ABSTRACT

Title of Dissertation: THE BENEFITS OF METRO RAIL IN MUMBAI, INDIA: REDUCED FORM AND STRUCTURAL APPROACHES

Palak Suri
Doctor of Philosophy, 2022

Dissertation Directed by: Professor Maureen Cropper
Department of Economics

This dissertation studies the welfare effects of introducing metro rail in the city of Mumbai, India using a combination of reduced form and structural econometric approaches. Mumbai is one of the most densely populated cities in the world. To supplement its extensive, but overcrowded, network of Suburban Railway, over 300 km of metro rail lines are planned. Each chapter in this dissertation looks at a different dimension of the benefits of metro rail in the city and is a standalone paper.

In Chapter 1, I analyze the benefits of the introduction of metro rail in Mumbai by computing the value of travel time savings to households. I estimate preferences for commute time from residence to work using two discrete choice models: a commute mode choice model assuming fixed residence and work locations for short-term analysis, and a combined housing and commute mode choice model assuming fixed work location for medium-term analysis. Using the expected compensating variation measure, I value travel time savings due to: (i) Line 1 (11.4 km), ope-

rational since 2014, and (ii) three upcoming lines (92 km). The value of short-term benefits for an average beneficiary under either project is Rs. 71-99 per month (9-14% of the average out-of-pocket cost). The medium-term benefits are an order of magnitude higher than the short-term benefits due to the possibility of household re-sorting. Women, college educated workers, and high-income households receive greater benefits. Benefits of the upcoming network accrue to more individuals and are more dispersed, both spatially and demographically, than the benefits of Line 1. A limitation of the partial equilibrium models in this chapter is that they capture benefits only to households and only through the channel of time savings. In Chapter 2, I address this by studying the net benefits of Metro Line 1.

In Chapter 2 (co-authored with Maureen Cropper), we study the impact of Metro Line 1 in Mumbai on property prices using difference-in-differences in an event study framework. We use administrative data on assessed land values from 2011-18 for 726 sub-zones in the city. Comparing areas within 1 km of the metro with those beyond 1 km but within 3 km, we estimate the effects on property values for commercial, industrial, and residential properties. We find a significant and persistent increase in prices for all land use categories in the treated areas relative to the control areas after Metro Line 1. The price increase ranges from 13% for commercial properties to 17% for residential. We show that improvements in employment accessibility and other location amenities are plausible mechanisms underlying these effects.

In Chapter 3 (co-authored with Maureen Cropper), we study the effects of the introduction of Metro Line 1 in Mumbai on air pollution. We use data on daily average levels of nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and particulate matter (PM_{10}) from ground monitoring stations in an event study framework to identify the changes in pollution levels following the opening of Metro Line 1. We find a robust and significant reduction in the level of NO_2 and no evidence of changes in PM_{10} and SO_2. We also find a decline in the level of Aerosol Optical Depth measured using satellite data at 1 km resolution.

THE BENEFITS OF METRO RAIL IN MUMBAI, INDIA: REDUCED FORM AND STRUCTURAL APPROACHES

by

Palak Suri

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Maureen Cropper, Chair
Professor Sebastian Galiani
Professor Guido Kuersteiner
Associate Professor Joshua Linn
Professor Sergio Urzúa

Foreword

Chapters 2 and 3 of this dissertation contain work that is jointly authored with Maureen Cropper. I contributed substantially to this research and have followed all university guidelines with respect to the inclusion of jointly authored material in my dissertation.

Dedication

To my parents.

Acknowledgments

I am deeply indebted to my dissertation chair and advisor, Maureen Cropper, without whose support this dissertation would not have been possible. I am grateful to her for introducing me to the field of urban economics and for so generously sharing her knowledge and resources with me. Working with her has enabled me to understand the most significant aspects of research as well as life. Her unconditional kindness and belief in my abilities were instrumental in keeping me motivated and energized throughout the process. I am fortunate to have her as a mentor and a friend.

I would like to express my gratitude to Guido Kuersteiner and Sergio Urzúa for their feedback and discussions over the years. Their guidance has allowed me to confidently explore and work with complex econometric methods, which has also shaped my research interests. I am thankful to Sebastian Galiani, Jessica Goldberg, Judy Hellerstein, and Peter Murrell for training me. I also thank Josh Linn for his thoughtful comments on my dissertation.

I am thankful to Vickie Fletcher and Terry Davis for helping me tackle daunting administrative tasks, and for feeding into my optimism from time to time. Thanks also to Jianguo Ma for facilitating access to ArcGIS Pro and promptly responding to my queries. I am thankful to Yi Jiang, Rana Hasan, Marjorie Villanueva Remolador, and Jade Laranjo for their invaluable time and help during the early phases of this project. I thank the World Bank for sharing the household survey data used in Chapters 1 and 2 of this dissertation and the Asian Development Bank for sharing relevant travel time data for Mumbai used in Chapter 1.

I would like to thank my colleagues and friends for their undying support throughout the process. Thanks to Claire, Danny, John, Matt, Mrin, Prateik, Tomo, and Rosa for being there. Thanks also to my friends Dushyant, Gen, Prasad, Rohit, and Tais for their effort in keeping my social life active. Special thanks to my childhood best friend, Aastha, for her consistent effort in helping me stay sane.

I would like to thank Misty Heggeness for forcing me to think about the big picture in research and life. I am grateful to have her as a co-author and a friend.

Most importantly, I would like to express my loving gratitude to my family, especially my nieces, whose presence made everything easier. They provided research and emotional support.

Lastly, I would like to acknowledge all the exogenous shocks that made this dissertation possible.

Table of Contents

List of Tables

List of Figures

Chapter 1: Public Transit Infrastructure and Employment Accessibility: The Benefits of the Mumbai Metro

1.1 Introduction

Urban transit projects have important effects on the spatial structure of a city's economy, and consequently on its growth and development. For example, the introduction of the steam railway in 19th century London allowed workers to live away from their jobs, thereby promoting agglomeration economies and spurring the growth of modern London (Heblich et al. (2020)). Similarly, the introduction of Bus Rapid Transit in Bogotá, Colombia (Tsivanidis (2019)) and Buenos Aires, Argentina (Warnes (2020)) by improving access between workers and firms, generated significant welfare gains. What will be the benefits of building 100 km of Metro rail in Mumbai, which doubles the city's existing rail network?

Mumbai, the financial capital of India, has had an extensive passenger rail network since the 19th century, but its transit infrastructure has struggled to keep up with the demands imposed by the city's growing economy. The existing rail and bus infrastructure is old and over-crowded, and private vehicle ownership has increased four-fold in the last two decades. This has resulted in severe traffic congestion problems and consequently constrained intra-city commutes. In 2019, the median residence-work commute in the city was only 2.5 km. To improve intra-city commuting and alleviate congestion, over 300 km of Metro rail lines have been planned to supplement the historic railway network. The first part of the Metro network, an 11.4 km east-west link (Line 1) opened in 2014. Ninety-two additional km of Metro (Lines 2, 3 and 7) are scheduled to open in 2022. In this paper, I study the benefits of these two infrastructure projects (shown in Figure 1.1).

I use two approaches to measuring the benefits of Metro rail to households in Mumbai: (i) a commute mode choice model, and (ii) a combined housing and commute mode choice model. The commute mode choice model, which holds residence and workplace locations fixed, yields an estimate of the short-term benefits due to reductions in travel time by rail. The combined housing and commute mode choice model, by allowing households to re-sort across the city, yields an estimate of the medium-term benefits due to reductions in travel time by rail. Commute mode choice models have been traditionally used to estimate the value of time as a measure of direct commuting-related welfare due to transport projects (Small et al. (2007)). Recent literature has explored the use of spatial general equilibrium models to estimate the welfare effects on workers and firms (Heblich et al. (2020), Severen (2021), Tsivanidis (2019), Warnes (2020), Khanna et al. (2021)).[1] The combined housing and mode choice model serves as an intermediate approach by allowing adjustment along the housing margin.

My approach to valuing the time savings benefits of the Mumbai Metro allows me to answer four questions: (1) Relative to a traditional commute mode choice model, how much larger is the value of time savings due to the Metro projects when households are allowed to adjust their housing locations? (2) How do the time saving benefits vary by gender, education

[1]These models account for changes in land prices, wages and amenities that accompany improved commuting access between workers and firms.

and income levels of workers? (3) How do the benefits due to the two Metro projects compare? (4) How do these benefits compare with the construction cost of the Metro? These questions can inform transit policy debates by highlighting the trade-offs between a widely accessible transit network and a smaller one placed close to a commercial center in terms of the spatial and demographic distribution of benefits and the construction cost-recovery in the short vs long run. This paper adds to the limited literature focused on evaluating the nature of benefits of transit policies, especially their distributional effects. It also adds to the literature on the value of time by providing preference estimates for in-vehicle and out-of-vehicle times for a developing country using well-measured travel time information.

Using the commute mode choice model, I estimate preferences for in-vehicle time, out-of-vehicle time and out-of-pocket costs accounting for average unobserved mode-category-specific preferences. Data for estimation of the model come from a representative household survey of travel demand conducted by the World Bank in January-February 2019. The survey provides information on residential and workplace locations, worker characteristics and usual commute modes. I implement a network program to compute travel time along the shortest path for rail and walking options. I use travel time for driving trips from Google Maps API to account for traffic conditions, and combine data from HERE Transit API and Google Maps API to obtain travel time for bus trips. To measure the benefits of the Metro, I calculate changes in rail travel time for the following counterfactual scenarios: (i) removing Line 1 from the existing network, and (ii) adding Lines 2, 3 and 7 to the existing system, and value them using Hicksian expected compensating variation (Small and Rosen (1981)). I study the demographic heterogeneity in these benefits by estimating the model separately for sub-samples of commuters by gender, education and income levels.

In the combined housing and mode choice model, which holds the work location of the primary worker fixed, household utility is a function of proximity to work location via different travel modes, observed housing amenities, housing cost and average unobserved utility for houses and modes. I estimate this model using the same household survey data along with travel time information for each potential residence-work commute trip by rail, motorized road transportation and walking, calculated using the network program. Using estimated preferences and counterfactual travel time savings by rail for each potential residence-work commute, I compute expected compensating variation to value the medium-term time savings benefits due to the two Metro projects. I also summarize the spatial and demographic heterogeneity in these benefits. I compare benefits based on this model with those obtained from the commute mode choice model to highlight the influence of the fixed residence assumption.

The short-term benefits implied by the commute mode choice model accrue to only those workers whose commute time between their existing residence and work location is reduced. This could be due to newly accessible Metro stations, or more efficient transit routes. I find that the out-of-vehicle time savings are greater than the in-vehicle time savings under both Metro projects and that individuals have a greater distaste for out-of-vehicle time than for in-vehicle time. This implies that most of the observed benefits are due to out-of-vehicle time savings. The mean expected compensating variation conditional on benefits being positive is Rs. 71-99 per month (9-14% of the average out-of-pocket commuting costs) for each project. This implies that the average beneficiary would be willing to pay 9-14% more than their current out-of-pocket expenditure for these time savings benefits.

In the combined housing and mode choice model, since households can change residential location, time savings benefits accrue via improved accessibility between the household's primary worker's work location and any house that the household has a positive probability of selecting. The mean expected compensating variation implied by this model is Rs. 189 per month for Line 1 (2% of the average monthly rent); and Rs. 618 per month for Lines 2, 3 and 7 (6.3% of the average monthly rent).

An average beneficiary's medium-term benefits are 2.7 (6.8) times the short-term benefits due to Line 1 (Lines 2, 3 and 7). In the short run, 23% of commuters have a positive willingness to pay for Line 1, and 56% of commuters have a positive willingness to pay for Lines 2, 3 and 7. In the medium run, practically every household has a positive willingness to pay for both systems. Therefore, in terms of aggregate benefits, the annual medium-term benefits for Line 1 (Lines 2, 3 and 7) are 5-10 times (11 times) the short-term benefits. A comparison of benefits implied by the two models highlights the importance of the time horizon assumption, which is of interest for policy evaluation.

Who benefits from public transportation is an important policy consideration when evaluating transport projects. The spatial pattern of expected compensating variation implied by the short-term and medium-term models reveals that households close to the Metro stations receive greater time savings benefits. In the short term, women, workers with college or above education, and those with above median income receive greater benefits than their counterparts. The two Metro projects generate similar travel time savings for different sub-groups of individuals in the sample. Therefore, the heterogeneity in benefits is due to differences in preferences captured by the models. Medium-term benefits exhibit a similar pattern of heterogeneity.

Despite the much smaller extent of Line 1 (11.4 km) compared to Lines 2, 3 and 7 (92 km), for an average beneficiary, the short-term (medium-term) benefits of Line 1 are 80-98% (31%) of those due to the upcoming lines, highlighting the importance of its strategic location. Line 1 provided the first east-west rail link in the city, while most of the upcoming network runs north-south, parallel to the existing Suburban Railway network (Figure 1.1). The aggregate benefits of Line 1 are 30-40% of those due to the upcoming lines both in the short and medium terms, although the upcoming network benefits many more individuals. I also find that the dispersion in benefits across individuals due to upcoming Lines 2, 3 and 7 is greater than that due to Line 1. This is consistent with the general claim that widely accessible transport projects are more equitable than strategically placed ones that focus on a smaller area.[2]

A back-of-the-envelope comparison of aggregate benefits and the annualized capital costs suggests that the medium-term aggregate benefits of both projects outweigh the equivalent annualized capital cost (EACC). The short-term aggregate benefits of Line 1 are close to the EACC, but this is not the case for the upcoming network.

This paper is organized as follows. Section 1.2 presents stylized facts about Mumbai and the city's transport network, including existing and planned rapid transit. Section 1.3 presents a description of travel patterns and characteristics of the sample used to estimate both models. Section 1.4 discusses the commute mode choice model used to compute short-term time savings benefits. Section 1.5 discusses the combined housing and commute mode choice approach for computing medium-term time savings benefits. Section 1.6 compares the results from the two modeling approaches with the existing literature, followed by the conclusion in Section 1.7.

[2] Reddy et al. (2010) compares subways in Hong Kong, London, Paris, Singapore, Taipei and U.S. in terms of their accessibility and policymakers' objectives. All but the Hong Kong subway focus on maximizing accessibility and providing equitable benefits.

1.2 Context

This paper is focused on the Greater Mumbai Region (henceforth, Mumbai), a subset of the Mumbai Metropolitan Region (MMR). With a population of over 20 million, MMR is one of the most populous metropolitan areas of the world. Mumbai is the core of MMR, with a population of 12.5 million in 2011 (the last available Census) in an area of 603.4 sqkm. It is located along the central-western coast of India, surrounded by the Arabian Sea on the east, west and south. The city's habitable areas and the existing road and rail network are shown in Figure 1.2. Mumbai is divided into 24 administrative wards that comprise 6 zones (Appendix Figure A.1). The southern tip of the city (Zone 1) is the traditional city center. Zone 3 is a newly developed commercial and employment center. Zones 4, 5 and 6 constitute the suburban area.

There has been a northward movement of employment and households in the city over time, made possible by the presence of public transit, and lower property prices in the suburbs. Generally, population and employment are concentrated along the main rail lines. Figure 1.4 shows a map of employment density at the section level, based on the 2013 Economic Census.[3] 80% of workers in formal jobs are concentrated in Zones 1-4. Figure 1.3 shows a map of population density in the city based on the 2011 Census. About 70% of individuals live in Zones 1-4.

Mumbai Suburban Railway, spread across 100 km, is at the heart of the city's public transit network. It has two types of rail corridors, fast and slow, that operate on parallel tracks. The faster corridor has fewer stops and allows an average speed of 45-50 kmph as compared to the average speed of 35 kmph on the slower one. Suburban trains have a severe overcrowding problem–

[3] Mumbai is divided into 88 sections for the purposes of Census calculations.

there are about 14-16 passengers per sqm. of floor space (Hindustan Times (2017)). These trains generally do not have AC, but owing to rising temperatures and public demand, some AC trains have been introduced since December 2017.

There is also an extensive network of public buses that complements the rail system, but its importance has been declining over time. Average daily bus ridership in 2019 was 2 million passengers (DNA India (2019)), as compared to 4.2 million in 1997-98 (Korde (2018)). This is likely due to the traffic congestion in the city, and poor upkeep of public buses combined with rising household incomes that have led to a sharp increase in private vehicle ownership over time. Two-wheeler and car population in the city increased by 340% and 200% from 2000 to 2017, further contributing to the traffic congestion.[4] The average speed in the city during morning rush hours is 22 kmph, with a peak traffic speed of 7 kmph.[5]

The Metro rail project was planned to alleviate Mumbai's congestion problems. The existing and planned lines are shown in Figure 1.1. The first Metro rail line, Line 1 became operational in 2014. It is 11.4 km long and provided the first east-west rail link in the city. Three additional lines, Lines 2, 3 and 7, 92 km in length, are expected to become operational in 2021-22. Line 2b will provide another east-west rail link, while the remaining parts of the upcoming network are running parallel to the Mumbai local railway network. The average operating speed of the Metro is lower than the suburban trains at 25-35 kmph. However, with a much better ambiance relative to the traditional transit services, it is expected to attract train and bus users as well as those using other transportation modes such as auto-rickshaw, private vehicles, and taxis.

Other improvements to the city's transport infrastructure over the years include new flyovers

[4]There were 407,306 two-wheelers and 303,108 cars in Mumbai is 2000. Their population increased to 1,784,657 and 911,856 by 2017, respectively. (Source: Department of Motor Vehicles, Maharashtra)

[5]Slow speeds and congestion in Mumbai are due to traffic as well as the city's road infrastructure (Akbar et al. (2018)). Its shape and coastal location also constraint its development (Harari (2020)).

and the monorail (20 km in length). The demand for Monorail was not as high as the planners expected and hence, has been deemed unsuccessful by the policymakers. Future extensions to the monorail were also scrapped (Gangan (2017)).

1.3 Mode Choices and Individual Characteristics

This paper uses information on the residential location of households, household members' residence-to-work commute patterns, including workplace locations, from a representative household survey of Mumbai. The survey was conducted by the World Bank in January-March 2019 to study gender and travel patterns in the city (Alam et al. (2021)).[6] 3,024 households were sampled in proportion to the population at the ward level. Two members were interviewed in each household, an adult male and female (ages 18-45) with priority given to primary earners and/or decision makers of the household. Sampled households are shown in Appendix Figure A.2. While I do not know the exact work location of individuals in the survey, I denote the work location by a randomly selected post office that has the same pin code as the survey-reported pin code of the work location.[7] In this section, I describe the commuting patterns based on this survey that will form the basis for estimating the models in Sections 1.4 and 1.5.

The existing pattern of residential and employment locations in the city determines the extent of short-term benefits to commuters due to the Metro rail. Table 1.1 shows the zone-level commute patterns between residence and work location. 72% of sampled workers work in Zones 1-4, while 68% live in Zones 1-4. Commute trips are generally short in the city. 50% of the

[6] This survey was a sequel to another survey conducted by the World Bank in 2004 and follows the same sampling and questionnaire design (Baker et al. (2005)).

