Initiative task force (September 2020 – May 2021).

Louviere, J. J. and Hensher, D. A. (1982). On the design and analysis of simulated choice or allocation experiments in travel choice modelling. *Transportation research record*, 890(1):11–17.

Louviere, J. J. and Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research*, 20(4):350–367.

Lubitow, A., Liévanos, R. S., Carpenter, E., and McGee, J. A. (2021). Transformative transportation survey methods: Enhancing household transportation survey methods for hard-to-reach populations. *Transportation Research Part D: Transport and Environment*, 98:102953.

Lum, S. K. (2002). Market fundamentals, public policy and private gain: house price dynamics in singapore. *Journal of Property Research*, 19(2):121–143.

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470.

McLaren, J. (2020). Racial disparity in covid-19 deaths: Seeking economic roots with census data. Working Paper 27407, National Bureau of Economic Research.

Meginnis, K., Burton, M., Chan, R., and Rigby, D. (2021). Strategic bias in discrete choice experiments. *Journal of Environmental Economics and Management*, 109:102163.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2):197–210.

Mulley, C., Ma, L., Clifton, G., Yen, B., and Burke, M. (2016). Residential property value impacts of proximity to transport infrastructure: An investigation of bus rapid transit and heavy rail networks in brisbane, australia. *Journal of Transport Geography*, 54:41–52.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Muth, R. F. (1969). Cities and housing; the spatial pattern of urban residential land use.

Nash, S. and Mitra, R. (2019). University students' transportation patterns, and the role of neighbourhood types and attitudes. *Journal of Transport Geography*, 76:200–211.

Odeck, J., Rekdal, J., and Hamre, T. N. (2003). The socio-economic benefits of moving from cordon toll to congestion pricing: the case of oslo. In *TRB Annual Meeting*.

Olszewski, P. and Wibowo, S. S. (2005). Using equivalent walking distance to assess pedestrian accessibility to transit stations in singapore. *Transportation research record*, 1927(1):38–45.

Olszewski, P. and Xie, L. (2005). Modelling the effects of road pricing on traffic in singapore. *Transportation Research Part A: Policy and Practice*, 39(7-9):755–772.

Parry, I. W. H. and Bento, A. (2002). Estimating the welfare effect of congestion taxes: The critical importance of other distortions within the transport system. *Journal of Urban Economics*, 51(2):339–365.

Pigou, A. C. (1920). *The economics of welfare*. Palgrave Macmillan.

Qian, H., Miao, T., Liu, L., Zheng, X., Luo, D., and Li, Y. (2020). Indoor transmission of sars-cov-2. *Indoor Air*, 00:1–7.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, 80(4):647–657.

Rogers, T. N., Rogers, C. R., VanSant-Webb, E., Gu, L. Y., Yan, B., and Qeadan, F. (2020). Racial disparities in covid-19 mortality among essential workers in the united states. *World Medical Health Policy*, 12(3):311–327.

Shah, M., Sachdeva, M., and Dodiuk-Gad, R. P. (2020). Covid-19 and racial disparities. *Journal of the American Academy of Dermatology*, 83(1):e35.

Shen, Y., Li, C., Dong, H., Wang, Z., Martinez, L., Sun, Z., Handel, A., Chen, Z., Chen, E., Ebell, M. H., and Wang, F. (2020). Community outbreak investigation of sars-cov-2 transmission among bus riders in eastern china. *JAMA Internal Medicine*, 180(12):1665–1671.

Singapore Department of Statistics (2010). *Census of Population 2010 Advance Release*.

Standen, C., Greaves, S., Collins, A. T., Crane, M., and Rissel, C. (2019). The value of slow travel: Economic appraisal of cycling projects using the logsum measure of consumer surplus. *Transportation research part A: policy and practice*, 123:255–268.

Steimetz, S. S. and Brownstone, D. (2005). Estimating commuters' "value of time" with noisy data: a multiple imputation approach. *Transportation Research Part B: Methodological*, 39(10):865–889.

Sun, C. and Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing covid-19 transmission. *Sustainable Cities and Society*, 62:102390.

Sælensminde, K. (2006). Causes and consequences of lexicographic choices in stated choice studies. *Ecological Economics*, 59(3):331–340.

Tabasi, M., Siripanich, A., Khan, N. A., Auld, J., and Rashidi, T. H. (2024). Gps-supported smartphone app-based integrated travel diary and time-use data collection: challenges and lessons learned. *Transportation*, pages 1–26.

Train, K. and Weeks, M. (2005). *Discrete choice models in preference space and willingness-to-pay space*. Springer.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Truong, T. P. and Hensher, D. A. (1985). Measurement of travel time values and opportunity cost from a discrete-choice model. *The Economic Journal*, 95(378):438–451.

Tsoleridis, P., Choudhury, C. F., and Hess, S. (2022). Deriving transport appraisal values from emerging revealed preference data. *Transportation Research Part A: Policy and Practice*, 165:225–245.

Tveter, E. (2023). The value of travel time: a revealed preferences approach using exogenous variation in travel costs and automatic traffic count data. *Transportation*, 50(6):2273–2297.

Vickrey, W. S. (1963). Pricing in urban and suburban transport. *The American Economic Review*, 53(2):452–465.

Wang, K. (2014). How change in public transportation usage reveals fear of the sars virus in a city. *PLoS ONE*, 9(3):e89405.

Weber, S. (2021). A step-by-step procedure to implement discrete choice experiments in qualtrics. *Social Science Computer Review*, 39(5):903–921.

Wei, L. and Wehby, G. L. (2020). Community use of face masks and covid-19: Evidence from a natural experiment of state mandates in the us. *Health Affairs*, 39(8):1419–1425.

Wilson, P. W. (1988). Welfare effects of congestion pricing in singapore. *Transportation*, 15(3):191–210.

Wu, J., Xu, F., Zhou, W., Feikin, D. R., Lin, C.-Y., He, X., Zhu, Z., Liang, W., Chin, D. P., and Schuchat, A. (2004). Risk factors for sars among persons without known contact with sars patients, beijing, china. *Emerging infectious diseases*, 10(2):210.

Yu, P., Zhu, J., Zhang, Z., and Han, Y. (2020). A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. *The Journal of Infectious Diseases*, 221(11):1757–1761.

Appendix A

Chapter 2 Appendix

A.1 DCE Supplementary Materials

A.1.1 DCE Scripts

Introduction:

The next 3 questions ask you to choose an option for taking a trip. These options might not currently exist or seem realistic, but they might exist in the future. The travel time includes the wait time (which is time spent not moving), and the trip cost includes parking costs. "Self-driving Car" is like Uber or Lyft but without a human driver. "Metro Micro" is like Uber Pool and is an on-demand shared ride service for short trips. Please choose the best among the 3 options for each trip.

DCE Task Prompts:

Which option do you prefer for a 12 mile trip?

(display first set of DCE alternatives)

Which option do you prefer for a 7 mile trip?

(display second set of DCE alternatives)

Which option do you prefer for a 2 mile trip?

(display third set of DCE alternatives)

Would you prefer to walk instead of the travel method you chose for the last hypothetical trip?

Yes	No

Full DCE Design

Table A.1: List of alternatives, alternatives, and blocks used in the discrete choice experiment design. For WiFi, 0 = not available and 1 = available.