[7] There are 88 unique pin codes in Mumbai. The number of post offices per pin code ranges from 1 to 9, with the median being 4. Any measurement error due to this assumption is likely to be random.

workers who commute for work traveled less than 2.5 km to get to their work location.[8,9] 75% of all workers work in the same zone as their residence.

The spatial pattern of commute trip length at the ward level is shown in Figure 1.5. Each commuting worker's household location is matched with the centroid of the ward of work location. Clusters of segments near the ward centroids indicate that most workers (58%) are working in the same ward as their residence. The 25% of the commuting workers who are traveling to Wards K/W, K/E and L are likely to benefit the most from Metro Line 1 in terms of improved first and last mile accessibility. Over 80% of workers live and work in Zones 1-5, and may directly benefit from the upcoming lines.

The success of new transit projects depends not only on their placement but also on individuals' commute mode choices. 8-10% of the workers work from home. Table 1.2 shows the main commute modes for the sample of workers who do commute to work. 33% travel by foot or bicycle, traveling an average distance of about 3 km. About 24% of the workers who commute use public transportation including bus and train with the share of bus being about half of that of train. 10% use auto-rickshaws or taxis, and 34% use private two-wheelers or four-wheelers. Distances traveled by these commuters are in the last row. This suggests that mobility is generally low in the city and that public transit facilitates longer commute distances.

Vehicle ownership and income are important determinants of commute mode choices, and consequently the demand for Metro rail. 50% of workers live in households with at least one vehicle, an increase from 20% in 2004. Columns 6 and 7 of Table 1.2 show modal shares by vehicle ownership. 70% of workers with a vehicle use it as their main commute mode, while the

[8]Over 40% of Mumbai's population is in the workforce (Census 2011).

[9]In 2004, the median commuter was traveling 2.9 km to get to their work location according to the 2004 round of this survey.

rest are equally divided between walking, and public transit or auto-rickshaw.

The differences in commuting patterns for different sub-groups of sample in Table 1.2 reflect the potential heterogeneity in the distributional effects of Metro rail. The difference in average commute distance between men and women is not statistically different, but women are less likely to use private motorized transport than men. On average, workers without a college degree live closer to their work location (3.8 vs 5.5 km) and are more likely to walk to work than workers with a college education (40% vs 14%). This is likely due to the commuting constraints faced by lower income individuals. College educated commuters are more likely to use train and private vehicles because of their potentially higher incomes and greater commute distances. Workers earning above median incomes travel significantly greater distances (4.9 km) than those earning median or below median incomes (4.2 km). They are more likely to travel via their private vehicle than walk or use road-public transit. The models discussed in the following sections use these observations to estimate the underlying preference parameters.

1.4 Short-term Travel Time Savings

This section estimates a structural model that rationalizes commuting workers' short-term commute mode choice decisions. In the short run, a worker with fixed residence and work location chooses a travel mode for commuting to work based on in-vehicle and out-of-vehicle travel times relative to travel costs as in McFadden (1978), Koppelman and Bhat (2006) and Small et al. (2007). I estimate this model under nested logit assumptions to obtain preferences underlying the observed commute mode decisions. Preference parameters are identified in this model based on the tradeoff between mode characteristics. Section 1.4.1 presents the model, Section 1.4.3 discusses the estimation and results, and Section 1.4.4 analyzes the heterogeneity in estimated benefits.

1.4.1 Model

Assuming fixed residence and work location, individual i chooses a travel mode $m \in \{\text{Walk, Bus, Train, Auto-rickshaw, Two-wheeler, Car}\}$ to maximize their utility. I classify these travel modes into $1 \leq K \leq 6$ mutually exclusive categories or nests, denoted by $B_1, B_2, ..., B_K$ based on various criteria of similarity. In the main analysis, $K = 2$ or 3.

Utility U_{imB_k} is assumed to be a function of in-vehicle and out-of-vehicle travel times, income minus out-of-pocket travel costs, scaled to the per trip level, unobserved nest-specific preference, and an individual specific idiosyncratic random shock.[10] I assume a linear additive random utility function,

$$U_{imB_k} = V_{imB_k} + \epsilon_{imB_k} \tag{1.1}$$

$$V_{imB_k} = \alpha_1 * t^{ivt}_{imB_k} + \alpha_2 * t^{ovt}_{imB_k} + \alpha_3 * (w_i - c_{imB_k}) + \delta_{B_k}$$

$t^{ivt}_{imB_k}$ and $t^{ovt}_{imB_k}$ denote the in-vehicle and out-of-vehicle travel times in minutes for individual i's commute trip taken via mode $m \in B_k$. c_{imB_k} denotes the per trip out-of-pocket cost, w_i the individual wage scaled to the per trip level and δ_{B_k} the mean utility for nest B_k. V_{imB_k} is the deterministic portion of the utility. $w - c$ enters the model linearly for computational simplicity.[11]

[10] I assume 22 working days and 2 trips per day, so the value of the monthly Hicksian bundle is divided by 44 to scale it to the per trip level.

[11] Allowing it to enter non-linearly as Cost/Wage lowers the estimated preferences for travel time slightly. Appendix Tables A.3 and A.4 show estimated parameters and predicted shares respectively.

ϵ_{imB_k} is an i.i.d. random utility shock following a generalized extreme value distribution.

$$\epsilon_{imB_k} \sim exp\Big(\sum_{k}^{K}\big(\sum_{m\in B_k} -e^{\epsilon_{imB_k}/\lambda_{B_k}}\big)^{\lambda_{B_k}}\Big) \tag{1.2}$$

In this specification, I assume that for any two alternatives m_1 and m_2 in nest B_k , $\epsilon_{im_1B_k}$ is correlated with $\epsilon_{im_2B_k}$. Any two alternatives across nests are assumed to be uncorrelated, i.e., $Cov(\epsilon_{imB_k}, \epsilon_{im'B_l}) = 0$ for $m \in B_k$ and $m' \in B_l$. The parameter λ_{B_k} represents the degree of independence among the alternatives in nest B_k. The probability of an individual choosing alternative $m \in B_k$ is given by

$$P_{im} = \frac{e^{V_{im}/\lambda_{B_k}}\big(\sum_{j\in B_k} e^{V_{ij}/\lambda_{B_k}}\big)^{\lambda_{B_k}-1}}{\sum_{l=1}^{K}\big(\sum_{j\in B_l} e^{V_{ij}/\lambda_{B_l}}\big)^{\lambda_{B_l}}} \tag{1.3}$$

The average monetary value of time is the marginal rate of substitution between time and cost. Therefore, the average value of in-vehicle time is $\dfrac{\alpha_1}{\alpha_3}$ and the value of out-of-vehicle time is $\dfrac{\alpha_2}{\alpha_3}$. This is the value of time savings (VTTS) measure commonly used in the literature (Koppelman and Bhat (2006), Small et al. (2007), Tsivanidis (2019), Craig (2019), Akbar (2020), Buchholz et al. (2020)). A rough estimate of the VTTS associated with an infrastructure project is computed by multiplying the changes in in-vehicle and out-of-vehicle times by the respective marginal value of travel times for users affected by the project. This does not, however, allow for the fact that modal shares may change in response to the policy.

Expected compensating variation is the preferred welfare measure for a change in the attributes of travel modes since it allows commuters to change their commute mode in probability (Small and Rosen (1981), Varian (1992), Small et al. (2007)). By definition, compensating variation is the change in the value of the Hicksian bundle after a policy change so that expected utility is equal to what it was pre-policy. It is defined as follows.

$$E\big[\max_m U(t^{ivt,0}_{imB_k}, t^{ovt,0}_{imB_k}, w^0_i - c^0_{imB_k}, \delta^0_{B_k})\big] = E\big[\max_m U(t^{ivt,1}_{imB_k}, t^{ovt,1}_{imB_k}, w^0_i - c^0_{imB_k} - CV_i, \delta^0_{B_k})\big] \tag{1.4}$$

The superscript 0 indicates baseline variable values and the superscript 1 indicates variables changed by the policy. Due to the linear-in-parameters additive random specification with income also entering linearly, expected compensating variation has an exact formula (Kling and Thomson (1996)) given below.

$$\frac{1}{\alpha_3}\left[ln\big[\sum_k \big(\sum_m e^{(V^1_{imB_k}/\lambda_{B_k})}\big)^{(\lambda_{B_k})}\big] - ln\big[\sum_k \big(\sum_m e^{(V^0_{imB_k}/\lambda_{B_k})}\big)^{(\lambda_{B_k})}\big]\right] \tag{1.5}$$

For sensitivity analysis, I also consider a model where there is only one nest $K = 1$, i.e., a conditional logit model; and a correlated random coefficients model (or mixed logit model) allowing for parameters α_1,α_2, and α_3 to vary by individual and follow a joint Gaussian distribution. Nested logit is the preferred model because of the unrealistic substitution patterns implied by the conditional logit model, and because empirically, it fits the data better than a mixed logit model (discussed in Section 1.4.3).[12]

[12]A conditional logit model with all mode-specific intercepts is not preferred because of insufficient statistical power to identify preferences from the variation in in-vehicle travel times that remains after accounting for mean preferences for modes.

1.4.2 Data

The data from the 2019 World Bank household survey used for this analysis has information on up to three chosen modes for a typical residence-work commute, along with time spent on each mode. The chosen travel mode in this model is the 'main commute mode' defined as the motorized mode with maximum duration, or the non-motorized mode with maximum duration if that is the only reported travel mode. Metro rail users are not well identified in the data because of the small sample size and due to the absence of a separate response category for 'Metro rail' in the question that asks about the usual commute mode.

To estimate preferences for travel time, I need information on travel time and costs for travel between residential and work locations for all travel modes that are feasible, not only the chosen mode as observed in the survey data. I compile this information using multiple sources.

First, I implement a network program to compute travel time along the shortest duration path for each residence-work commute trip in the sample by rail and walking.[13] In addition to computing travel times under the existing rail network including Suburban Railway and Metro Line 1, I also calculate travel times with the upcoming Metro project added to the existing network and without any Metro line for counterfactual welfare calculations.

Second, I obtain a dataset with travel times for shortest duration drive and transit trips for 500,000 and 250,000 randomly selected origin-destination pairs, respectively.[14] I use origin-destination pairs from this dataset that are within 1 km of the survey households' origin-destination points. The median distance between survey households and the origin point of a matched trip

[13]The program, implemented in Python, uses origin and destination locations, maps of the road and rail networks from Open Street Maps and speeds to compute the travel time along the shortest duration path between an origin-destination pair using Dijkstra's algorithm. Details are in Appendix Section A.2.

[14]This dataset from 2018 was compiled and generously shared by researchers at the Asian Development Bank.

in this dataset is 148 meters.[15] Google Maps API gives step-by-step detailed information for any trip but this dataset has overall travel durations only. As a result, it is not possible to distinguish between train and bus trips in the transit data. The main advantage of these data is that travel times for driving trips account for traffic conditions and allow me accurately model the tradeoff between rail and road transport, which is critical because of the traffic problems in Mumbai.

Third, I use HERE API to obtain detailed step-by-step information about transit trips by train or bus for each residence-commute trip. This information allows me to identify access time, transfer time and the in-vehicle travel time for transit options separately. Most of these trips are by bus, therefore, these data also allow me to identify travel time by bus separately from that by rail.

In constructing the in-vehicle time variable, travel time by train is always from the network program. Travel time by bus is from HERE data, whenever the information is available. In the absence of valid data from HERE, the maximum of Google Maps transit and Google Maps drive time is used.[16] This happens in 17% of cases (481 trips). While HERE data allows the identification of transfer time for transit, in the main analysis, out-of-vehicle time refers to the initial access time, and in-vehicle time includes transfer time unless otherwise stated.[17]

The out-of-vehicle time variable for train and bus is the walk time from a household to the nearest railway station or bus stop. This is computed using the network program assuming a walking speed of 5 kmph. For non-motorized trips, this is the walk time to the post office chosen as the work location. For auto-rickshaw, this value is taken from the survey data. In Section 1.4.3,

[15]The median distance between the post office and the destination of a matched trip is 717 meters; but, since the post office is not the exact work location, this is simply classical measurement error.

[16]Sometimes HERE queries resulted in valid trips but missing travel times, while sometimes they returned completely empty results.

[17]Since the exact work location is not known, and destination of the commute trip is a randomly chosen post office in the pin code of the work location, including last mile access time only introduces measurement error.

I alter these definitions to test the sensitivity of estimated preference parameters. In-vehicle time for non-motorized trips, and out-of-vehicle time for car and two-wheeler is always zero.

Out-of-pocket costs for bus, train and auto-rickshaw are calculated using the per km official fare rules relevant for a single-trip in 2019.[18] For two-wheeler and car, assuming a mileage of 26 kilometer per liter (kmpl) and 12 kmpl, I calculate the cost per trip km using the prevailing petrol price in Mumbai at the time (Rs. 86.16 per liter). I multiply the commute distances by the cost per km to calculate out-of-pocket costs.

Table 1.3 presents data on travel time, out-of-pocket cost, distance to work location, and individual income by the main commute mode chosen. Both the average travel time and distances are the greatest for train commuters, while the cost per trip is the lowest for train commuters. On average, individuals commuting via two-wheeler and cars have a higher income than the remaining sample. The distribution of average monthly incomes for train users in the last panel is indicative of the relatively low levels of segregation among these commuters as compared to bus and auto-rickshaw commuters.

1.4.3 Estimation and results

I estimate parameters for the nested logit model in equations 1.1 and 1.2 using maximum likelihood estimation.[19] I classify commute modes into nests based on two criteria– similarity in scheduling flexibility or general accessibility and autonomy, as measured by the private or public nature of the travel mode. Bus and train are the least flexible of the publicly available options because of their fixed schedule. Auto-rickshaw and walking are less flexible than using

[18] I conduct robustness checks using fares for commuters with a monthly or quarterly pass for bus and train.

[19] In all estimations in this Section, car and two-wheeler are assumed to be feasible only for individuals who own them and walking is feasible for commute distances up to 7 km.

a two-wheeler or a car because of their logistical infeasibility in certain locations or for longer distances.[20] Commuters are also more likely to choose train for longer commute distances (Table 1.3) which could indicate an absence or unreliability of buses on certain routes.

Due to the arbitrariness in classifications, I present estimated parameters based on three different nesting structures (Table 1.4). The nesting structure in Column 1 is {(Car, Two-wheeler), (Bus, Train), (Walking, Auto-rickshaw)}, in Column 2 is {(Walk, Car, Two-wheeler), (Train, Bus, Auto-rickshaw)}, and in Column 3 is {(Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus)}. The specifications in Columns 1 and 3 are preferred because they fit the data better and generate predictions closer to the true shares. Therefore, they will be the focus of the following analysis. The modal share predictions generated by these models are in Table 1.5.[21]

Results in Table 1.4 indicate a distaste for longer commute time.[22] The distaste for out-of-vehicle time is greater than the distaste for in-vehicle time, a common finding in this literature. Estimated nest-intercepts indicate that the mean utility for private and flexible modes is higher than that for bus or train. On average, the value of in-vehicle time is Rs 0.4-0.6 per minute, and the value of out-of-vehicle time is Rs. 1.4 per minute. This is about 20-31% and 72% of the mean wage, respectively.

These estimates lie in the range of estimates in the existing literature that are commonly used for transport policy analysis. The U.S. Department of Transportation recommends setting the value of time equal to 50% of the wage. Concas and Kolpakov (2009) report a range of estimates from the literature ranging from 20% to over 100% of the wage, with 50% being

[20] For example, auto-rickshaws are not allowed in South Mumbai.

[21] To obtain predictions of modal shares consistent with equation 1.3, I constraint the dissimilarity parameters to 1 whenever they exceed 1 by a significant magnitude.

[22] Note that in multinomial choice models, the sign of the estimated coefficients can be different from the sign of the marginal effects, but throughout this paper, they match. I do not report marginal effects since the focus is on monetary values obtained by computing the marginal rate of substitution.

considered a reasonable estimate (Small (2012)). Craig (2019) estimates the value of time in British Columbia to be 58% of the mean wage. Buchholz et al. (2020) estimates the value of waiting time in Prague to be equal to the mean wage.

These preference estimates are also robust to different measures of in-vehicle and out-of-vehicle travel time. I estimate Models 1 and 3 using (i) transit time information from HERE so that transfer times for bus and rail are included in out-of-vehicle, and excluded from in-vehicle time, and (ii) out-of-vehicle time exclusively from the household survey based on the rationale that households behave according to their perceptions of travel time, and not necessarily the actual time. Results for Model 1 are in Table 1.6, and for Model 3 in Appendix Table A.1. The values of time implied by these definitions are slightly smaller, but are in the same range as the main results. As a percentage of wage, the values of in-vehicle and out-of-vehicle time are 5-13 percentage points smaller. Table 1.7 compares predicted modal shares for the three models in Table 1.6 and Appendix Table A.1 with the true shares. Models 2 and 3 predict modal shares slightly better than Model 1, when alternative definitions of out-of-vehicle time are used.

For comparison, I also estimate a correlated random coefficients (mixed logit) model. This model allows for more flexible substitution patterns and random individual specific taste heterogeneity that may be correlated. Correlated taste heterogeneity captures the possibility that individuals who have a greater distaste for in-vehicle travel time may also have a greater distaste for out-of-vehicle time and a lower marginal utility of money. Table A.2 compares modal shares predicted by the nested logit Models 1-3 from Table 1.5 with a mixed logit model. The nested logit models perform better than the mixed logit specification in terms of predictions.

1.4.4 Counterfactual Analysis: Time Savings Benefits of Metro Rail

The existing rail network in Mumbai consists of the Suburban Railway and Line 1. I compute counterfactual in-vehicle and out-of-vehicle travel times via rail using the network program described in Section 1.4.2 by (i) removing Line 1 from the existing rail network map, and (ii) adding Lines 2, 3 and 7 to the existing rail network map. Using estimated preference parameters from Models 1 and 3 in Table 1.4, I compute expected compensating variation measure defined in equation 1.5 to value the time savings benefits due to Line 1 and the upcoming Lines 2, 3 and 7. Note that these benefits are an underestimate of the true value of benefits because the model does not account for the differences in infrastructure between the Suburban Railway and the Metro. The distaste for existing public transit infrastructure is captured in the nest-specific intercept that is assumed to be fixed in these calculations.

In terms of time savings, Line 1 reduces the in-vehicle commute time for 9% of commuters, while Lines 2, 3 and 7 reduce it for 30% of commuters. Conditional on positive time savings, the average time savings is 13 minutes for Line 1 and 9 minutes for Lines 2, 3 and 7. Line 1 reduces the out-of-vehicle travel time for 14% of commuters, while Lines 2, 3 and 7 reduce it for 41% of commuters. Average out-of-vehicle time savings conditional on positive savings is 21 minutes for Line 1, and 12 minutes for Lines 2, 3 and 7.

The monetary benefits of travel time savings are presented in Table 1.8. 19% of commuters have a positive willingness to pay for benefits due to Line 1, while 51% have a positive valuation of benefits due to Lines 2, 3 and 7. Conditional on the value of time savings being positive, the mean value of time savings implied by Model 1 is Rs. 71 per month for Line 1, Rs. 85 per month for Lines 2, 3 and 7. The value of benefits due as a proportion of the average out-of-pocket commuting cost imply that the average beneficiary is willing to pay an additional 9% (12%) for the time savings benefits due to Line 1 (Lines 2, 3 and 7).[23] The corresponding value of time savings implied by Model 3 are Rs. 97 and Rs. 99 per month for the existing Line 1, and the upcoming Lines 2, 3 and 7, respectively. These are 12-14% of the out-of-pocket costs for the average beneficiary. Metro Line 1 affects fewer individuals than Lines 2, 3 and 7, given its length. However, due to its placement, the average benefits conditional on benefits being positive are similar to those under the upcoming network. The majority of monetized benefits due to time savings are attributable to reductions in out-of-vehicle time: individuals value out-of-vehicle time more than in-vehicle time and the reduction in out-of-vehicle time is greater than the reduction in in-vehicle time.

The spatial distribution of expected compensating variation (Figure 1.6) sheds further light on the nature of travel time benefits. Many more commuters benefit from Lines 2, 3 and 7 due to the wide accessibility of the upcoming network (92 km) than under Line 1 (11.4 km). Commuters in the vicinity of metro stations experience the maximum benefits, mainly due to reductions in out-of-vehicle access times. But commuters in other parts of the city also experience non-trivial benefits, due to improved transit connections. This is especially so for the 11.4 km long Line 1 which provided the first east-west rail link in the city.[24]

Equity is an important consideration in transit infrastructure planning. Depending on the spatial structure of a city, certain groups are ex-ante more likely to benefit due to improved transit infrastructure (Baum-Snow and Kahn (2000), Glaeser et al. (2008), Akbar (2020)). To understand which groups receive greater benefits due to Metro rail, I estimate equation 1.1 for

[23]Note that the average beneficiary could be using any mode.

[24]Anecdotally, Metro Line 1 greatly reduced travel time by eliminating the need for commuters from the north of Mumbai to change trains at the Dadar railway station, one of the major interchange stations of the Mumbai Suburban Railway.

various subgroups of individuals distinguished by gender, education and income level. Table 1.8 shows the value of in-vehicle and out-of-vehicle travel times for each group and the value of travel time benefits due to Line 1 and Lines 2, 3 and 7.

Conditional on benefits being positive, women experience 16-23% greater benefits than men under both Metro projects.[25] The reduction in both in-vehicle and out-of-vehicle times is similar for men and women under both metro projects. However, women have a greater distaste for travel time, especially out-of-vehicle time, compared to men, as implied by their marginal rate of substitution. This is also reflected in the fact that a greater proportion of women use publicly available modes despite traveling similar distances as men (Table 1.2).[26] The value of time savings conditional on positive savings due to Line 1 (Lines 2, 3 and 7) is Rs. 85-115 (Rs. 98-113) per month for women and Rs. 68-94 (Rs. 85-99) for men.

The reduction in travel times for workers with and without a college degree are similar, but the former group experiences greater benefits under both Metro projects. The value of benefits due to Line 1 (Lines 2, 3 and 7) to college educated workers is over 100% (82%) higher than the benefits to workers without a college degree. This is because workers with a college degree have a stronger distaste for both in-vehicle and out-of-vehicle travel times. They also commute a longer distance compared to individuals with less than college education (Table 1.2). Their expected compensating variation for Line 1 (Lines 2, 3 and 7) is Rs. 117-169 (Rs. 131-147) per month as compared to Rs. 57-73 (Rs. 72-82) per month for workers below college education. College educated workers are also more likely to have above median incomes in the sample.

Commuters with above median income experience greater benefits than their counterparts.

[25]This difference is significant at the 89-93% confidence level because of the limited sample size of women beneficiaries. All other differences reported in this section are significant at the 99% level.

[26]The difference in value of time is due to differences in preferences as implied by the models' nesting structures.

The value of savings due to Line 1 (Lines 2, 3 and 7) is 100-135% (20-85%) higher for above median commuters. Both these groups experience similar time savings, indicating that this pattern is due to a difference in preferences. The mean value of time savings conditional on positive benefits for Line 1 (Lines 2, 3 and 7) is Rs. 90-165 (Rs. 87-113) per month for the above median group, and Rs. 38-82 (Rs. 47-94) for the subsample with median or lower income. I also test for heterogeneity in benefits by vehicle ownership using the main specification and summarizing the value of benefits separately by vehicle ownership. Those without a vehicle experience benefits that are three times the value of benefits experienced by vehicle owners.[27] The dispersion in monetized benefits accruing to different sub-groups is higher under the upcoming network than under Line 1, suggesting that the upcoming project is more equitable than Line 1, due to its wider accessibility.[28]

1.5 Medium-term Travel Time Savings

This section presents a structural model that explains individuals' housing and commute mode choices. In the medium run, households simultaneously select a housing location and a commute mode for their primary worker assuming an exogenously fixed work location. This model is based on the framework of McFadden (1973), McFadden (1978) and Berry et al. (1995) as modified for the context of residential location choice in Bayer et al. (2004) and Bayer et al. (2007).

This framework improves upon the short-term commute mode choice model by allowing households to optimize their housing location in addition to commute mode. Section 1.5.1

[27]There is limited statistical power to separately estimate the model for these subsamples.
[28]This is based on a two-sample variance comparison test.

presents the model and discusses its estimation. Data and characteristics of households in the estimation sample are discussed in Section 1.5.2. Section 1.5.3 discusses the estimation results. Section 1.5.4 discusses the implied value of time savings benefits due to the two Metro rail projects.

1.5.1 Model and Estimation

Assuming a fixed work location, a household i chooses a house h from the set of housing types in its feasible set along with a commute mode $m \in \{\text{Walk, Rail, Road}\}$ to maximize their utility.[29,30] I assume a linear-additive structure for the random utility function, U_{ihm} given by the following,

$$\begin{aligned} U_{ihm} &= V_{ihm} + \epsilon_{ihm} \\ V_{ihm} &= \gamma_m + \beta_K * K_{ihm} + \beta_Z * Z_{ih} + \alpha_p * \mathrm{P}_h + \alpha_x * X_h + \nu_h \end{aligned} \tag{1.6}$$

V_{ihm} refers to the deterministic portion of the utility function. ϵ_{ihm} is the household-specific idiosyncratic shock component assumed to follow an i.i.d. Type I extreme value distribution.

γ_m is the mode specific constant that captures the mean utility for mode m common across households. K_{ihm} denotes the mode m-specific travel time between house h and the fixed work location of the primary worker of household i. Z_{ih} denotes housing attributes that vary by household such as the proportion of households within 2 km of house h that have the same religion or language as household i. P_h is the monthly rental price of housing h. X_h denotes housing

[29] In this model, mode definitions are broader than the commute mode choice model to reflect the medium-term decision horizon.
[30] The option of walking is modeled separately because of the high prevalence of it in the context of Mumbai. This is not the case in Craig (2019) and Akbar (2020).

characteristics other than rental price that do not vary across households, such as floor space, condition of roof, presence of an indoor toilet, and neighborhood characteristics such as access to jobs in the city, prevalence of crimes against women, whether the neighborhood is a slum area, and distance to the coast. ν_h captures unobserved preferences for housing h.

β_K and β_Z capture average preferences for attributes in Z_{ih} and K_{ihm}. In some specifications, I use individual characteristics such as education, income level, and vehicle ownership as taste shifters. In those cases, β is composed of two components, one that is constant across all individuals, and one that is constant across all individuals within a specific income or education category but varies across categories. For example, $\beta_Z = \bar{\beta_Z} + \beta_Z^{edu}$. This heterogeneity in preferences captures sorting based on income, education and mobility. α_p and α_x capture average preferences for rental price and housing-specific characteristics in X_h.

I estimate the model described above in two stages following the two-step approach of Berry et al. (1995) as modified for residential location choice in Bayer et al. (2004) and Bayer et al. (2007). In the first stage, I estimate the parameters of the following equation using maximum likelihood estimation.

$$U_{ihm} = \gamma_m + \beta_K * K_{ihm} + \beta_Z * Z_{ih} + \delta_h + \epsilon_{ihm} \tag{1.7}$$

The housing-specific variables in equation 1.6 are subsumed in δ_h, the housing specific constant that captures the mean utility for house h. It will be decomposed in a second stage to study the tradeoff between different housing specific characteristics. γ_m and δ_h capture both observed and unobserved preferences for m and h, respectively that are common across individuals. Their presence allows for the consistent estimation of β_K and β_Z as long as there aren't any other unobservables that are correlated with Z_{ih} and K_{ihm}.[31]

In the second stage, $\hat{\delta_h}$ is decomposed using a linear model with random errors to estimate preferences for house-specific attributes that do not vary by individual.

$$\hat{\delta_h} = \alpha_p * \mathrm{P}_h + \alpha_x * X_h + \nu_h \tag{1.8}$$

Each house observed in the household survey is assumed to represent a housing type. Housing type or house h is feasible for a household i if the survey-reported monthly rental cost of h is lower than the monthly income of household i. My estimation sample has 2,209 households choosing among 2,209 houses and two or three modes per housing location.[32] Since the number of alternatives available per household is large, for computational reasons, I take a random sample of the feasible set in estimating the model. When the feasible set is large, taking a random sub-sample of the feasible set does not bias the results in the conditional logit model (McFadden (1978)).[33]

I use the estimated housing-specific intercepts from the first stage to estimate equation 1.8 using two-stage least squares. Unobserved housing attributes omitted from this equation contained in ν_h are likely correlated with rental price. To identify α_p and α_x, I need an instrumental variable for the survey-reported monthly rental price. I instrument for rental price using administrative sale value assessment of properties in the neighborhood of h. Neighborhoods for the purposes of price assessments, called sub-zones, are defined by the municipal government based

[31] For example, if some commute category were systematically preferred for certain housing locations by certain households for reasons other than travel time, β_K would not be consistent. It is difficult to think of such a scenario since the commute modes modeled here are fairly broad categories.

[32] For each individual, rail and motorized transport options are always available. Walking is not an option when individual's work location is more than 7 km away from a housing location.

[33] This simplification leads to a slight loss in precision in the first-stage estimates but not enough to outweigh the computational gains.

on a variety of factors including historical boundaries, land regulations, market values and market potential. These sale value assessments are made for the purposes of collecting appropriate transaction and property taxes, and may differ from the market price of the houses. For house h in sub-zone s_h, rental price can be written as a function of prevailing assessed values in sub-zone s_h.

$$\mathrm{P}_h = \omega * \text{Assessed sale value}_{s_h} + \alpha_x * X_h + \zeta_h \tag{1.9}$$

This instrument uses the sale value of properties in the vicinity of h but not necessarily the value of h itself.[34] This instrument is valid as long as it is not correlated with unobserved correlates of estimated house intercepts. Formally, the exclusion restriction for instrument validity requires $\mathrm{Cov}(\text{Assessed sale value}_{s_h}, \nu_h) = 0$. This is likely to be true given the heterogeneity in housing types within a sub-zone s that are captured by the amenities in X_h. Since the survey-reported rental price and the assessed values are different from the true market value, the correlation between unobserved amenities and these prices may be limited.[35] This is also likely to be true since the correlation between observed amenities and assessed values is limited. The same argument, however, suggests that the instrument may be weak. I show robustness to the weak instrument problem using the inference criteria suggested in Lee et al. (2021).

To value time savings benefits using this model, I compute the expected compensating

[34] Another option for an instrument is as constructed in Bayer et al. (2007) using a two-step approach. In the first step, variables describing housing stocks and land use beyond 3 miles are used to instrument for prices within 1-mile, 2-mile and 3-mile rings. In the second step, parameter estimates from this equation are used along with the assumption that there are no housing level attributes excluded from the equation to obtain a predicted price vector that clears the market. Validity relies on the assumption that predicted price captures only exogenous variation in pricing, i.e., there were no unobservables in the first step.

[35] Survey reports of rental price are susceptible to measurement error on account of some houses not being rental properties, some rental agreements being subject to a lock-in price, and due to recall bias.

variation for a reduction in house-workplace commute time via rail. For a conditional logit model with income entering the utility function linearly, the exact formula for expected compensating variation is given below (Small and Rosen (1981)).

$$\frac{1}{\alpha_p}\left[ln\big(\sum_{hm} e^{V_{hm}(\gamma_m^0, P_h^0, X_h^0, Z_{ih}^0, K_{ihm}^1)}\big) - ln\big(\sum_{hm} e^{V_{hm}(\gamma_m^0, P_h^0, X_h^0, Z_{ih}^0, K_{ihm}^0)}\big)\right] \tag{1.10}$$

Superscript 0 denotes baseline values of travel time under the existing rail network, while 1 denotes the travel time in the absence of Metro Line 1 or with the addition of Lines 2, 3 and 7.

To compute expected compensating variation for counterfactual policies that reduce travel time via rail assuming everything else to be constant, I use the formula in equation 1.10. Compensating variation for each individual is a function of their idiosyncratic utility shock, and the expectation of this distribution is obtained for ease of interpretation. A household will have positive benefits by this measure if the travel time by rail is reduced between their primary worker's work location and any house in their feasible set that has a positive probability of being chosen.

1.5.2 Data

The model described above is estimated using the 2019 household survey data conducted by the World Bank. The final estimation sample has 2,209 households.[36] The following sub-sections discuss the estimation sample, and the construction of a general employment accessibility index.

[36]This is due to dropping households for whom information on housing amenities used in the main specification was missing.

1.5.2.1 Household and Housing Characteristics

To estimate preferences for commute time, I need travel times for each potential residence-work-location pair. I implement the network program as in Section 1.4 to compute travel time along shortest duration travel path between each potential origin-destination point in the sample via walking, rail, and motorized road transportation (described in Appendix Section A.2).

Table 1.9 describes the mode choices, distances traveled and commute times of the households in the sample. The median household lives 2.6 km away from their primary worker's work location. 55% of the primary workers choose some form of road transportation, private or public, 16% travel by rail, while 29% walk. The median primary worker commutes for 15 minutes one-way.

The heterogeneity in mode choices by household characteristics is also described in Table 1.9. Households whose primary worker has a college degree (30% of the sample) have a longer residence-work commute compared to households whose primary worker does not have a college degree; however, they travel by faster modes, and have roughly similar commute times as the latter. Similarly, households with higher income travel slightly longer distances but using mostly road transportation or rail, and hence, have slightly lower travel times. Half of the households in the sample own a vehicle. There is no significant difference in commute distance by vehicle ownership, but those with vehicles have shorter travel times. This is because 80% of households with a vehicle choose road transportation for the primary worker's commute, while about 70% of households without a vehicle rely on rail or walking.

Table 1.10 describes household demographic characteristics. There is some clustering in households' chosen locations by religion and language. The two main religions in the sampled households are Hinduism (79%) and Islam (16%). 53% of households state Hindi as their mother tongue, while 36% state Marathi. On average, within a 2 km radius around each household, 45% of households have the same language, and 68% of households have the same religion. To identify household preferences for living close to other households with same language and religion, I compute for each house, the proportion of households within a 2 km radius with a given main religion and language. For each house in a household's feasible set, I identify the relevant proportion of similar households based on the household's own main language and religion.

Housing characteristics included in the model are summarized in the second panel of Table 1.10. The average survey-reported monthly rent for the sample is Rs. 9,704 ($485 PPP). The average floorspace area of the houses in this sample is 262 sqft, with the median house having only a single room. 60% of the houses have a separate kitchen space. Many residential housing structures in the city do not have a provision for toilet or bathroom inside the house, and households living in these places must rely on communal facilities. Access to public transit is typically good in the city. Mean distance to the nearest railway station is 1.5 km, which is an 18 minute walk assuming a walking speed of 5 kmph. The nearest bus stop from most houses in the sample is within a 5 minute walk. I estimate preferences for a general employment accessibility index as well, discussed in the following subsection.

I also obtain administrative data on land value assessments for residential land to instrument for the rental price and to consistently estimate the model parameters in the second stage as mentioned above. For the purposes of land valuation, the city of Mumbai is divided into 725 sub-zones. For each sub-zone, property transactions in a given year determine the official land value for the following year. These assessments are made for the purposes of accurate determination of property taxes and transaction taxes. However, these are typically lower than the market price of a given residence.

1.5.2.2 Employment Accessibility Index

To assess the importance of general employment accessibility as an amenity, I construct an index using a proxy for relative wages across pin codes in the city, and travel time between each house and a randomly chosen post office location in each pin code. This index is a travel time weighted average of the potential attractiveness of pin codes as employment locations as implied by the existing spatial structure. The importance of accessibility in determining urban land use and prices is the foundation of the urban economics literature. Although the application of indices such as the one used here can be traced back to Hansen (1959) in the urban planning literature, their use in the urban economics literature has only been explored in Osland and Thorsen (2008), Ahlfeldt (2011) and Tsivanidis (2019).

Let j be the possible number of work locations in the city. The employment accessibility index for a house h is defined below.

$$\text{EA}_h = \sum_j \left(\frac{w_j}{d_{hj}}\right) \tag{1.11}$$

w_j is the wage obtainable at location j, d_{hj} is the commuting cost from house h to location j. In the absence of data on wages, I use the method in Kreindler and Miyauchi (2021) to construct a proxy for w_j.[37] $d_{hj} = exp(\kappa * t_{hj})$ is the iceberg commuting cost. t_{hj} is the travel time between h and j. $\kappa > 0$ is a decay parameter specifying the semi-elasticity of commuting costs d_{hj}

[37] Household and personal incomes in the survey are not preferred for this measure because they are recorded as categorical variables in a fairly small number of categories, and can not capture fine levels of variations that are obtainable using the method in Kreindler and Miyauchi (2021).

to commuting times t_{hj}. Its value has been estimated structurally within a general equilibrium framework in Ahlfeldt et al. (2015), Severen (2021) and Tsivanidis (2019); and using a reduced-form approach in Kreindler and Miyauchi (2021), which I follow in this paper. The estimation is discussed in Appendix Section A.3. I use the standardized values of EA_h as a housing specific amenity.

1.5.3 Results

Preference parameters in equation 1.6 estimated using a conditional logit specification are presented in Table 1.11. Households have a significant distaste for longer commute times and a preference for living close to other households with the same language and religion. Estimated intercepts for commute category indicate a positive mean preference for walking and road transport relative to taking the train. Mean preferences for walking are the highest, indicating that most households prefer houses within walking distance of the primary worker's work location. Mean utility for rail is the lowest as in Section 1.4, capturing the average distaste for public transit.