Scenario	Block	Task	Option	Mode	Travel time	Travel cost	Wait time	WiFi
1	1	4	1	Train	40	1	5	0
1	1	4	2	E-Bike or scooter rental	70	4	5	0
1	1	4	3	Car	50	5	0	0
2	1	30	1	E-Bike or scooter rental	40	4	5	0
2	1	30	2	Bus	50	1	5	1
2	1	30	3	Train	30	1	15	0
3	1	67	1	Self-driving Car	20	10	10	0
3	1	67	2	Bus	12	0	5	1
3	1	67	3	Metro Micro	15	1	5	0
1	2	5	1	E-Bike or scooter rental	60	5	0	0
1	2	5	2	Self-driving Car	50	10	15	0
1	2	5	3	Train	55	0	10	1
2	2	56	1	Bus	30	0	10	1
2	2	56	2	Train	50	2	5	0
2	2	56	3	E-Bike or scooter rental	50	4	10	0
3	2	61	1	Bus	17	2	10	0
3	2	61	2	Self-driving Car	20	10	5	1
3	2	61	3	E-Bike or scooter rental	10	2	5	0
1	3	32	1	Train	70	1	15	1
1	3	32	2	Self-driving Car	40	5	5	0
1	3	32	3	Bus	40	2	10	0
2	3	36	1	Train	30	1	10	1
2	3	36	2	Bus	50	0	10	0
2	3	36	3	E-Bike or scooter rental	40	5	5	0
3	3	52	1	Bus	17	1	5	0
3	3	52	2	Self-driving Car	10	10	5	1
3	3	52	3	Metro Micro	15	0	5	1
1	4	1	1	Car	40	5	0	0
1	4	1	2	Bus	70	2	15	0
1	4	1	3	Train	40	1	5	1
2	4	25	1	E-Bike or scooter rental	50	4	10	0
2	4	25	2	Train	30	2	15	1
2	4	25	3	Bus	50	0	5	0
3	4	45	1	Metro Micro	10	0	5	1
3	4	45	2	Bus	17	1	5	0
3	4	45	3	Car	30	10	0	0
1	5	15	1	Self-driving Car	50	10	5	1
1	5	15	2	E-Bike or scooter rental	70	3	5	0
1	5	15	3	Bus	40	2	15	0
2	5	27	1	Self-driving Car	45	4	15	1

Table A.1 continued from previous page

Scenario	Block	Task	Option	Mode	Travel time	Travel cost	Wait time	WiFi
2	5	27	2	Train	30	0	5	0
2	5	27	3	E-Bike or scooter rental	40	5	5	0
3	5	53	1	Metro Micro	15	2	10	1
3	5	53	2	Self-driving Car	20	7	5	0
3	5	53	3	Bus	12	0	5	1
1	6	19	1	Car	30	15	0	0
1	6	19	2	Train	40	1	15	0
1	6	19	3	Bus	70	0	5	1
2	6	40	1	Self-driving Car	35	4	15	1
2	6	40	2	Bus	50	1	15	0
2	6	40	3	Train	30	2	5	0
3	6	66	1	Metro Micro	15	2	10	0
3	6	66	2	Bus	12	0	10	1
3	6	66	3	Self-driving Car	20	10	10	1
1	7	23	1	Train	70	2	5	1
1	7	23	2	Bus	55	1	15	0
1	7	23	3	E-Bike or scooter rental	60	3	10	0
2	7	33	1	Train	50	2	5	0
2	7	33	2	E-Bike or scooter rental	30	4	10	0
2	7	33	3	Self-driving Car	35	4	10	0
3	7	69	1	E-Bike or scooter rental	20	3	0	0
3	7	69	2	Self-driving Car	30	7	10	1
3	7	69	3	Metro Micro	10	1	5	0
1	8	8	1	Self-driving Car	30	10	15	0
1	8	8	2	Car	30	5	0	0
1	8	8	3	Bus	70	2	10	1
2	8	29	1	Bus	50	1	10	1
2	8	29	2	Train	40	2	5	1
2	8	29	3	Self-driving Car	25	4	15	0
3	8	63	1	E-Bike or scooter rental	10	2	0	0
3	8	63	2	Self-driving Car	20	7	5	1
3	8	63	3	Metro Micro	25	2	15	1
1	9	22	1	Train	40	2	5	1
1	9	22	2	Self-driving Car	50	5	10	0
1	9	22	3	Bus	55	1	15	0
2	9	47	1	Self-driving Car	35	7	15	0
2	9	47	2	E-Bike or scooter rental	50	3	0	0
2	9	47	3	Bus	30	2	10	1
3	9	62	1	Self-driving Car	10	4	5	1
3	9	62	2	Bus	7	2	5	0
3	9	62	3	E-Bike or scooter rental	30	3	0	0
1	10	55	1	E-Bike or scooter rental	70	3	0	0
1	10	55	2	Train	40	2	15	1
1	10	55	3	Self-driving Car	50	10	5	1
2	10	57	1	Train	50	1	15	1
2	10	57	2	Bus	50	2	5	0