In the second stage, the estimated house-specific intercepts from the first stage are regressed on housing-specific characteristics in a two-stage least squares regression following equation 1.8. Log of annual assessed land value for residential properties is used as an instrument for monthly rental price.[38] Table 1.12 presents these regressions for different sets of control variables. With standard errors clustered at the level of sub-zones or a more aggregate level, the first-stage F-statistic$\sim$40.[39] Given the possibility of this instrument being weak, I use an adjusted critical value for t-statistic for inference at the 95% confidence level following Lee et al. (2021). These

[38] Results are not shown here but similar estimates are obtained when using assessed land values for commercial or industrial property types.

[39] According to Lee et al. (2021), first stage F-statistic below 100 may indicate a weak instrument problem.

are reported in the last row of Table 1.12. The coefficient on rental price has the expected sign, and is significant and robust across these specifications. The corresponding first-stage estimates for these two-stage least square regressions are in Table A.5.

Table 1.12 indicates that households have a preference for lower rents, better housing infrastructure, proximity to the coast, areas with less crime, places further away from railway stations, and houses that have a higher accessibility to potentially attractive work locations. While the consistency of estimates in the first stage is independent of the second stage specification, examining the sensitivity of the coefficient on rental price is important to understand the sensitivity of the value of benefits of Metro rail implied by the model. Column 1 has a housing amenity index which is the first principal component of housing amenities available in the survey.[40] Columns 2-5 have the following additional variables: distance to coast, nearest railway station, slum classification of the residential area, number of reports of crimes against women, an index of employment accessibility, an index for proximity to private or government doctor and hospitals.[41] The coefficient on distance to nearest railway station captures the average disamenity associated with being close to a congested transit access point. In so far as access to a transit stop improves the minimum travel time required to reach a potential employment location, the utility due to improved access is captured by the employment accessibility index.[42] Column 4 is the preferred specification: it allows the greatest number of controls without the loss of sample

[40]It includes floorspace, number of rooms and dummy variables for good roof, separate kitchen, separate toilet and bathrooms inside the house, access to piped water, presence of footpath in the neighborhood, and the house being well-connected to the rest of the city. Factor loading of each of these variables in shown in Appendix Table A.6. Except for the presence of footpath and perceived well-connectedness of the house, almost all variables contribute roughly equally to this index.

[41]The index of proximity to health services is the first principal component of related variables from the survey, including categorical variables for the walk duration to the nearest private doctor, private hospital, government hospital. Factor loadings indicating the importance of each of these variables in the constructed index are in Table A.7.

[42]Note that travel time used in this index is the lesser of drive and rail time.

size due to missing observations.

To test the sensitivity of preference parameters to the first-stage specification, I estimate specifications allowing for the preferences for travel time and proximity to households with similar language and religion to vary by income, education and vehicle ownership. The results from the first-stage of these models are in Table 1.13. The second-stage results are similar to those for the main specification and are not shown here for brevity. The coefficients in the first and second stage are similar across these specifications, highlighting the robustness of the model. Unlike the mode-choice model, preferences for commute time do not differ significantly by whether the primary worker of the household has a college degree. However, as before, higher income households have a greater distaste for travel time. Households that own a vehicle also have a greater distaste for travel time than their counterparts.[43]

1.5.4 Counterfactual Analysis: Time Savings Benefits of Metro Rail

I compute counterfactual travel time via rail using the network program by (i) removing Line 1 from the existing rail network, and (ii) adding Lines 2, 3 and 7 to the existing rail network. A household experiences benefits as long as the commute time is reduced between their primary worker's work location and a house in their feasible choice set, and there is a positive probability of traveling via rail. For an average household in the sample, the reduction in commute time via rail between all feasible houses and the primary work location is an average of 5 minutes due to Line 1, and an average of 6 minutes due to the upcoming Lines 2, 3 and 7.

The monetary value of time savings benefits as measured using expected compensating

[43] These comparisons are made in monetary terms using the marginal rate of substitution obtained by dividing the coefficient on travel time by the coefficient on rental price from the second stage.

variation (equation 1.10) is a function of the estimated parameter α_p of the second stage. The second panel in Table 1.14 presents the expected compensating variation under different specifications for the second stage. The average value of time savings benefits due to Line 1 is Rs. 137-189 per month. This is about 1.4-2% of the monthly rent. In comparison, the average value of time savings due to Lines 2, 3 and 7 is Rs. 448-618, over three times the value of benefits due to Line 1. This is 4.6-6.3% of the monthly rental value.[44]

I test the sensitivity of expected compensating variation to the first-stage specification by estimating four different combinations of taste-shifters, and then computing the mean expected compensating variation for each case separately. The results are in Table 1.15. The implied value of benefits for both Metro projects are similar to the results of the main model discussed above.

Spatial variation in expected compensating variation for Line 1 and Lines 2, 3 and 7 as predicted by specification (1) in Table 1.11 is shown in Figure 1.7. As in the case of short-term benefits, the greatest benefits of each system accrues to households living close to Metro rail. For both systems, households in the northern part of the city receive greater benefits than those in the south. In contrast to the short-term benefits in Figure 1.6, medium-term benefits are more dispersed and positive for almost all households in the city.

Different groups of households may benefit differently in the medium-run depending on their work location, their feasible set and their preferences. To study the heterogeneity in benefits implied by this model, I estimate the main specification for the entire sample, and study the variation in expected compensating variation for different sub-samples. Recall that in Section 1.4, to study the heterogeneity in preferences and valuation, I estimated the model separately

[44] Note that these values are lower than the simple value of time estimate implied by the marginal rate of substitution because the expected compensating variation accounts for the probability of changing mode or housing choice and not every individual's commute would benefit equally due to the Metro.

for each sub-group.[45] This approach was preferred in Section 1.4 to prevent confounding of preferences. In this case, because the estimated mean preferences for housing in the first stage are used to compute preferences for housing-specific amenities in the second stage, dividing the sample into smaller sub-samples may not produce credible samples for the second-stage.

As shown in Table 1.16, households with a college educated primary worker have a 43% and 17% higher valuation for time savings relative to households with workers below college education for Line 1 and Lines 2, 3 and 7, respectively. Households owning a vehicle do not benefit significantly differently compared to households without a vehicle due to the Metro.[46]. Average benefits of households in the median income category are 24% higher compared to the average benefits of below median income households.[47]

Medium-term time savings benefits due to the upcoming network of Lines 2, 3 and 7 are significantly more dispersed than the benefits due to Line 1.[48] The proportional differences in the value of benefits received by different sub-groups of sample are also lower for the upcoming network than for Line 1, highlighting the potentially greater equity in benefits due to the upcoming lines.

1.6 Discussion

The short-term benefits of the Metro rail computed using the commute mode choice model accrue to only a fraction of commuters– 19% for Line 1 and 51% for Lines 2, 3 and 7. When

[45] Similar results were obtained using this approach.

[46] This is likely due to not accounting for vehicle ownership while constructing feasible sets. This choice was made to reflect flexibility in mode choice that is generally present in the medium-term.

[47] The difference in average benefits experienced by household above the median income are not statistically distinguishable from those experienced by households in the median income category or below median income categories.

[48] This is based on a two-sample variance comparison test.

households are allowed to adjust their housing location as in the medium-term model, practically every household experiences positive benefits. Conditional on benefits being positive, the average medium-term benefits of Line 1 are 2-3 times the short-term benefits, and the corresponding figure for Lines 2, 3 and 7 is 6-7 times. Relaxing the assumption of fixed housing location has a larger impact on the benefits of the upcoming project with Lines 2, 3 and 7 because of its much larger extent compared to Line 1 (92 km vs 11.4 km).

In a back-of-the-envelope calculation, I scale individual-level mean expected compensating variation using ward-level sample weights to obtain annual aggregate benefits at the city level.[49] Total annual short-term benefits of Line 1 are \$100-200 Million (PPP), while the medium-term benefits are 5-10 times this value. Total annual short-term benefits due to Lines 2, 3 and 7 are \$ 300 Million (PPP), and their medium-term benefits are 10 times the short-term value. Despite the much smaller extent of Line 1, its aggregate benefits are about 30-40% of those from Lines 2, 3 and 7, highlighting the consequences of its strategic placement. The benefits per km generated by Line 1 are higher than those generated by Lines 2, 3 and 7.

A comparison between short and medium-term benefits highlights the influence of the decision time horizon assumption, which is of interest for policy evaluation (Small (2012)). Specifically, for a cost-benefit analysis, the conclusions can differ based on how the benefits are computed. This is especially important when policymakers have political incentives that are tied to the period over which the benefits of infrastructure projects are realized. Using publicly available information on the construction of the Metro rail projects, I compute the equivalent annualized capital cost (EACC) based on various plausible assumptions about the social discount

[49] Households in the sample were chosen such that there was at least one working member, and one male and one female respondent was available. The overall no-response rate was around 5-7%. I scale individual-level expected compensating variation with a factor measuring the relative proportion of worker population in a ward to the number of households in the sample from that ward to obtain population-level benefits.

rate and asset life. One option is to use the interest rate on the original and refinanced loan borrowings of the construction company, 12% (Prasad (2015)).[50] Another option is to use an interest rate that is closer to the long-term yield on government bonds; therefore, I also show results using rates of 10% and 8%. I compare the annual aggregate short-term and medium-term benefits of both Metro projects with their EACC in Table 1.17. The construction cost of Line 1 is $ 2.03 Billion (PPP), and its EACC ranges from $200-$300 Million (PPP). The *projected* construction cost of Lines 2, 3 and 7 is $ 22 Billion (PPP).[51] The equivalent annualized capital cost (EACC) ranges from $ 1.9-3 Billion (PPP). The aggregate medium-term benefits of both projects outweigh the EACC. The aggregate short-term benefits of Line 1 are close to the EACC, but this is not so for Lines 2, 3 and 7. This highlights the potential tradeoff faced by policymakers when prioritizing or choosing between projects.

The distributional effects of transport projects are an important determinant of their success. Transport infrastructure in a city is a central feature determining the spatial distribution of economic activity. Therefore, such projects have strategic importance in enabling certain sub-groups of the population to participate more in economic activity. For example, the presence of high-speed Metro rail has been linked to an increase in women's workforce participation in South Korea (Kwon (2020)). In the context of Mumbai, transport availability may not be the biggest constraining factor for women's workforce participation (Alam et al. (2021)), but the results in this paper indicate that the marginal benefits of Metro rail received by women workers are greater than those received by men, suggesting a potential effect along the intensive margin.

Workers of different skill groups may benefit differently depending on the spatial distribu-

[50] Metro Line is operated by a Public Private Partnership.

[51] Cost of Line 2 is $ 8.2 Billion, Line 3 is $ 10.9 Billion, and Line 7 is Rs. $ 2.9 Billion.

tion of industries and occupations. The introduction of Bus Rapid Transit (BRT) in Bogota led to slightly greater benefits for high-skilled workers as implied by the general equilibrium model in Tsivanidis (2019).[52] In the context of Mumbai, I find that college educated workers receive greater short-term and medium-term benefits due to Metro rail. In Buenos Aires, the general equilibrium welfare gains of BRT accrue similarly to high and low skilled workers (Warnes (2020)). Using a partial equilibrium framework similar to the medium-term model in this paper, in the context of U.S., Akbar (2020) finds that rail transit improvements lead to greater benefits for higher-income groups in cities with relatively high baseline transit speeds, and for lower-income groups in cities with relatively slower baseline transit speeds. Similarly, in Mumbai, higher income households receive greater benefits due to Metro rail. In the context of consumption-related travel in Singapore, high-income workers are found to benefit more due to the 41.9 km long Downtown (Metro rail) Line as implied by the general equilibrium model in (Tan and Lee (2020)).

Caveats to this analysis: An important limitation of the partial equilibrium models used in this paper is that they capture benefits only to households and only through the channel of time savings. There is evidence of important general equilibrium effects of infrastructure projects that improve accessibility between spatially distant workers and firms (Ahlfeldt (2011), Tsivanidis (2019), Severen (2021), Heblich et al. (2020), Warnes (2020)).[53] Despite this, the aggregate benefits of Metro rail implied by the combined housing and commute mode choice model exceed

[52]Incidentally, Tsivanidis (2019) draws the opposite result when using a simple time savings model like the one in Section 1.4. The author attributes this difference in finding to the change in prices due to the 'large infrastructure change'. This could also be due to different parametric assumptions and welfare measures.

[53]Tsivanidis (2019) shows that the value of time savings due to the introduction of Bus Rapid Transit in Bogota as captured by a mode choice model accounts for 60-80% of the total welfare benefits implied by a general equilibrium model. However, the relative importance of benefits arising through different channels is contextual. For example, the ease with which economic actors can adjust in response to infrastructure projects depends on the economic and political environment of a city which is likely to vary across contexts.

at least the equivalent annualized construction cost. Chapter 2 studies the impact of Metro Line 1 on property values and the relative attractiveness of residential and employment locations in the city. Other types of benefits of urban transport projects include the impact on air pollution (Chen and Whalley (2012), Goel and Gupta (2017), Bauernschuster et al. (2017), Bel and Holst (2018), Gendron-Carrier et al. (2018), Guo and Chen (2019), Li et al. (2019), Cropper and Suri (2022)), population changes (Glaeser et al. (2008), Heblich et al. (2020), Pathak et al. (2017), Khanna et al. (2021)), innovation (Koh et al. (2021)), vehicle ownership (Mulalic and Rouwendal (2020)), congestion (Gu et al. (2021)) and trade and economic growth (Donaldson (2018), Banerjee et al. (2020)).

Second, the models in this paper do not allow for employment decisions that may be taken simultaneously with housing and mode choice decisions. The reason for this modeling choice is the lack of reliable data on employment opportunities and potential wages. The only available source of official data on employment opportunities to the best of my knowledge is the Economic Census 2013 which has information on the number of workers by gender and industry for formal establishments in the city. To investigate the implications of this modeling assumption, I use the data from the Economic Census to estimate a multinomial choice model in which an individual chooses an employment location out of 80+ broad localities (either pin codes or section wards) using a similar two-step approach as in Section 1.5. Assuming fixed household location, employment location choice is a function of travel time and the share of workers in the sample that have a similar education level working in that location. I find that the marginal utility of travel time implied by this model is similar to that implied by the combined housing and mode choice model. However, because of a lack of other observables to characterize employment location, the influence of unobserved factors on the estimated marginal utility of money is substantial. This calls into question the value of time savings benefits implied by this model.

Another limitation is that preferences estimated using data from one period may not be relevant for policy analysis for a different time period because of changes in preferences. However, this does not seem to be true in this context. Using comparable data from a similar household survey conducted by the World Bank in 2004, I estimate the simultaneous housing and mode choice model and find that as a proportion of monthly rental price, the value of benefits of each metro project is quantitatively similar to the results reported in this paper. This suggests that the preference estimates reported here are not specific to the time period of data collection.

The counterfactual analyses in this paper can be further improved as more institutional information becomes known. For example, the current set of results do not account for expected future fares for Metro, suburban railway, buses, auto-rickshaws, and petrol/diesel prices. COVID-19 is also likely to change preferences for traveling via congested public transit modes, and potentially differentially across income and education groups. By modifying the mode-specific intercepts, the models in this paper can also be used to simulate individual responses to transport changes under counterfactuals with an increased distaste for congested public transit alternatives.

1.7 Conclusion

In this paper, I estimate structural models of commute mode choice and combined housing and commute mode choice using data from a 2019 representative household survey of Mumbai. I use these estimates to compute the short-term and medium-term monetary values of time savings benefits due to an existing and an upcoming Metro rail project in Mumbai. The existing Metro Line 1 (11.4 km) started operations in June 2014 and provided the first east-west rail link in the city. The upcoming network evaluated here consists of Lines 2, 3 and 7 (92 km) scheduled to open during 2022. The intention behind the introduction of Metro rail in the city was to alleviate the overcrowding problems facing the historic suburban railway network and road traffic congestion by motivating substitutions away from existing modes towards Metro rail. This is expected to improve intra-city commuting accessibility.

Both Metro projects generate substantial benefits, but the short-term benefits are a fraction of medium-term benefits. In the short term, conditional on benefits being positive, the average individual value of time savings is Rs. 71 per month for Line 1 and Rs. 85 per month for Lines 2, 3 and 7 amounting to 9-12% of the average out-of-pocket costs. The average medium-term values are an order of magnitude higher, valued at Rs. 189 and Rs 618 for Line 1 and Lines 2, 3 and 7, respectively. This is about 2-6% of the average monthly rental price. In the short-term 19% of commuters have a positive willingness to pay for Line 1, while 51% of commuters have a positive willingness to pay for Lines 2, 3 and 7. In the medium-term, virtually every household experiences positive benefits. Therefore, in terms of aggregate benefits, the implied annual medium-term benefits for Line 1 (Lines 2, 3 and 7) are 5-10 times (10 times) the short-term benefits.

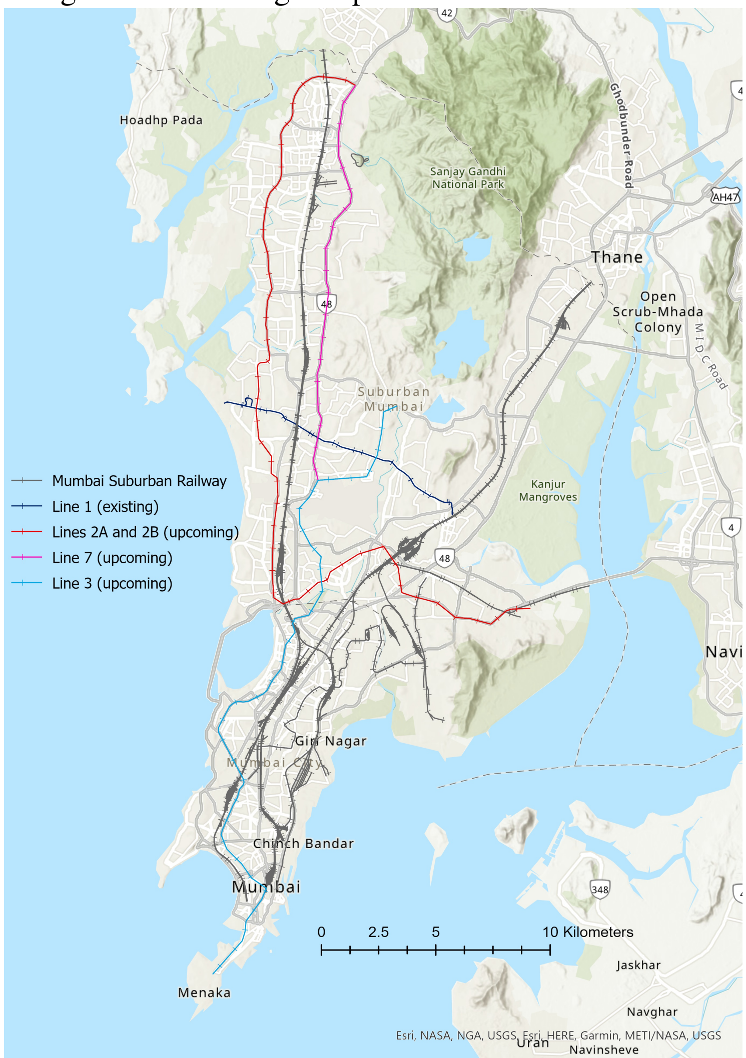
To compute short-term benefits, I use estimated preferences of commuters for in-vehicle travel time, out-of-vehicle travel time, travel costs and mean preferences for commute mode categories while assuming fixed residence and work locations. I compute medium-term benefits using estimated preferences of households for the commute time of primary workers, average unobserved mode-specific preferences, observed amenities and average unobserved housing-specific preferences while assuming fixed work location. Household sorting allowed by this model leads to the aggregate value of medium-term benefits being substantially higher than the short-term benefits.

There are important spatial and demographic heterogeneities in the benefits. Consistent with findings in the limited literature evaluating heterogeneity, I find that workers living close to the Metro stations, women, workers with college education and high incomes receive disproportionately greater short-term benefits than their counterparts. This is due to a difference in preferences as each Metro project generates similar travel time reductions for each group. Medium-term benefits exhibit a similar pattern. The benefits of Lines 2, 3 and 7 are more equally distributed than Line 1 because of their wider accessibility. But the benefits per km are higher for Line 1.

The aggregate annual medium-term benefits due to each project outweigh the equivalent annualized capital costs. This is despite the fact that these models capture only direct commuting-related welfare. To capture additional effects of Metro rail in Mumbai, Chapter 2 analyzes the effects of Metro Line 1 on property values and the spatial structure of economic activity in the city. I leave the estimation of a full general equilibrium model similar to the models in Redding and Rossi-Hansberg (2017) to future work because of data limitations.

1.8 Figures and Tables

Figure 1.1: Existing and planned rail lines in Mumbai



This map shows the existing Suburban railway network of Mumbai in grey, along with the existing Metro Line 1 in blue, and the three upcoming Metro Lines 2, 3 and 7 that are the focus of this paper in red, aqua and magenta, respectively.

Figure 1.2: Administrative zones in Mumbai with existing rail lines

This map shows the 6 administrative zones of Mumbai, the habitable areas, major roads, and the existing rail network. Municipal wards within each zone are shown in Appendix Figure A.1.

Figure 1.3: Population density
(People per sqkm)

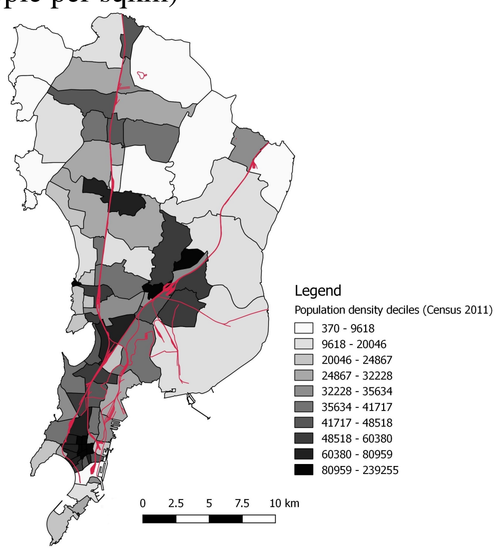
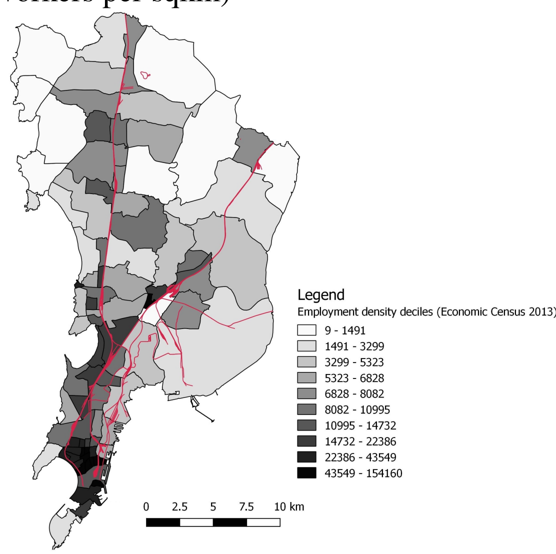
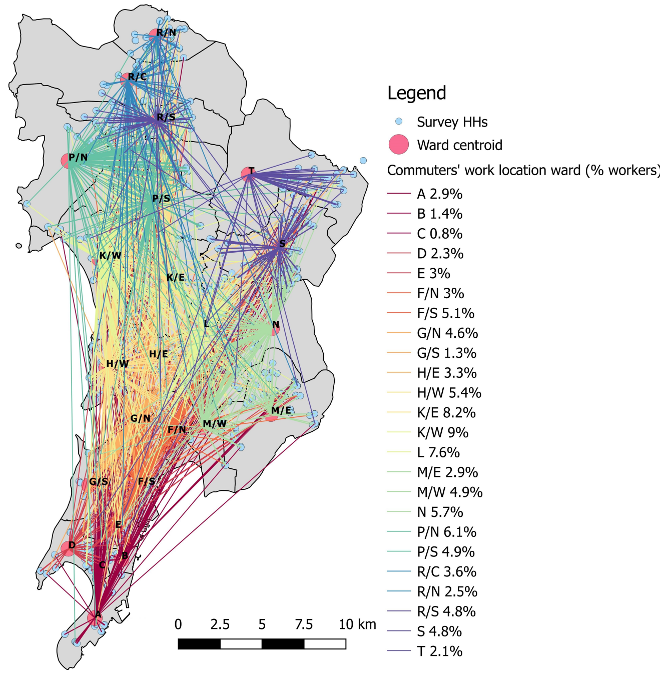


Figure 1.4: Employment density
(Workers per sqkm)



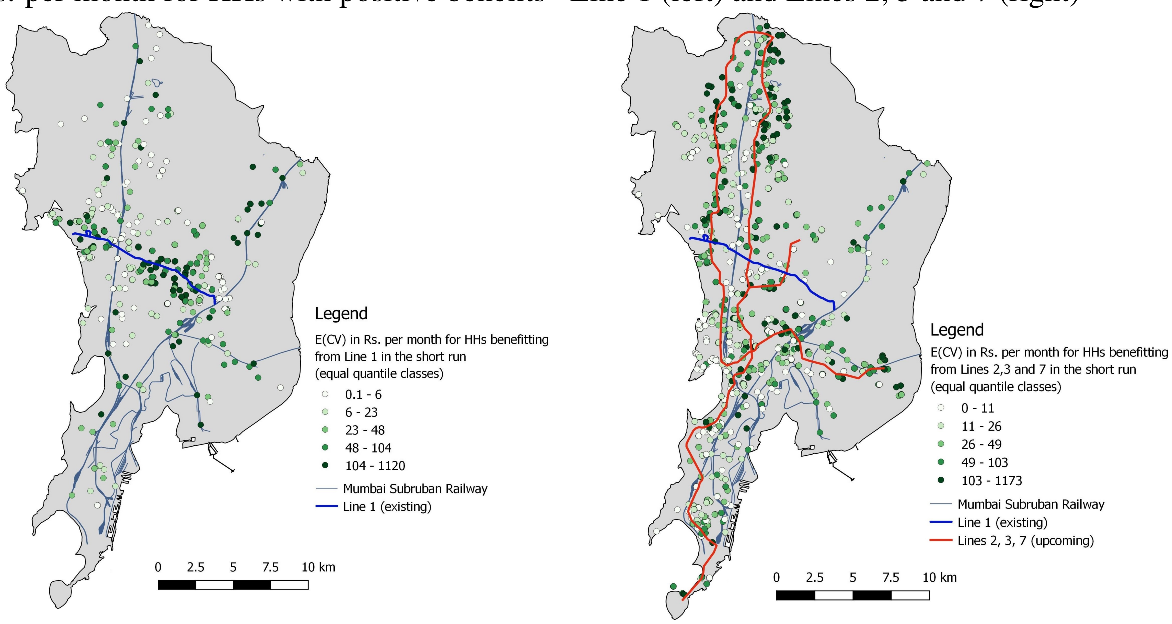
These maps show the 88 administrative sections in Mumbai shaded to reflect deciles of population per sqkm and workers per sqkm. Population density is calculated using population figures from Census 2011. Employment density is calculated using the number of workers employed in formal establishments from the Economic Census 2013. Area of each section is calculated using a digitized map of the city. The Suburban railway network of Mumbai is in red.

Figure 1.5: Sampled commuting workers' work location



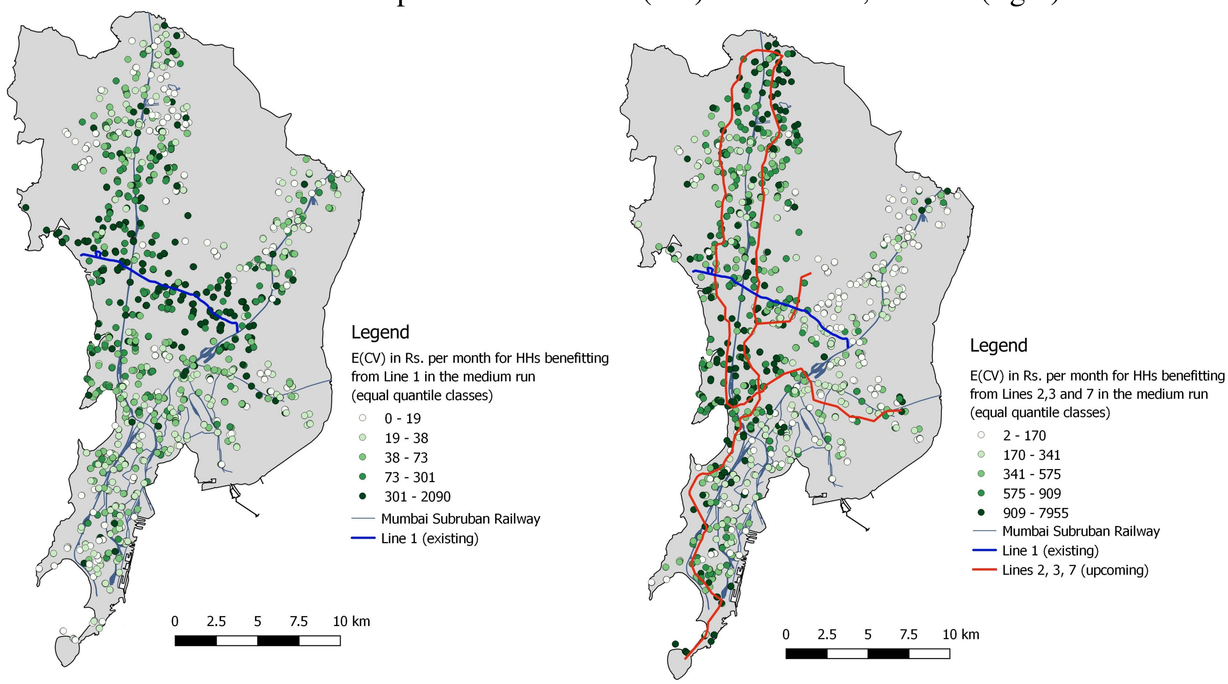
This map shows straight lines between household location and the centroid of their work location ward. Length of each line indicates the Euclidean distance in km. The share of workers traveling to each ward is indicated in the legend.

Figure 1.6: Spatial variation in the value of short-term benefits from the mode choice model in Rs. per month for HHs with positive benefits– Line 1 (left) and Lines 2, 3 and 7 (right)



These maps show the sampled households with positive expected compensating variation computed using the formula in equation 1.5.

Figure 1.7: Spatial variation in the value of medium-term benefits from the combined housing and mode choice model in Rs. per month– Line 1 (left) and Lines 2, 3 and 7 (right)



These maps show the sampled households' expected compensating variation computed using the formula in equation 1.10.

Table 1.1: Commuting pattern in Mumbai by Zones

Residence Zone	Work Zone 1	2	3	4	5	6	Total	Residence Zone % of residents
1	**72**	15.5	8.2	0.9	1.3	2.2	100	8.4
2	11.3	**67.6**	12.6	1.3	6.4	0.8	100	14.1
3	2.5	1.9	**81.9**	11.1	2.2	0.5	100	21.5
4	1.1	1.4	14.6	**80.8**	1.8	0.3	100	23.7
5	4.3	8.0	9.0	1.8	**69.7**	7.2	100	18.5
6	3.4	4.7	4.7	2.6	6.5	**78.1**	100	13.9
Zone % of workers	9.7	13.7	25.8	22.5	15.7	12.7	100	

This Table shows the commuting pattern of workers residing in different zones. The table is based on a sample of 2,767 workers for whom information is available to estimate the mode choice model in Section 1.4.

Table 1.2: Main mode chosen for commute to work– Shares in %

	Full sample	By worker gender		By worker's educ level		HH's vehicle ownership		By worker income	
		Men	Women	Below College	College	Does not own	Owns	$\leq$ median	$>$ median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Walk	32.2	30.4	42.3	40.4	13.9	48.8	14.9	37.8	15.9
Train	15.7	15.4	17.3	13.7	20.0	23.6	7.4	16.2	14.0
Bus	8.5	7.8	12.4	9.4	6.4	14.5	2.2	10.6	2.3
Auto-rickshaw	9.4	8.2	15.8	9.6	8.9	13.2	5.4	11.0	4.7
Own two-wheeler	29.9	33.5	9.3	25.5	39.9		61.2	23.1	49.8
Car	4.4	4.6	2.9	1.5	10.9		8.9	1.3	13.4
Observations	2,767	2,356	411	1,912	855	1,414	1,353	2,058	709
Mean distance (in km)	4.4	4.4	4.2	3.8	5.5	4.2	4.5	4.2	4.9

This Table shows the commute mode shares for different sub-groups of individuals in the sample. Mean distance reported here is the distance along the shortest path from commuter residence to a randomly chosen post office in the survey-reported pin code of their work location. It is computed using the network program and the map of road network. The mode 'bicycle' is included in the category 'walk' because of the small share of individuals whose main commute mode is 'bicycle'.

Table 1.3: Summary statistics by chosen travel mode for the sample used to estimate the commute mode choice models

Chosen travel mode	Mean	Std. Dev.	Min	Max
Road distance from residence to work location in km				
Walk	1.78	1.08	0.06	5.856
Train	9.73	6.85	0.29	28.58
Bus	4.73	3.69	0.20	17.98
Auto-rickshaw	3.62	4.45	0.14	30.42
Own two-wheeler	4.24	4.48	0.04	26.89
Car	5.89	5.14	0.19	22.98
Full Sample	**4.36**	**4.97**	**0.04**	**30.42**
In-vehicle time (IVT) in minutes				
Walk	0	0	0	0
Train	33.72	18.27	1.39	109.6
Bus	36.99	21.68	5	134
Auto-rickshaw	15.94	13.61	2.62	95.2
Own two-wheeler	17.8	12.9	2.5	74.93
Car	20.91	16.41	2.95	74.55
Full Sample	**16.15**	**18.48**	**0**	**134**
Out-of-vehicle time (OVT) in minutes				
Walk	21.3	12.92	0.70	70.27
Train	15.93	9.26	0.99	50.62
Bus	4.10	2.74	0.30	14.44
Auto-rickshaw	6.85	3.33	5	20
Own two-wheeler	0	0	0	0
Car	0	0	0	0
Full Sample	**10.34**	**12.39**	**0**	**70.27**
Cost of one-way trip (c) in Rs.				
Walk	0	0	0	0
Train	5.79	1.82	5	10
Bus	14.31	4.83	8	28
Auto-rickshaw	58.86	60.01	10.18	394.7
Own two-wheeler	13.64	12.29	1.65	84.46
Car	51.85	44.29	6.92	216.7
Full Sample	**13.98**	**28.17**	**0**	**394.7**
Average monthly income in Rs.				
Walk	18,659	10,587	2,500	75,000
Train	21,882	11,945	7,500	75,000
Bus	17,287	6,615	2,500	37,500
Auto-rickshaw	18,813	7,864	7,500	37,500
Own two-wheeler	28,379	16,129	7,500	125,000
Car	44,897	24,297	7,500	125,000
Full Sample	**23,117**	**14,544**	**2,500**	**125,000**

This Table presents summary statistics of variables used in the estimation of mode choice model in Section 1.4 for the estimation sample with 2,767 workers. Bicycle is included in the category 'walk' since the share of commuters who bicycle is very small. Road distance is computed using network program. In-vehicle time for train is from the network program; for bus, it is from HERE Transit API and Google Maps API; and for the remaining options, it is from Google Maps API. Out-of-vehicle time for walk, train and bus are from the network program; for auto-rickshaw, it is from the survey. Cost is computed using 2019 fare rules. The income variable in the survey is a categorical variable (Alam et al. (2021)); average income is computed using the median value of each category.

Table 1.4: Commute mode choice models estimated under Nested logit assumption: comparing three different nesting structures

	Model 1	Model 2	Model 3
Income-Cost	0.025***	0.011***	0.026***
	(0.002)	(0.001)	(0.001)
IVT	-0.016***	-0.016***	-0.011**
	(0.003)	(0.003)	(0.003)
OVT	-0.034***	-0.049***	-0.036***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Train, Bus)	Omitted		
(Train, Bus, Auto-rickshaw)		Omitted	
(Car, Two-wheeler)	2.096***		0.719***
	(0.087)		(0.077)
(Walk, Auto-rickshaw)	1.271***		Omitted
	(0.096)		
(Walk, Car, Two-wheeler)		1.896***	
		(0.064)	
(Train)			-1.353***
			(0.086)
(Bus)			-2.037***
			(0.112)
Dissimilarity parameters:			
(Car, Two-wheeler)	1		1
	(constrained)		(constrained)
(Bus, Train)	0.632***		
	(0.093)		
(Bus, Train, Auto-rickshaw)		1	
		(constrained)	
(Walk,Auto-rickshaw)	0.643***		0.604***
	(0.069)		(0.058)
(Walk, Car, Two-wheeler)		1.036***	
		(0.081)	
(Train)			1.000
			(286146.890)
(Bus)			1.000
			(419923.363)
Individuals	2767	2767	2767
LR chi2	304	502	353.8
Log likelihood	-2894.3	-2824	-2866.6
Value of IVT (Rs. per minute)	0.65	1.41	0.41
Value of OVT (Rs. per minute)	1.35	4.34	1.38
Value of IVT (% wage)	33	72	20
Value of OVT (% wage)	70	224	72

This Table presents estimated preference parameters for the nested logit model in equation 1.1. Std. errors are in parentheses. IVT and OVT are per trip in-vehicle and out-of-vehicle times (in minutes), respectively. Income-Cost is the value of monthly Hicksian bundle scaled to per trip level. It is obtained by subtracting monthly out-of-pocket travel cost from monthly income and dividing by the number of working days in a month (22) and number of trips in a day (2). Nesting structure in Model 1: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus); Model 2: (Walk, Car, Two-wheeler), (Train, Bus, Auto-rickshaw); Model 3: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus). Dissimilarity parameter is constrained to be ≤ 1 so that predictions are consistent with equation 1.3. 'Walk' also includes 'bicycle'.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.5: Predicted mode shares of nested logit models from Table 1.4

Travel modes	True shares	Model 1	Model 2	Model 3
Walk	32.2	28.1	34.3	27.4
Train	15.6	12.3	10.7	15.6
Bus	8.5	11.9	14.5	8.5
Auto-rickshaw	9.4	13.4	8.3	14.1
Two-wheeler	29.9	30.8	28.3	30.8
Car	4.4	3.5	3.9	3.5

This Table compares the predicted mode shares under the three nested logit models in Table 1.4 with the true sample shares. Nesting structure in Model 1: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus); Model 2: (Walk, Car, Two-wheeler), (Train, Bus, Auto-rickshaw); Model 3: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus).