Table A.1 continued from previous page

Scenario	Block	Task	Option	Mode	Travel time	Travel cost	Wait time	WiFi
2	10	57	3	Self-driving Car	25	7	10	0
3	10	64	1	Self-driving Car	20	7	5	0
3	10	64	2	Bus	7	0	5	1
3	10	64	3	Metro Micro	25	2	15	0
1	11	6	1	E-Bike or scooter rental	60	5	10	0
1	11	6	2	Bus	70	0	5	0
1	11	6	3	Train	55	0	15	1
2	11	43	1	Bus	50	2	15	1
2	11	43	2	Self-driving Car	25	7	10	0
2	11	43	3	E-Bike or scooter rental	50	4	0	0
3	11	60	1	Bus	12	0	5	1
3	11	60	2	Self-driving Car	10	10	5	0
3	11	60	3	Metro Micro	25	1	10	1
1	12	31	1	Bus	40	0	10	0
1	12	31	2	Train	55	2	10	1
1	12	31	3	Self-driving Car	50	5	5	0
2	12	44	1	Car	35	7	0	0
2	12	44	2	Train	40	0	15	0
2	12	44	3	Self-driving Car	45	10	5	1
3	12	48	1	Metro Micro	15	0	5	1
3	12	48	2	Bus	7	2	5	1
3	12	48	3	Self-driving Car	30	7	15	0
1	13	2	1	Bus	55	2	5	1
1	13	2	2	Car	40	10	0	0
1	13	2	3	Self-driving Car	40	5	15	0
2	13	59	1	Self-driving Car	25	4	10	0
2	13	59	2	Car	45	7	0	0
2	13	59	3	Bus	40	2	15	1
3	13	70	1	Metro Micro	25	1	5	0
3	13	70	2	Bus	7	1	5	1
3	13	70	3	E-Bike or scooter rental	10	3	5	0
1	14	26	1	Bus	70	0	15	1
1	14	26	2	E-Bike or scooter rental	70	5	10	0
1	14	26	3	Self-driving Car	30	10	5	0
2	14	28	1	Self-driving Car	25	4	10	0
2	14	28	2	Train	40	2	15	1
2	14	28	3	Bus	50	2	5	1
3	14	42	1	E-Bike or scooter rental	30	2	10	0
3	14	42	2	Self-driving Car	20	10	10	1
3	14	42	3	Bus	7	2	5	0
1	15	12	1	Self-driving Car	50	15	15	0
1	15	12	2	Bus	40	1	10	1
1	15	12	3	Train	55	0	5	0
2	15	49	1	E-Bike or scooter rental	40	5	5	0
2	15	49	2	Self-driving Car	35	4	15	1
2	15	49	3	Bus	40	1	5	0