Table 1.6: Sensitivity of the commute mode choice model parameters to different definitions of IVT and OVT for Model 1 in Table 1.4

	(1)	(2)	(3)
OVT definition	Survey+NetworkX	Survey+HERE (includes transfer time)	Survey
IVT definition	GM+HERE +NetworkX	GM+HERE (excluding transfer time)	GM+HERE +NetworkX
Income-Cost	0.025***	0.023***	0.023***
	(0.002)	(0.001)	(0.001)
IVT	-0.016***	-0.012***	-0.009***
	(0.003)	(0.002)	(0.002)
OVT	-0.034***	-0.027***	-0.026***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Train, Bus)	Omitted	Omitted	Omitted
(Car, Two-wheeler)	2.096***	1.947***	1.902***
	(0.087)	(0.078)	(0.074)
(Walk, Auto-rickshaw)	1.271***	1.615***	1.563***
	(0.096)	(0.084)	(0.080)
Dissimilarity parameters:			
(Car, Two-wheeler)	1	1	1
	(constrained)	(constrained)	(constrained)
(Bus, Train)	0.632***	0.475***	0.333***
	(0.093)	(0.059)	(0.037)
(Walk,Auto-rickshaw)	0.643***	0.139***	0.125***
	(0.069)	(0.015)	(0.013)
Individuals	2767	2754	2725
LR chi2	304.040	299.442	298.189
Log likelihood	-2894.262	-2855.071	-2791.160
IVT value (Rs. per minute)	0.651	0.544	0.380
OVT value (Rs. per minute)	1.348	1.197	1.157
Value of IVT (% wage)	33	28	20
Value of OVT (% wage)	70	62	60
Mean IVT (minutes)	16.15	14.37	16.24
Mean OVT (minutes)	10.34	16.60	15.47

This Table presents estimated preference parameters for the nested logit model in equation 1.1 for different definitions of in-vehicle time and out-of-vehicle time. In-vehicle time in Columns (1) and (3) for train is from the network program; for bus, it is from HERE Transit API and Google Maps API; and for the remaining options, it is from Google Maps API. In these two columns, out-of-vehicle time measures the first mile access. In Column (1), out-of-vehicle time for walk, train and bus are from the network program; for auto-rickshaw, it is from the survey. In Column (3), out-of-vehicle time is from the survey. Column (2) is the same as Column (1) except that out-of-vehicle time includes transfer time for bus and train from HERE API, and the same is excluded from in-vehicle time. Std. errors are in parentheses. Estimated parameters are based on the nesting structure in Model 1 of Table 1.4: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus). Estimated parameters based on Model 3 are in Appendix Table A.1. Dissimilarity parameter is constrained to be $\leq$1 so that predictions are consistent with equation 1.3.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.7: Predicted mode shares of nested logit models from Table 1.6 and Appendix Table A.1

		(1)	(2)	(3)
OVT definition		Survey+NetworkX	Survey+HERE (includes transfer time)	Survey
IVT definition		GM+HERE +NetworkX	GM+HERE (excluding transfer time)	GM+HERE +NetworkX
Commute modes	True shares	Predicted shares	Predicted shares	Predicted shares
Nested logit Model 1:				
Walk	32.2	28.1	29.7	29.7
Train	15.6	12.3	12.5	12.8
Bus	8.5	11.9	11.7	11.6
Auto-rickshaw	9.4	13.4	11.8	11.6
Two-wheeler	29.9	30.8	30.7	30.8
Car	4.4	3.5	3.6	3.6
Nested logit Model 3:				
Walk	32.2	27.4	29.6	29.6
Train	15.6	15.6	15.6	15.7
Bus	8.5	8.5	8.5	8.6
Auto-rickshaw	9.4	14.1	11.9	11.7
Two-wheeler	29.9	30.8	30.8	30.9
Car	4.4	3.5	3.5	3.5

This Table compares the predicted mode shares from specifications using different definitions of in-vehicle and out-of-vehicle travel times for two nesting structures with true shares. Estimated parameters are in Table 1.6 and Appendix Table A.1.

Table 1.8: Heterogeneity in the value of short-term time savings implied by nested logit models

	Full sample	Men	Women	$<$ College education	$\geq$ College education	$\leq$median income	$>$ median income
Nested logit Model 1:							
Individuals	2767	2356	411	1912	855	2058	709
IVT Value (Rs./min)	0.65	0.69	0.53	0.45	1.3	0.06	1.77
IVT Value (% of wage)	33.47	33.49	39.62	27.84	49.04	4.52	48.28
OVT Value (Rs./min)	1.35	1.31	1.51	1.2	1.94	0.6	1.83
OVT Value (% of wage)	69.27	63.91	113.04	73.36	73.44	44.54	49.76
% sample with positive E(CV) Line 1	23.31	23.26	23.6	20.35	30.29	22.93	24.54
Mean E(CV) Line 1—Positive E(CV) (Rs./month)	70.87	68.09	84.57	56.9	116.72	38.22	89.87
% sample with positive E(CV) Lines 2,3,7	56.45	56.71	54.99	55.33	59.77	56.41	56.56
Mean E(CV) Lines 2,3,7—Positive E(CV) (Rs./month)	85.12	84.51	98.14	71.9	131.41	46.82	86.61
Nested logit Model 3:							
Individuals	2767	2356	411	1912	855	2058	709
IVT Value (Rs./min)	0.41	0.44	0.45	0.3	0.52	0.29	1.16
IVT Value (% of wage)	21.27	21.56	33.74	18.53	19.58	21.23	31.6
OVT Value (Rs./min)	1.38	1.36	1.56	1.22	2.05	1.26	1.86
OVT Value (% of wage)	71.07	66.29	116.81	74.77	77.58	93.32	50.62
% sample with positive E(CV) Line 1	23.42	23.39	23.6	20.35	30.29	22.93	24.82
Mean E(CV) Line 1—Positive E(CV) (Rs./month)	96.99	94.26	115.36	72.73	169.34	81.87	165.11
% sample with positive E(CV) Lines 2,3,7	56.7	57	54.99	55.33	59.77	56.41	57.55
Mean E(CV) Lines 2,3,7—Positive E(CV) (Rs./month)	98.71	99.05	115.35	81.49	146.59	93.72	113

This Table presents the marginal rate of substitution and the mean expected compensating variation for Line 1 and Lines 2, 3 and 7 computed using estimated parameters from a nested logit model (equation 1.1) estimated separately for the subsamples indicated in the columns. Preference parameters for the full sample are in Table 1.4. Individuals indicate the number of individuals in each of these estimation samples. Nesting structure in Model 1: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus); Model 3: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus). There is limited statistical power to separately estimate the model for the subsamples by vehicle ownership; so, I summarize the mean E(CV) values obtained for the full sample by vehicle ownership to test heterogeneity in this case. Mean E(CV) for commuters without a vehicle is three times the value of those with a vehicle.

Table 1.9: Summary of chosen mode categories for the sample used to estimate the combined housing and mode choice model

				Mode Category Shares (in %)				
		By primary worker educ			**By HH income**		**By HH Vehicle ownership**	
Mode Category	**Full Sample**	$<$ College	$\geq$ College	$<$ median income	Median income	$>$ median income	No vehicle	Owns vehicle
Walk	29.5	36.9	12.5	38.9	21.1	10.9	45.1	13.9
Train	15.5	14.6	17.6	16.6	14.8	11.8	24.5	6.4
Road	55.0	48.5	69.9	44.4	64.0	77.3	30.4	79.7
				Travel Time (in minutes)				
		By primary worker educ			**By HH income**		**By HH Vehicle ownership**	
Mode Category	**Full Sample**	$<$ College	$\geq$ College	$<$ median income	Median income	$>$ median income	No vehicle	Owns vehicle
Walk	23.7	24.0	21.7	25.0	19.9	24.2	23.4	24.5
Train	34.3	34.2	34.6	34.6	33.8	34.3	34.3	34.4
Road	14.0	12.9	15.7	13.1	14.7	14.5	13.5	14.2
Average	20	19.8	20.1	21.3	18.6	17.9	23	16.9
				Travel distance (in km)				
		By primary worker educ			**By HH income**		**By HH Vehicle ownership**	
Mode Category	**Full Sample**	$<$ College	$\geq$ College	$<$ median income	Median income	$>$ median income	No vehicle	Owns vehicle
Walk	2.0	2.0	1.8	2.1	1.7	2.0	2.0	2.0
Train	10.7	10.4	11.3	10.7	10.4	11.4	10.6	11.0
Road	4.7	4.3	5.2	4.4	4.9	4.8	4.5	4.7
Average	4.8	4.3	5.9	4.5	5	5.3	4.8	4.8

This Table shows the commute mode category shares, travel time and distance for different sub-groups of households in the sample of 2,209 households used to estimate the combined housing and mode choice model in Section 1.5. Bus, auto-rickshaw, taxi, two-wheeler and car are included in Road transport. Walk includes bicycle. Travel time for all modes is from the network program.

Table 1.10: Summary statistics for the sample used to estimate the combined housing and mode choice model

Variables	Mean	Std. Dev.
Household characteristics		
Income in Rs.	31,330	20,582
Main religion: Hindu	0.79	
Main religion: Muslim	0.17	
Main religion: Other	0.04	
Main language: Hindi	0.53	
Main language: Marathi	0.36	
Main language: Gujarati	0.06	
Main language: Others	0.05	
Households in the neighborhood with same religion	0.68	
Households in the neighborhood with same language	0.45	
Housing characteristics		
Monthly rental price in Rs.	9,742	7,140
Distance to nearest railway station in km	1.48	1.12
Standardized employment accessibility index	0	1
Floorspace in sqft.	262.26	162.80
Good Roof	0.71	
Number of rooms (Median)	1	0.59
Kitchen is separate	0.59	
Toilet inside the house	0.65	
Bathroom inside the house	0.76	
Piped water	0.76	
Footpath in the neighborhood	0.75	
Slum Classification	0.44	
Distance to coast (in km)	3.57	2.48
Reports of crimes against women (Median)	35	18.26
Walk time to the nearest pvt. doctor	8.09	6.04
Walk time to the nearest govt. hospital	19.97	8.85
Walk time to the nearest pvt. hospital	17.36	8.83

This Table presents summary statistics of variables used in the estimation of housing and commute mode choice model in Section 1.5. No. of Households= 2,209, except in the case of last three variables which are for sample sizes 2,167, 2,181 and 2,080, respectively. Standard deviation is not shown for binary variables. 'Other' religions include Christianity, Sikhism, Jainism, Buddhism, and Zoroastrianism (Parsi). 'Other' languages include Tamil, Telugu, Marwari, Kannada, Konkani, Punjabi, Sindhi, English, Bengali, Bhojpuri, and Odia. Proportion of households with the same language and religion are defined within a 2 km radius around the household's location. Mean walk time to health facilities is calculated by averaging the median of categories with survey-reported times. Mean monthly household income Rs 31,330 = $1,472 (PPP); mean monthly rent Rs. 9,742 = $458 (PPP).

Table 1.11: Combined housing and commute mode choice model– Preference estimates from First stage

	Parameter estimates
Proximity to work location (in minutes)	-.09139***
	(0.0019)
Proportion of HHs with same language	1.997***
	(0.2407)
Proportion of HHs with same religion	2.809***
	(0.3781)
Mean preferences for mode category:	
Walk	2.634***
	(0.0712)
Road transport	0.6623***
	(0.0642)
Individuals	2,209
Wald chi2	3953
Log likelihood	-8895

This Table presents estimated preference parameters for the first-stage (equation 1.6) of the combined housing and mode choice model in Section 1.5.1. Standard errors are in parentheses. Proximity to work location is the one-way travel time in minutes via walking, rail or using road transport from the network program. Proportion of HHs with the same language and religion as the chooser are defined within a 2 km neighborhood around the house. Walking is not a feasible alternative whenever distance between a chooser's work location and a house in the feasible set is greater than 7 km. In estimating mode-specific intercepts, train is the omitted category.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 1.12: Second stage 2SLS regression of mean preferences for housing on amenities

	(1)	(2)	(3)	(4)	(5)
Rental Price	-0.000124***	-0.000117***	-0.000090***	-0.000090***	-0.000115***
	(0.00003)	(0.00004)	(0.00003)	(0.00003)	(0.00003)
	[-3.586]	[-3.300]	[-3.054]	[-3.246]	[-4.041]
Housing amenities index	0.356153***	0.346996***	0.299431***	0.301683***	0.348696***
	(0.07261)	(0.07247)	(0.06074)	(0.05747)	(0.06346)
	[4.905]	[4.788]	[4.930]	[5.250]	[5.495]
Distance to coast in km		0.027921	0.001788	-0.048086***	-0.034070**
		(0.01797)	(0.01345)	(0.01655)	(0.01635)
		[1.554]	[0.133]	[-2.905]	[-2.084]
Slum classification dummy		-0.147291**	-0.104903*	-0.109721**	-0.113037**
		(0.07323)	(0.05688)	(0.05254)	(0.05709)
		[-2.011]	[-1.844]	[-2.088]	[-1.980]
Crimes against women (No. of reports)		-0.003408	-0.006329**	-0.006684***	-0.003547
		(0.00320)	(0.00246)	(0.00224)	(0.00252)
		[-1.063]	[-2.571]	[-2.983]	[-1.407]
Distance to nearest station in km			0.312727***	0.181853***	0.141643***
			(0.03240)	(0.04337)	(0.04657)
			[9.652]	[4.193]	[3.042]
Standardized employment accessibility index				0.314743***	0.300498***
				(0.06103)	(0.06216)
				[5.157]	[4.834]
Index for proximity to doctor					0.018388
					(0.02698)
					[0.682]
F(excluded IV)	40.601	39.133	37.765	38.388	38.768
Observations	2,209	2,209	2,209	2,209	2,024
Critical value for —t— at 95% level (Lee et al. (2021))	2.22	2.23	2.24	2.24	2.24

This Table presents 2SLS parameter estimates of the second stage of the combined housing and mode choice model (equation 1.8, Section 1.5.1). Robust std. errors clustered at the sub-zone level are in parentheses. t-statistics are in brackets. Dependent variable is the vector of estimated intercepts from the first stage conditional logit model presented in Table 1.11). Log of assessed land value for residential land in the sub-zone of a house is used as an instrument for monthly rental price in Rs. To provide evidence for instrument strength, critical t-values using adjusted standard errors are noted in the last row following Lee et al. (2021) for valid inference at the 95% level. Star marks reflect conventional inference values.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.13: Preferences for commute time from the combined housing and mode choice models with taste shifters

	(1)	(2)	(3)	(4)
Base category	-0.09***	-0.09***	-0.09***	-0.08***
	(0.002)	(0.002)	(0.002)	(0.002)
$\geq$ college educ		-0.004		
		(0.004)		
median HH income			-0.02***	
			(.004)	
$\geq$ median HH income			-0.02***	
			(0.006)	
HH owns vehicle				-0.04***
				(0.004)
Mode constant	Yes	Yes	Yes	Yes
House constant	Yes	Yes	Yes	Yes
Taste shifter	None	Educ	Income	Vehicle ownership
Households	2209	2209	2209	2209

This Table shows preferences for travel duration estimated in the first-stage of the combined housing and mode choice model using specifications with taste-shifters. Proximity to work location is the one-way travel time in minutes via walking, train or using road transportation. Walking is not a feasible alternative whenever distance between a chooser's work location and a house in the feasible set is greater than 7 km. Column (1) is the base model without any taste-shifters. Base category is households where the primary worker has below college education in Column (2); below median income household in Column (3); and households that do not own a vehicle in Column (4). Parameters are not shown here but all specifications account for preferences for the proportion of HHs with the same language and religion within a 2 km radius of the feasible house.

Table 1.14: Mean expected compensating variation for time savings due to Metro rail implied by various second-stage specifications

	(1)	(2)	(3)	(4)
Mean E(CV) Line 1 in Rs. per month	137	145	189	189
Mean E(CV) Line 1 (% mean monthly rent)	1.4	1.5	2	2
Mean E(CV) Lines 2, 3, 7 in Rs. per month	448	475	618	618
Mean E(CV) Lines 2, 3, 7 (% mean monthly rent)	4.6	4.9	6.3	6.3
Controls:				
Housing amenities index	✓	✓	✓	✓
Distance to coast in km	✗	✓	✓	✓
Slum classification dummy	✗	✓	✓	✓
Crimes against women (Reports)	✗	✓	✓	✓
Distance to nearest station	✗	✗	✓	✓
Employment accessibility index	✗	✗	✗	✓

This Table presents the mean expected compensating variation for Line 1 and Lines 2, 3 and 7 (equation 1.10) computed using the first stage estimates in Table 1.11 and various second-stage specifications (Table 1.12), the controls for which are indicated.

Table 1.15: Mean expected compensating variation sensitivity to first-stage specification (in Rs. per month)

	Line 1	Lines 2, 3, 7
Main Model	189	618
Models with taste shifters:		
Income	220	709
Education	186	606
Vehicle ownership	179	586

This Table presents the mean expected compensating variation for Line 1 and Lines 2, 3 and 7 for first-stage specifications with and without taste-shifters (Table 1.13) and using the preferred second-stage specification (Column (4) of Table 1.12). Main model is the one without any taste-shifters.

Table 1.16: Heterogeneity in the value of medium-term time savings implied by the combined housing and mode choice model

Group	Line 1 (in Rs.)	Lines 2,3, 7 (in Rs.)
Full sample	189	618
Primary worker has $<$ college education	167	587
Primary worker has $\geq$ college education	238	688
Below median income HH	173	600
Median income HH	214	637
Above median income HH	183	642
HH does not own vehicle	185	606
HH Owns vehicle	197	640

This Table presents the mean expected compensating variation for Line 1 and Lines 2, 3 and 7 for sub-groups of households indicated in the rows. Expected compensating variation is computed using the first-stage specification of the combined housing and mode choice model without taste-shifters (Table A.5) and using the preferred second-stage specification (Column (4) of Table 1.12) for the full sample with 2,209 households and averages are calculated for each sub-group.

Table 1.17: Aggregate benefits of Metro lines compared to equivalent annualized capital costs (EACC) in $ Billion (PPP)

	EACC			**Short-term Benefits**		**Medium-term Benefits**
	(1)	(2)	(3)	Model 1	Model 3	
Line 1	0.3	0.2	0.2	0.1	0.2	1
Lines 2, 3 and 7	3	2.3	1.9	0.3	0.3	3.2
EACC assumptions:						
Life of asset (years)	20	30	35			
Interest rate (in %)	12	10	8			

This Table presents the equivalent annualized capital costs under three assumptions indicated in the second panel; and the annual aggregate short-term and medium-term benefits of Metro Line 1 and Lines 2, 3 and 7 computed by scaling the individual-level estimates. Model 1 and 3 correspond to the commute mode choice models with nesting structures in Columns 1 and 3 of Table 1.4. Total construction cost of Line 1 is $2.03 Billion (PPP). Total *projected* cost of Lines 2, 3 and 7 is $22 Billion (PPP). Exchange rate: $1 PPP = Rs. 21.283

Chapter 2: What are the Benefits of a Subway in Mumbai, India?

2.1 Introduction

There is a large literature in urban economics on the benefits of transport projects using ex-ante and ex-post methods of evaluation. Part of this literature estimates the effects of transport projects on land values. Such estimates provide an ex-post measure of the benefits realized by individuals since net benefits of transport infrastructure get capitalized into property values as individuals internalize potential present and future benefits and disbenefits of the public good. They also provide a measure of the impact on property tax revenues. We extend this literature by measuring the impact of the first line of the Mumbai Metro on property values and then using insights from the new urban economics literature (Ahlfeldt et al. (2015), Redding and Turner (2015), Redding and Rossi-Hansberg (2017) Heblich et al. (2020), Tsivanidis (2019), Warnes (2020), Kreindler and Miyauchi (2021)) to explain the sources of these benefits.

Line 1 of the Mumbai Metro became operational in 2014. It is only 11.4 km in length but provided the first east-west rail link in a city with an extensive Suburban Rail network. We measure its impact on property prices using administrative data on assessed values in 726 sub-zones of the city, from 2011 to 2018, for various land-use categories: residential, commercial shop, commercial office, industrial, and open-use land. Using difference-in-differences in an event study framework, we compare areas that are within 1 km of Line 1 with control areas that are beyond 1 km but within 3 km of Line 1 before and after the opening of Line 1. We find that two years before the opening of Line 1, the change in property prices in areas within 1 km of Line 1 was 5-6% higher than the change in areas between 1 and 3 km from Line 1, reflecting anticipatory effects of the policy. After 2014, prices for commercial and residential properties within 1 km of the Metro increased between 13% and 17% relative to the control group.

The assumption implicit in our analysis is that the change in prices over time would have been the same in sub-zones up to 3 km from Line 1 in the absence of Line 1. To test the sensitivity of our results, we repeat this analysis with different treatment and control group definitions. We find that the magnitude of effects declines as the treatment area is expanded to include sub-zones further away from Line 1, suggesting that the influence of Line 1 is positively associated with spatial proximity to the line.

We then examine the source of these benefits. An important source of benefits is the time savings to commuters that Line 1 provides. Chapter 1 has studied the magnitude of these savings using a static commute mode choice model that assumes fixed residential and job locations and a commute mode choice model that allows households to adjust their residential locations but holds job locations constant. The estimated city-level annualized value of commute time savings due to Line 1 through improvements in travel time and household resorting in Mumbai is $1 Billion PPP. Twenty percent of these estimated benefits accrue to households living in our treatment area.

These benefits, however, fail to account for the changes in the location of jobs that have occurred in Mumbai over time. We measure the impact of these changes on the benefits of Line 1 by computing a measure of changes in employment accessibility between 2004 and 2019. This measure is intuitively similar to the measure of changes in commuter market access, which captures general equilibrium changes in access to wages across the city as households and firms re-sort in response to new infrastructure projects. Changes in commuter market access have been shown as the major source of welfare in general equilibrium models in the context of the construction and removal of the Berlin wall (Ahlfeldt et al. (2015)), the introduction of the steam railway in London (Heblich et al. (2020)), the introduction of Bus Rapid Transit (BRT) in Bogota (Tsivanidis (2019)), and the introduction of BRT in Buenos Aires (Warnes (2020)). Our approach provides a reduced-form way to relate the spatial changes implied by the canonical spatial equilibrium models to the net benefits observed in property price appreciation, which is particularly useful in data sparse settings.

Our measure of employment accessibility is a commuting-cost-weighted average of effective wages obtainable across locations within the city from a given sub-zone. Lacking detailed information on wages, we follow the approach in Kreindler and Miyauchi (2021) to infer relative wages from commute flows. We obtain data on commute flows from two representative household surveys of travel demand conducted by the World Bank in 2004 and 2019. Intuitively, commute flows between a pair of locations are a function of origin or residential location amenities, destination or work location amenities, and the cost of commuting between the origin-destination pair. In a regression of commute flows on commute time, origin fixed effects, and destination fixed effects, the estimate of destination fixed effects describes the relative work location amenities, or effective wages, across the city.[1]

We demonstrate that the benefits of Line 1 are slightly higher in treatment areas that have experienced the largest improvements in employment accessibility between 2004 and 2019. Improvements in employment accessibility may reflect reductions in commuting times to high-wage locations and/or increases in effective wages in more accessible locations, holding com-

[1] Similarly, the estimated vector of origin fixed effects describes relative residential location amenities.

muting times constant. To the extent that changes in employment accessibility are due to factors unrelated to Metro Line 1, these results underscore the importance of the strategic placement of transport investments.[2] We emphasize this by evaluating the impact of the Mumbai Monorail (8.9 km), which also opened in 2014, on property prices. We fail to find any significant changes in property prices due to Monorail, which is explained by the location of Monorail— in the southeastern part of the city in an area that has not experienced significant improvements in employment accessibility.

2.2 Context

Mumbai is the financial capital of India and one of the most densely populated cities in the world. In the Greater Mumbai Region, 12.5 million people (Census, 2011) live in an area of 603.4 sqkm.[3] Mumbai faces enormous challenges with shortages of land, housing, infrastructure, and social services that have not kept up with the growing demands of the city. Some of Asia's largest slums, including Dharavi, with a population of over one million, are located in Mumbai. An estimated 42% of the city's poor lives in slums, many located along railway tracks (Census 2011).

Mumbai has an extensive rail network. Mumbai Suburban Railway is comprised of 100 km of track in the Greater Mumbai Region and 465 km across the entire Mumbai Metropolitan Region (see Figure 2.1). It is one of the busiest commuter rail systems in the world. Suburban trains are, however, faced with an acute overcrowding problem– there are about 14-16 passengers

[2] The similarity between our main estimates and those obtained when using treatment and control areas that experienced similar changes in location amenities and employment accessibility also suggests the validity of our empirical strategy.

[3] The Greater Mumbai Region is the core of the larger Mumbai Metropolitan Region, which has a population of 22.88 million in an area of 6,355 sqkm.

per sqm of floor space during typical rush hour times (Hindustan Times (2017)). Although some AC trains have been introduced since December 2017, few of Mumbai's suburban trains are air-conditioned. Mumbai also has an extensive network of public buses that complements the rail system.

The 2011 Census reports that 50% of Mumbai's commuters used either rail or bus to get to work (Table 2.1); however, this share has been falling. As incomes have risen, commuters have switched to private vehicles. Between 2000 and 2017 the number of two-wheelers in Mumbai increased by 340%; the number of cars increased by 200%.[4] This has led to huge traffic congestion problems in the city, and a fall in bus ridership. Bus ridership declined from 4.2 million per day in 1997-98 (Korde (2018)) to 2 million in 2019 (DNA India (2019)).

An extensive metro rail project comprised of 300 km of metro lines has been planned to alleviate Mumbai's congestion problems in an environmentally friendly way (Chacko (2018)). Metro Line 1 opened in 2014 and an additional 92 km of Metro are currently under construction. Metro Line 1, shown in Figure 2.1, is the first east-west rail link in the city. A monorail of 20 km has also been planned for the southeastern part of Mumbai. The first line of the monorail, 8.9 km in length (see Figure 2.1), was opened to the public in 2014, four months before Metro Line 1. Pre-pandemic, the daily ridership of Monorail was about 19,000, in contrast to 450,000 for Metro Line 1.

It is important to put both infrastructure projects in the context of Mumbai's development. Historically, both population and employment in the GMR have been organized around the Suburban Rail lines, with maximum concentration in the city's southernmost part, where the

[4]There were 407,306 two-wheelers and 303,108 cars in Mumbai in 2000. Their population increased to 1,784,657 and 911,856 by 2017, respectively. (Source: Department of Motor Vehicles, Maharashtra)

original central business district was located. Over time, both population and employment have been moving northward, and there is a newly developing industrial center in the middle of the GMR. Line 1 goes through this region. Figures 2.2 and 2.3 show the population and employment density in the GMR according to the 2011 Census and the 2013 Economic Census, respectively. There is high population and employment density near Line 1, suggesting the potential benefits of this location.

We measure the magnitude of realized benefits by estimating the capitalization of Line 1 into property values. Each year, the Municipal Council of Greater Mumbai (MCGM) uses information on property transactions conducted during the year to set assessed property values for the following year for geographical sub-zones. The entire city is divided into over 726 sub-zones. The rationale is to bring assessed values close to market prices to maximize the city's earnings from transactions and property taxes. Known as the Ready Reckoner Rates, annual assessed values, in Rs. per sqm., are published for the following categories of properties, based on floor space use: open land, residential property, commercial office, commercial shop, and industrial property. In the absence of a consistent dataset on the market values of sale and rental prices of properties in Mumbai, Ready Reckoner rates are the best available proxy for property prices.

A summary of assessed prices in the city for different property types is in Table 2.2. Commercial shop is the most expensive floorspace category, followed by commercial office, industrial, residential, and open-use type.[5] Average residential land assessed price in 2015-18 was Rs. 165,393 per sqm (about $230 per sqft) while the highest price was Rs. 713,800 per sqm

[5] This is not true for the 2013-14 period for average residential prices for the entire city because of an outlier.

(about \$990 per sqft).[6] Average commercial shop price in the city during the same period was Rs. 257,905 per sqm, 56% higher than the average residential price. The increase in prices from 2011 and 2012 is higher for sub-zones within 1 km of Line 1 than for sub-zones at a greater distance. Yearly trends in these prices are shown in Figure 2.4 for sub-zones within 1 km of Line 1 and those beyond 1 but within 3 km of Line 1. The sharp divergence in prices after 2014 coincides with Line 1 becoming operational.

2.3 Effects of Line 1 on Property Prices

We study the net benefits of Line 1 by estimating the effects of the opening of Line 1 on property prices. We also estimate the anticipatory price effects of Line 1. The impacts of Line 1 on property prices are then contrasted with the impacts of the monorail on property values.

2.3.1 Data and Methods

To estimate the impact of Line 1 on property values we use georeferenced administrative data on assessed property prices for each sub-zone in Mumbai for the period 2006-2018. The Municipal Council of Greater Mumbai (MCGM) annually publishes detailed information on localities that fall under each sub-zone along with Ready Reckoner Rates. Maps of sub-zone boundaries were constructed manually based on this published information.[7] We use this information to compile a panel dataset of property prices for 726 consistently defined sub-zones in the city for the period 2011-2018.[8] Sub-zone boundaries are drawn based on historical factors

[6] Exchange rate used for this calculation: \$1=Rs. 67 which is the rough average for that period.

[7] Boundaries were digitized by AInsight Technologies Pvt. Ltd. https://www.ainsighttech.com

[8] Data for 2010 is not available. 2008-2009 have been excluded because the assessed prices were artificially controlled to mitigate the influence of the global recession. Since these prices are administratively set, some areas were likely more protected than others. 2006 was the first year in which the current system of assessing prices was

and regional policy requirements, which may differ each year. We harmonized the sub-zone boundaries over time manually starting with the boundaries in 2015 and matching boundaries in other years to them.[9]

We estimate the effects of Line 1 on property prices using a difference-in-differences framework that leverages the spatial and temporal variation in property prices across sub-zones. The benefits of Line 1 in terms of agglomeration economies and improved access to the rest of the city are likely to be greatest in areas close to the line.[10] To define our treatment and control sub-zones, we compute the shortest distance via the road network between each sub-zone centroid and Line 1 using network analysis in ArcGIS.[11] In our main analysis, sub-zones with a distance less than 1 km are defined as treated while those beyond 1 km and within 3 km are defined as controlled. This is because we believe that these sub-zones are most likely to be similar and satisfy the common trends assumption. We estimate the following regression specification.

$$\text{logP}_{st} = \alpha_s + \tau_t + \zeta * \text{Treated}_s * \text{Year 2013/2014}_t + \delta * \text{Treated}_s * \text{Post-2014}_t + \epsilon_{st} \qquad (2.1)$$

logP_{st} indicates the property price in sub-zone s and year t. α_s and τ_t represent sub-zone and year fixed effects, accounting for aggregate shocks at the sub-zone and year levels, respectively. Since the assessed prices in year t reflect market conditions in year t-1, we choose the years 2011 and 2012 as our reference period, years 2013 and 2014 as the period for measuring anticipatory effects, and years 2015-18 as the period for measuring the effects of Line 1 becoming operational.

put in place.

[9] We use QGIS 2.18.14 to manually fix errors in boundaries and perform the harmonization. Note that there are multiple ways of performing this harmonization and since there is no theoretically correct way to do this, we confirm the robustness of our conclusions to a few other matching strategies.

[10] Chapter 1 notes that most of the time savings benefits of Line 1 in the short run accrue through improvements in access times.

[11] For distance calculation, Line 1 is converted into nodes to calculate distance between two points.

ζ is the estimate of changes between treatment and control areas up to two years before the opening of Line 1 relative to 2011-12 and indicates the anticipatory effects of Line 1 on property prices. Our anticipatory effects are restricted to two years because of evidence in the literature that anticipatory effects generally do not show up more than two years before a transit project (Diao et al. (2017), Gupta et al. (2020)) and because of the observed trends in price changes (shown in Figure 2.5, discussed in the following subsection). δ is the main coefficient of interest representing the effect of the opening of Line 1 on prices. It captures the cumulative effects on prices. ϵ_{st} represents idiosyncratic shocks at the sub-zone-year level. Standard errors are clustered at the sub-zone and year levels.[12]

A spatial difference-in-differences analysis based on proximity to the metro cannot account for the changes in property prices that may have occurred due to improved commuter market access in areas beyond 1 km of the metro. But this empirical approach has the advantage of satisfying the identification assumptions more robustly as compared to an analysis of citywide properties which will inevitably be based on stronger identifying assumptions. With this in mind, we repeat the analysis with other treatment and control group definitions to learn more about the gradient of the effects on property prices.

Other definitions of treatment and control groups in our analysis include sub-zones within 1 km vs those between 1 and 5 km and sub-zones within 3 km vs those between 3 and 5 km. We also compare sub-zones within 1 km with all other sub-zones in the city beyond 1 km. We note that the possibility of general equilibrium effects in the property markets in the vicinity of Line 1 implies a possible violation of the stable unit treatment value assumption (SUTVA). While it is not possible to completely address this in a reduced-form spatial difference-in-differences

[12]Results are robust to clustering at a more aggregated neighborhood level than the sub-zone level.

framework, we believe that limiting the geographical scope of our analysis reduces the likelihood of this problem.

In future work, we will collect additional data to implement a synthetic difference-in-differences strategy (Arkhangelsky et al. (2021)). We will match sub-zones on pre-treatment characteristics and construct a synthetic control group for treated sub-zones. This will allow us to expand our treatment definition to areas that are not necessarily in the immediate vicinity of Line 1 but may have experienced improvements through labor or product market access.[13] Synthetic controls will also allow us to expand our control group definition when the treated sub-zones are those within 1 km of Line 1 and will thus serve as a robustness check on our main empirical strategy.

2.3.2 Results

We first examine the relative changes in property prices in treatment and control groups over time compared to the difference in 2011 (shown in Figures 2.5-2.8). Each point on these graphs shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following variation of equation 2.1.

$$\text{logP}_{st} = \alpha_s + \tau_t * \text{Year}_t + \beta_t * \text{Treated}_s * \text{Year}_t + \epsilon_{st} \tag{2.2}$$

[13]Another possible strategy that would allow us to do this is the recentered instrumental variables strategy of Borusyak and Hull (2020) that relies on randomization inference, but its suitability for our context is not unambiguous.

$\log P_{st}$ is the property price in sub-zone s and year t. α_s and τ_t represent sub-zone and year fixed effects as before. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). Treated_s is an indicator for treated sub-zones. β_t is the main coefficient of interest reflecting the relative difference between treated and control sub-zones in the year t relative to this difference in 2011. A plot of β_t is in Figure 2.5. We see that in 2012, the average difference in prices in sub-zones within 1 km and those beyond 1 km but within 3 km was no different than this difference in 2011. In 2013-14, prices in treatment areas were 5-6% higher and post-2014, prices were 10-25% higher, with the price increase being persistent over time. The price increase in 2013-14 represents the anticipatory effects of Line 1 and the increase post-2014 is the effect of opening of Line 1.

We also examine these trends for other control group definitions. Comparing sub-zones within 1 km of Line 1 with those beyond 1 km but within 5 km of Line 1, Figure 2.6 shows that the magnitude of anticipatory effects is somewhat smaller (4-5%) and the effects of the opening of Line 1 somewhat larger (13-25%). Figure 2.7 compares sub-zones within 1 km with all the remaining sub-zones in the city. Both, the anticipatory effects, as well as the effects of the opening of Line 1, are uniformly lower in this case, with the former in the range of 4-5% and the latter 4-16%. This suggests that other parts of the city not close to Line 1 were experiencing much higher increases in property prices, likely not due to Line 1. This also suggests that sub-zones closer to Line 1 are more appropriate controls for sub-zones within 1 km of Line 1.

Since our treatment definition is based on the proximity of the sub-zone centroids to the nearest point on Line 1 via the road network, measurement error is likely present. So sub-zones that we classify as being within 3 km may be treated in reality. Figure 2.8 compares changes in property prices between sub-zones within 3 km and sub-zones beyond 3 km but within 5 km.

In this case, we find no evidence of anticipatory effects, but increases in property prices after the opening of Line 1 ranging between 9-18%. This is strong evidence of the robustness of our empirical design and the magnitude of the estimates.