Table A.1 continued from previous page

Scenario	Block	Task	Option	Mode	Travel time	Travel cost	Wait time	WiFi
3	15	58	1	Bus	7	1	5	0
3	15	58	2	Metro Micro	10	1	5	1
3	15	58	3	Self-driving Car	30	7	10	1
1	16	10	1	Train	70	0	10	0
1	16	10	2	Bus	40	2	15	1
1	16	10	3	E-Bike or scooter rental	70	3	0	0
2	16	38	1	Train	50	1	15	0
2	16	38	2	Bus	30	0	5	1
2	16	38	3	Self-driving Car	35	10	15	1
3	16	72	1	Bus	17	2	15	0
3	16	72	2	Metro Micro	15	0	5	1
3	16	72	3	Self-driving Car	10	7	5	0
1	17	7	1	Bus	40	0	5	0
1	17	7	2	E-Bike or scooter rental	60	4	0	0
1	17	7	3	Self-driving Car	50	15	15	1
2	17	14	1	Car	45	7	0	0
2	17	14	2	Self-driving Car	35	4	15	0
2	17	14	3	Train	30	2	5	1
3	17	54	1	Self-driving Car	30	4	5	1
3	17	54	2	Bus	7	2	5	0
3	17	54	3	Metro Micro	15	1	5	1
1	18	9	1	Bus	55	0	5	1
1	18	9	2	Train	70	2	5	0
1	18	9	3	Self-driving Car	30	5	15	1
2	18	17	1	Bus	40	0	5	0
2	18	17	2	E-Bike or scooter rental	30	4	5	0
2	18	17	3	Self-driving Car	45	10	10	1
3	18	21	1	E-Bike or scooter rental	20	4	5	0
3	18	21	2	Metro Micro	10	1	5	1
3	18	21	3	Bus	17	0	5	0
1	19	13	1	Self-driving Car	50	10	10	1
1	19	13	2	Train	55	0	10	0
1	19	13	3	Car	30	15	0	0
2	19	24	1	Bus	30	2	5	1
2	19	24	2	Self-driving Car	45	4	15	0
2	19	24	3	Train	40	1	10	1
3	19	34	1	E-Bike or scooter rental	20	4	5	0
3	19	34	2	Self-driving Car	30	4	15	1
3	19	34	3	Metro Micro	15	1	10	0
1	20	20	1	Train	70	0	15	0
1	20	20	2	Self-driving Car	30	15	5	1
1	20	20	3	Car	40	10	0	0
2	20	39	1	E-Bike or scooter rental	50	4	10	0
2	20	39	2	Self-driving Car	35	10	5	0
2	20	39	3	Bus	40	0	5	1
3	20	41	1	Self-driving Car	10	10	5	0

Table A.1 continued from previous page

Scenario	Block	Task	Option	Mode	Travel time	Travel cost	Wait time	WiFi
3	20	41	2	Metro Micro	15	0	5	1
3	20	41	3	Bus	17	0	10	1
1	21	16	1	Bus	55	1	10	1
1	21	16	2	Train	55	2	15	0
1	21	16	3	E-Bike or scooter rental	60	3	0	0
2	21	35	1	Self-driving Car	35	10	10	1
2	21	35	2	E-Bike or scooter rental	50	4	0	0
2	21	35	3	Bus	30	0	15	0
3	21	65	1	Metro Micro	10	1	5	0
3	21	65	2	Self-driving Car	30	10	10	1
3	21	65	3	Bus	12	0	5	1
1	22	11	1	Bus	55	2	15	0
1	22	11	2	Train	70	0	5	1
1	22	11	3	Self-driving Car	30	10	5	0
2	22	37	1	Self-driving Car	25	10	10	1
2	22	37	2	Train	50	1	10	0
2	22	37	3	E-Bike or scooter rental	40	3	10	0
3	22	68	1	Metro Micro	15	0	10	1
3	22	68	2	E-Bike or scooter rental	30	4	0	0
3	22	68	3	Bus	7	2	5	0
1	23	18	1	Self-driving Car	30	5	5	0
1	23	18	2	Train	70	2	10	0
1	23	18	3	E-Bike or scooter rental	70	4	10	0
2	23	50	1	Self-driving Car	25	7	5	1
2	23	50	2	E-Bike or scooter rental	40	3	10	0
2	23	50	3	Train	50	2	10	0
3	23	71	1	Metro Micro	25	1	15	0
3	23	71	2	Car	10	4	0	0
3	23	71	3	Self-driving Car	10	10	5	1
1	24	3	1	Train	40	0	5	1
1	24	3	2	E-Bike or scooter rental	80	4	10	0
1	24	3	3	Self-driving Car	50	15	5	0
2	24	46	1	Bus	30	2	15	0
2	24	46	2	Self-driving Car	25	4	5	1
2	24	46	3	Train	50	1	5	0
3	24	51	1	Self-driving Car	20	10	15	1
3	24	51	2	E-Bike or scooter rental	30	3	5	0
3	24	51	3	Metro Micro	10	0	5	0

A.2 Supplementary Tests and Results

A.2.1 Test for MMNL Model Appropriateness

It is also important to formally test whether mixing is appropriate for the travel time parameters. McFadden and Train (2000) lay out a simple procedure using a Lagrange multiplier test from estimating the MNL model, which facilitates computation time. The procedure starts with evaluating the choice probabilities from the MNL estimation, given in Equation (2.2), which can be denoted L_C for individual n's choice set, C[1]. To further facilitate notation, suppose that the travel time variables (which are separate by scenario) form a component t of the attribute-specific variables contained in X_{ni}, and are denoted x_{it} for ease of notation. Finally, artificial variables are created using the following formula:

$$z_{ti} = (x_{ti} - x_{tC})^2, \text{where } x_{tc} = \sum_{j \in C} x_{tj} \cdot L_C(j|X, \hat{\beta}). \tag{A.1}$$

The artificial variables, z_{ti} constructed from scenario-specific travel are then added to the MNL specification and refit. A Wald test or likelihood ratio test may then be used to determine if the estimated coefficients on the artificial variables are jointly significant, which is tested against the null hypothesis of no mixing among the model parameters in question. These results are presented in Table A.2, and for brevity, only the estimates of the coefficients on the artificial variables are shown.

[1]This procedure is being conducted within the choice scenario k, so this notation is dropped.

Table A.2: MNL coefficient estimates on artificial variables and joint significance tests.

Artificial Variable	β	SE
$z_{TT,\text{ long}}$	0.001	(0.0004)
$z_{TT,\text{ med}}$	0.002	(0.001)
$z_{TT,\text{ short}}$	0.003	(0.001)
Log pseudolikelihood	-3291.621	
Wald test χ^2 (d. f.)	12.51**	(3)

Individually, two coefficients are statistically significant: the coefficient on the artificial variable constructed from the short scenario travel times and the same variable corresponding to long trip scenario travel times. However, the coefficient on the artificial variable corresponding to the medium trip scenario travel time can be considered significant at the 10% level ($p = 0.087$). Importantly, the resulting χ^2 test statistic from the Wald test is significant at the 1% level, suggesting evidence of joint significance. This result, combined with the statistical significance of the standard deviation of the random travel time coefficients in Table 2.5, suggest there is evidence that mixing is present.

A.2.2 Supplementary Results

Table A.3: Regression results for the multinomial logit and mixed multinomial logit models, after excluding straight-lining responses.

Variable	**MNL**		**MMNL**	
	β	SE	β	SE
Travel Attributes				
Travel time (long trip)	-0.017	(0.004)	-0.024	(0.005)
Travel time (medium trip)	-0.039	(0.005)	-0.042	(0.005)

Travel time (short trip)	-0.021	(0.006)	-0.018	(0.007)
Travel cost	-0.046	(0.013)	-0.048	(0.014)
Wait time	-0.010	(0.006)	-0.011	(0.007)
WiFi availability	0.245	(0.049)	0.264	(0.052)
$\sigma_{\text{TravelTime}}$	—	—	0.053	(0.009)
Individual Variables and ASCs (Base = Car)				
Bus				
Alternative-specific constant	-1.357	(0.269)	-1.435	(0.287)
Medium trip	0.788	(0.306)	0.854	(0.326)
Short trip	0.871	(0.326)	0.811	(0.346)
Has GoPass	0.537	(0.238)	0.582	(0.255)
Driver's license	-1.027	(0.232)	-1.103	(0.250)
E-bike or scooter rental				
Alternative-specific constant	-3.717	(0.388)	-3.840	(0.403)
Medium trip	1.471	(0.422)	1.564	(0.439)
Short trip	2.192	(0.449)	2.193	(0.464)
Has GoPass	-0.237	(0.303)	-0.227	(0.318)
Driver's license	-0.212	(0.294)	-0.265	(0.308)
Metro Micro				
Alternative-specific constant	-1.145	(0.432)	-1.157	(0.447)
Has GoPass	0.160	(0.283)	0.184	(0.298)
Driver's license	-0.346	(0.276)	-0.384	(0.291)
Self-driving Car				
Alternative-specific constant	-0.538	(0.237)	-1.157	(0.263)
Medium trip	0.014	(0.305)	0.018	(0.324)
Short trip	0.571	(0.314)	0.526	(0.332)
Has GoPass	-0.538	(0.237)	-0.545	(0.253)
Driver's license	0.186	(0.229)	0.189	(0.244)
Train				

Alternative-specific constant	-1.428	(0.274)	-1.550	(0.292)
Medium trip	0.334	(0.307)	0.373	(0.325)
Has GoPass	0.305	(0.243)	0.334	(0.260)
Driver's license	-0.771	(0.238)	-0.839	(0.256)
Walk				
Alternative-specific constant	0.660	(0.368)	0.460	(0.399)
Has GoPass	-0.248	(0.258)	-0.281	(0.292)
Driver's license	-0.108	(0.258)	-0.092	(0.284)
Number of individuals (N):	874		874	
Number of choice situations:	3,092		3,092	
Number of total observations:	10,158		10,158	
Log pseudolikelihood (Simulated for MMNL):	-2959.690		-2953.107	

Table A.4: Comparison of aggregate consumer surplus (in thousands of $) changes of different policy changes across trip scenarios, shown by LACCD campus.

College	Long	Medium	Short
Panel A: GoPass ($0 transit fare)			
Los Angeles Pierce College	13.5	16.7	4.8
Los Angeles Mission College	8.4	11.4	3.1
Los Angeles Valley College	15.2	13.8	4.8
Los Angeles City College	14.4	15.3	5.3
Los Angeles Trade Tech College	9.4	10.0	2.9
Los Angeles Southwest College	5.7	6.2	0.9
Los Angeles Harbor College	7.5	6.4	2.8
East Los Angeles College	28.8	31.0	10.2
West Los Angeles College	10.2	11.5	3.7
All campuses	**112.8**	**122.5**	**38.6**
Panel B: 5% travel time reduction			
Los Angeles Pierce College	61.2	67.2	4.7
Los Angeles Mission College	40.0	44.2	2.8
Los Angeles Valley College	62.6	68.4	5.1
Los Angeles City College	63.9	71.6	5.1
Los Angeles Trade Tech College	45.7	48.1	3.7
Los Angeles Southwest College	25.8	29.5	1.9
Los Angeles Harbor College	28.7	30.6	2.4
East Los Angeles College	124.5	138.3	9.4
West Los Angeles College	49.0	51.7	4.1

All campuses	**501.1**	**549.1**	**39.1**
Panel C: $5 congestion charge			
Los Angeles Pierce College	-43.4	–11.9	–9.4
Los Angeles Mission College	-24.6	–6.2	–9.2
Los Angeles Valley College	–35.7	–5.1	–11.5
Los Angeles City College	–34.1	–10.3	–16.7
Los Angeles Trade Tech College	–28.7	–5.9	–5.9
Los Angeles Southwest College	–19.5	–1.3	–2.6
Los Angeles Harbor College	–11.6	–6.8	–2.9
East Los Angeles College	–66.1	–27.6	–16.8
West Los Angeles College	–34.2	–9.8	–8.2
All campuses	**–297.0**	**–84.6**	**–84.6**

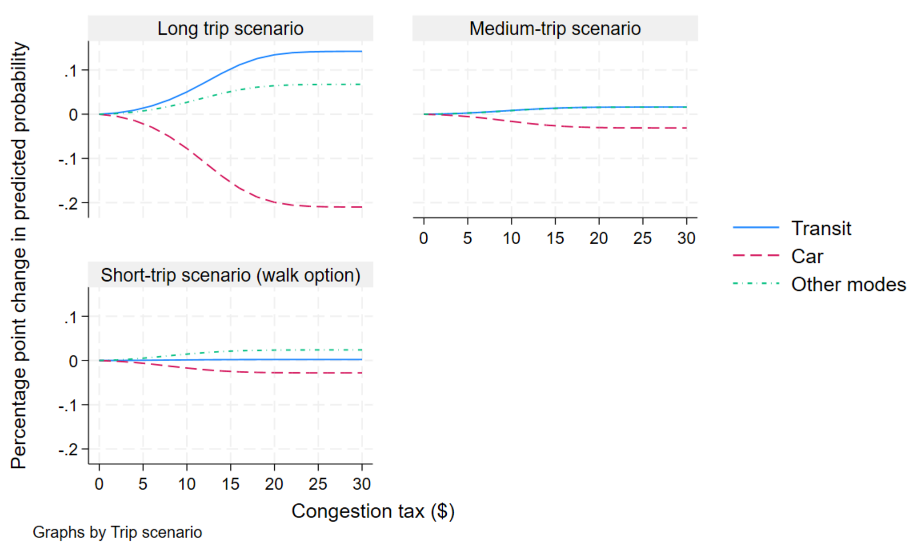


Figure A.1: Congestion tax scenario with tax expanded to $30.

Appendix B

Chapter 3 Appendix

B.1 Technical Description of Bus Ridership Imputations

To estimate ridership in each CSA for months where we did not directly obtain stop-level ridership, we first compute the share of riders in CSA i conditional on commuting on line l, from the most recent stoplevel data, which is provided for month t^*. This share will be denoted $s_{(i|l)t^*}$. Then, we use monthly perline ridership data to infer ridership specific to each line and CSA, which will be denoted B_{il}. Finally, to obtain overall CSA-level ridership, the sum of all values B_{il} is taken across all lines l passing through CSA i.

$$S_{(i|l)t^*} = \frac{B_{ilt^*}}{B_{lt^*}} \tag{B.1}$$

$$B_{it} = \sum_l B_{lt} S_{(i|l)t^*} \tag{B.2}$$

Equation (B.1) shows how the share of riders in CSA i that comprise line l is computed, using the most recent month t^*, before month t, for which ridership B_{it} is known. Equation (B.2) shows how known values of B_{lt} are multiplied by $s_{(i|l)t^*}$ and then added across all lines l to give the total average weekly boardings for month t and CSA i.

Next, we impute a measure of average passenger-miles to use as a proxy for exposure to COVID while riding transit using a few assumptions, which are mostly related to data constraints. Because passenger-miles are available for each line, we assume that the distribution of passenger-miles within each line is independent of the CSA i where riders are boarding and alighting. We will denote linespecific average passenger-miles for month $t, \overline{x_{\mathrm{lt}}}$. Unlike for imputing bus ridership, to impute the passenger-miles, we obtain the share of CSA ridership hailing from line l, which we denote $s_{(l|i)t}$ * Doing this allows us to compute imputed passenger-miles as follows:

$$\overline{x_{lt}} = \sum_{l} \overline{x_{(l|l)t}} s_{(l|i)t^*} = \sum_{l} \overline{x_{lt}} s_{(l|i)t^*} \tag{B.3}$$

Equation (B.3) shows that by assuming $\overline{x_{(l|l)t}} = \overline{x_{lt}}$, the computation of $\overline{x_{lt}}$ turns into a simple expected value computation using the conditional estimated values of the probability of a rider from CSA i choosing to commute on line l, using the most recent data from month t^*. Assuming that passenger-miles on line l are independent of the CSA where riders board and alight would not be required if the conditional values $\overline{x_{(l|l)t}}$ were known.