Our main estimates of the anticipatory effects and effects of the opening of Line 1 are in Table 2.3. We find that two years before the opening of Line 1, property prices in sub-zones within 1 km of Line 1 increased by 5-6% relative to sub-zones beyond 1 km but within 3 km of Line 1. Due to the opening of Line 1, property prices increased by 13- 17% depending on the land-use classification. The difference in magnitude of anticipatory effects and the effects of the opening of Line 1 is especially interesting given the delays in the completion and opening of infrastructure projects in India. It could be due to limited forward-looking behavior on the demand side, insufficient flexibility in the credit market, or a general mistrust of government announcements.

We show the robustness of these capitalization effects by estimating equation 2.1 for various control group definitions. Comparing sub-zones within 1 km with those beyond 1 km but within 5 km of Line 1, we find anticipatory effects ranging from 3.7-4.7% and the effects of the opening of Line 1 ranging from 16.5 to 19.9%, depending on the property type (Table 2.4). Comparing sub-zones within 1 km with the rest of the city, we find robust anticipatory effects but significant increases in prices due to the opening of Line 1 only for residential, commercial office, and industrial land-use properties. The average effect size is 3-4% for anticipatory effects, similar to other specifications. However, the effects of the opening of Line 1 are much smaller, lying in the range of 8.6-11.6% (Table 2.5).

It is possible that the effects of Line 1 extended beyond 1 km, so we also estimate equation 2.1 considering sub-zones within 3 km of Line 1 as treated and those beyond 3 km but within

5 km as control. We note the lack of significant differences in price changes in the pre-Line 1 period between these groups of sub-zones in Figure 2.8. Table 2.6 shows weak but negative anticipatory effects for industrial land (about 2%) and no significant anticipatory effects for other property types. However, the magnitude of price appreciation due to the opening of Line 1 is still in the range of 12.9-16.5%, which is similar to that noted for our main sample (1 km vs 3 km). In this case, the maximum response is noted for commercial office and commercial shop floorspace types instead of residential properties. This is likely a reflection of transit-oriented development, which would in principle be less constrained due to distance from Line 1.

2.3.3 Comparison with the Monorail

To understand the relative importance of Line 1 in Mumbai, we also estimate the effects of the monorail, the first phase of which became operational in Feb 2014, four months before the opening of Line 1. Figures 2.9 and 2.10 show trends in the average price difference in sub-zones within 1 km of Monorail and those beyond 1 km for but within 3 km and 5 km of Monorail, respectively. The corresponding average estimated effects are in Tables 2.7 and 2.8. We find significant negative anticipatory effects and no significant positive capitalization benefits after the opening of Monorail. The negative anticipatory effects are consistent across different property types.

A potential reason for the absence of property price appreciation due to Monorail is that its disamenity value outweighs any benefits to commuters, households, or firms in its vicinity. In the context of a commuter train connecting Montreal, Canada with its southern periphery, Dubé et al. (2013) finds an appreciation in property values for houses with increased connectivity to transit and no change in property values for houses that were along the transit route but did not experience an improved connection. We also find no significant effects when comparing subzones within 3 km and those beyond 3 km but within 5 km of Monorail (Table 2.9).

2.3.4 Discussion

The magnitude of effects we find for the impact of Line 1 is slightly higher than in the existing literature. In the context of Singapore, Diao et al. (2017) estimates the effects of a mass rapid transit line that opened in phases in 2009-11 on non-landed housing values in the vicinity of the transit line to be 8.6%. Gupta et al. (2020) estimates a 10% increase in property values in the vicinity of the Second Avenue Subway line that opened in New York City in 2017. Billings (2011) estimates a 4-11% increase in residential property values in response to a light transit line in Charlotte, North Carolina within 1 mile of the line, and no effects on commercial property values. Zhou et al. (2019) find a price appreciation close to 4% in response to the Line 6 of the Chinese metro.

To the best of our knowledge, there isn't currently economics literature studying the reasons behind heterogeneity in the magnitudes of land value capitalization effects across contexts. In theory, differences in tax structure, the spatial structure of the city, zoning policies, nature of labor, housing, and land markets are potential factors underlying these differences. We plan to systematically explore this in future work.

2.4 Sources of Benefits

One source of benefits delivered by transport projects is the travel time savings that they produce. Transport projects such as Metro rail provide benefits to commuters by reducing commute times from home to work. Chapter 1 measures the value of reductions in commute time associated with Line 1, holding residence and workplace fixed, by estimating a traditional commute mode choice model. The author also estimates a joint model of residential and commute mode choice, holding workplace location fixed. The annual time-saving benefits of Line 1 when households can adjust their residential location are approximately $1 billion PPP dollars. About 20% of these benefits accrue to households living in our treatment area.

Transport projects also improve the access of households to jobs within a city. This is one of the channels through which transport projects improve welfare and raise output in general equilibrium models such as Ahlfeldt et al. (2015) and Tsivanidis (2019). Empirically, the access of a worker in location i to jobs can be measured by an employment accessibility index. For each residential location (e.g. sub-zone) in the city, the index measures employment opportunities in every employment location in the city, weighted by the travel cost to get to that location. In this section, we parameterize an employment accessibility index for each of the 726 sub-zones in Mumbai for 2004 and also for 2019. We measure the change in employment accessibility for each sub-zone over this period and use it to study heterogeneity in the impact of Line 1 on property values. Specifically, we measure how the treatment effects of Line 1 vary depending on how the employment accessibility of a sub-zone changed between 2004 and 2019.

Before using our employment accessibility measure to study heterogeneity in the impacts of Line 1 on property values, we present descriptive evidence that employment accessibility has an impact on residential property values. We estimate hedonic regressions to explain variation in residential property prices across the Greater Mumbai Region. Both, our employment accessibility index and distance to the nearest rail station have significant effects on residential property prices.

2.4.1 Measuring Employment Accessibility

Our employment accessibility index is a commuting-cost-weighted average of effective wages obtainable across the city that are accessible from a given residential location. Effective wages reflect the attractiveness of locations as employment locations after accounting for commuting time and individual preferences for commuting. Let j index possible work locations in the city. The employment accessibility index for residential location i is

$$\text{EA}_i = \sum_j (\frac{W_j}{d_{ij}}) \tag{2.3}$$

W_j is the wage obtainable at location j. $d_{ij} = exp(\kappa * t_{ij})$ is the iceberg commuting cost from house i to location j. t_{ij} is the travel time between i and j. κ is a decay parameter specifying the semi-elasticity of commuting costs d_{ij} to commuting times t_{ij}.

We use the methodology in Kreindler and Miyauchi (2021) to obtain a proxy for W_j and estimate κ. In general equilibrium models of urban location, the utility that a worker living at location i receives from working at employment location j is given by

$$U_{ij}(\omega) = \frac{W_j * \epsilon_{ij}(\omega)}{d_{ij}} \tag{2.4}$$

W_j is the effective wage obtainable at j and each worker gets the same wage. $d_{ij} = exp(\kappa * t_{ij})$ is the iceberg commuting cost between i and j represented by an exponential function of commuting time t_{ij} times the semi-elasticity of commuting costs to time κ. $\epsilon_{ij}(\omega)$ is an an idiosyncratic utility shock assumed to follow an i.i.d. Fréchet distribution with shape parameter θ and scale parameter normalized to one. Equation 2.4 implies that the probability of a worker working in j conditional on living in i is given by

$$\pi_{ij|i} = \frac{(W_j/d_{ij})^{\theta}}{\sum_j (W_j/d_{ij})^{\theta}} \tag{2.5}$$

Equation 2.5 implies the following gravity equation of commute flow shares.

$$\log\pi_{ij|i} = -\kappa * \theta * t_{ij} + \theta * \log W_j - \log\big(\sum_j (W_j/exp(\kappa * t_{ij}))^{\theta}\big) \tag{2.6}$$

We estimate the following reduced-form gravity equation of commuter flows using a Poisson pseudo-maximum likelihood estimator.

$$N_{ij} = -\beta * t_{ij} + \psi_j + \gamma_i + \nu_{ij} \tag{2.7}$$

N_{ij} represent aggregate commute flows between i and j.[14] β captures the sensitivity of commuting decisions to commuting time. γ_i and ψ_j are origin and destination fixed effects that reflect residence and workplace amenities, respectively. Workplace amenities are termed as 'effective wages' in our analysis. ν_{ij} is the random error. We can also use this approach to obtain effective wages for workers based on college education by separately estimating the gravity equation using

[14] We use aggregate commute flows instead of shares as the outcome variable because it provides a better model fit without changing the results.

commute flows for each group.

2.4.2 Data and Methods

To estimate equation 2.7, we use two representative surveys of travel demand conducted by the World Bank in 2004 and 2019 administered to 6000 and 3000 households in the city, respectively. These data have information on geocoded household locations and pincode of the workers' work location, which describes a worker's usual commuting trip. As in Chapter 1, we use a randomly chosen post office in each pincode to proxy work locations. Using road and rail networks from OpenStreetMap in a network program, we compute travel time t_{ij} along the shortest duration travel path between household location i and the relevant post office proxies for work locations j.

We estimate equation 2.7 using data on commute flows from the household surveys, considering flows between residence and work location pincodes. There are 85 unique residential pincodes and 88 unique work location pincodes in the data, implying a possible 7480 unique flows. Travel time is the pincode-pair-level mean of the minimum travel time via road or transit between each household in the survey and their work location. We estimate this equation using a Poisson pseudo-maximum likelihood estimator. Gravity equation estimates using 2019 and 2004 survey data for all workers and workers by education level are in Table 2.10. Estimates of work location fixed effects, $\hat{\psi}_j$ are assumed to proxy the effective wage obtainable W_j. The correlation between $\hat{\psi}_j$ and average income from the 2019 survey data at the level of work location pincode is 0.24.

The sensitivity of commuting decisions to commute time, β is composed of two components:

the semi-elasticity of commuting shares to commute costs (θ) and the semi-elasticity of commuting costs to commuting time (κ). Following Kreindler and Miyauchi (2021), we infer the estimated Fréchet shape parameter $\hat{\theta}$ through the relation between actual wages aggregated and the level of pincode j and $\hat{\psi}_j$. We obtain $\hat{\theta}$ by inverting the coefficient from an OLS regression of average incomes across work locations on $\hat{\psi}_j$. We then obtain $\kappa = \dfrac{\hat{\beta}}{\hat{\theta}}$. Intuitively, $\hat{\psi}_j$ are model predicted wages and they deviate from actual wages in proportion to the variation in idiosyncratic shocks. Tsivanidis (2019) estimates κ using a traditional commute mode choice model and by taking into account the shift in choices before and after the introduction of BRT. Ahlfeldt et al. (2015) estimates κ by using the parameter estimate from the gravity equation and calibrating the variation in wages implied by the model and actual wages. Warnes (2020) simply assumes κ to have the same value as in Ahlfeldt et al. (2015) and Tsivanidis (2019) because they are very similar (0.01) despite coming from different approaches. We find κ=0.0107 using 2019 survey data and 0.0178 using 2004 survey data. Note that $\hat{\psi}_j$ does not have a fixed scale, so we standardize EA_h to be mean 0 with variance 1 within each survey year.

We also construct employment accessibility indices at the sub-zone level to investigate changes at more disaggregated levels than pincodes.

While the employment accessibility index is a housing-specific amenity, ψ_j is a direct measure of the attractiveness of locations as employment locations. Therefore, we also conduct heterogeneity analysis using changes in standardized $\hat{\psi}_j$ between 2004 and 2019. We additionally estimate equation 2.7 using the information on commute flows for workers with and without a college education, separately to obtain estimates of effective wages for these two employment types. We then conduct heterogeneity analysis using changes in $\hat{\psi}_j$ constructed separately for college-educated and less than college-educated workers.

In addition to estimating indices of employment accessibility, we estimate hedonic regressions for residential property prices. These provide descriptive evidence of the impact of employment accessibility and access to public transit on housing prices. We estimate the hedonic regression specified below based on household-level data from the 2019 World Bank survey.

$$\text{logPrice}_i = \alpha + X_i\beta + \nu_i \tag{2.8}$$

In the World Bank survey, logPrice_i is the log of the survey-reported rental price of i in Rs. per month. X_i represents a vector of amenities including distance from house i to the nearest rail station, distance to the coast, number of crimes against women in the Ward in which house i is located, a dummy for whether i is in an area classified as slum, the employment accessibility index, and a housing amenity index. The housing amenity index is the first principal component of the following housing amenities available in the survey: floorspace, number of rooms and dummy variables for a good roof, separate kitchen, separate toilet and bathrooms inside the house, access to piped water, presence of footpath in the neighborhood, and the house being well-connected to the rest of the city.

We also estimate Equation 2.8 with outcome variable as the log of the assessed value of residential property in the sub-zone in which house i is located, in Rs. per sqm.

2.4.3 Results

Estimates of the hedonic models of the determinants of rental prices and assessed residential property values are in Table 2.11. Housing amenities are a significant determinant of rental prices in the World Bank survey, but not when the price is measured at the sub-zone level, using Ready

Reckoner data. This is because the variation in basic housing amenities across sub-zones is limited. Being farther from the coast, in a slum, or a Ward with more reports of crimes against women reduce housing prices.

Employment accessibility and being closer to a rail stop both increase housing values. 1 km reduction in proximity to the nearest rail station is associated with a 4.7% increase in the rental price and a 5.6% increase in assessed sale price. For houses within 1 km of Line 1, the mean reduction in distance to the nearest station is 1.68 km, suggesting an average increase in rental and sale prices of 7.9% and 9.4%, respectively. This magnitude is smaller than the difference-in-difference estimate we obtain. In comparison, 1 standard deviation increase in the employment accessibility index is associated with a 7% increase in the rental price and a 4% increase in sale price.

We next look at changes in the employment accessibility index across the city from 2004 to 2019. Figure 2.11 shows changes in the employment accessibility index across the city using κ obtained using old and new surveys. The gradient indicates quantiles of changes in employment accessibility. The largest improvements occurred in areas in the center of the city, around Line 1. The smallest improvements occurred in the southernmost part of the city, which is the location of the traditional business district. This is consistent with the pattern of changes in population density in the city over the period.

Spatial changes in employment accessibility at the sub-zone level are shown in Figure 2.12. Broadly, the pattern is the same as in Figure 2.11, with maximum improvement taking place in the center of the city and the least amount of change in the periphery. However, in the map based on κ from 2019 survey data (right), we note that almost the entire area around Line 1 lies in the maximum change region. This also points to the appropriateness of our treatment and control group definitions based on proximity from Line 1. Finally, note that the area around Monorail exhibits only small improvements in employment accessibility. This helps to explain the lack of increases in property values associated with Monorail observed in the previous section.

In Figure 2.13 we look at the changes in employment accessibility as a function of distance from Line 1 using a local linear regression. The negative gradient is consistent with the patterns noted in the maps. These results suggest that, despite its size, Line 1 generated significant benefits because of its strategic location. Since employment accessibility is an important determinant of housing values, we further investigate its importance in driving the capitalization benefits estimated in the previous section.

We examine changes in our difference-in-differences estimates when we restrict our samples to sub-zones above a certain threshold of employment accessibility change. In Table 2.12, we show our main difference-in-difference specification (comparing sub-zones within 1 km of Line 1 vs those beyond 1 km but within 3 km of Line 1) for sub-zones where the changes in employment accessibility were above 10th, 25th, 50th, 75th, and the 90th percentiles. Compared to our main results in Table 2.3, both the anticipatory effects and the effects of the opening of Line 1 are slightly higher in this case and robust across different subsamples and specifications. Since these treatment and control areas experienced similar improvements in employment accessibility, the similarity of these estimates with our main estimates in Table 2.12 suggests the validity of our main results.[15] It also highlights that the effects observed in the previous section are driven by these sub-zones.

To test the robustness of our conclusion and make sure that it is not driven by a limited sample size or a selected definition of treatment and control areas, we repeat this analysis for

[15] This is because some of these improvements could be due to factors unrelated to Line 1.

sub-zones within 3 km vs beyond 3 km but within 5 km and for sub-zones within 1 km vs all sub-zones in the city. These are shown in Tables 2.13 and 2.14. When comparing sub-zones within 3 km vs 3-5 km, there is no evidence of anticipatory effects as before (Table 2.6) and the effects of the opening of Line 1 are slightly higher than those estimated without the sample restriction. This is especially true for the samples of sub-zones above the 75th and 90th percentiles of employment accessibility changes. When comparing sub-zones within 1 km with the rest of the city for samples above certain thresholds of changes in employment accessibility index, the magnitudes of the anticipation effects as well as those of opening of Line 1 are similar to the previous section.

Finally, we repeat this heterogeneity analysis by estimating equation 2.1 for subsamples based on changes in employment location amenities overtime for all workers, for workers with a college education, and for those who have below college education. Table 2.15 shows estimates of equation 2.1 comparing sub-zones within 1 km of Line 1 with those beyond 1 km but within 3 km of Line 1 estimated using subsamples of sub-zones based on the extent of changes in sub-zone-level employment location amenities or effective wages.[16] The first two panels have estimates for sub-zones that experienced an increase in effective wages from 2004 to 2019 that was above the median and 75th percentile. The differential effect of being close to Line 1 on property prices is higher for sub-zones above the 75th percentile changes in effective wages as compared to our main sample (Table 2.3). This is true for residential, commercial, and industrial property types but the difference is most prominent for commercial shop use. Given that the increase in prices is at least partly driven by demand, higher effective wages are consistent with

[16]We calculate this change by estimating gravity equation (equation 2.7) separately for the 2004 and 2019 sample using pincode level flows. We construct the following measure of change using estimates of destination fixed effects that have been standardized for each year separately: $\hat{\psi}_{j\,std,2019} - \hat{\psi}_{j\,std,2004}$. We then match them to sub-zones based on the pincode of the sub-zone centroid. We do this for the entire sample, and for samples of workers who do and do not have a college education.

the higher magnitudes of the property price increase. This is also consistent with the notion of transit-oriented development wherein commercial activities are especially encouraged in areas around transit.

We further examine this for sub-zones where the change in effective wages was driven by employment patterns of college-educated workers. The second and third panels of Table 2.15 restrict the sample to sub-zones with above-median and above 75th percentile improvements in effective wages for college-educated workers. The difference-in-difference coefficient for commercial office properties is prominently higher (at 0.207) than that observed when the sample restriction was made based on improvements in average effective wages (0.164). This means that businesses that employed college-educated workers led to the increase in property prices that we observed for sub-zones that experienced an increase in effective wages as noted in the second panel of Table 2.15. The last two panels present estimates for the sample of sub-zones that experienced above the median and above 75th percentile increases in effective wages for workers without a college education.

The price increase was not only due to businesses that employ college-educated workers. However, magnitudes suggest that the increase in commercial office property prices in Table 2.3 was driven by sub-zones that experienced an increase in effective wages of workers with a college education. We interpret the results from heterogeneity analysis based on effective wages as only suggestive for two reasons. First, these effective wages were calculated at the level of 85 pincodes which are much more aggregate than sub-zones. The matching of pincode to sub-zone is based on the location of the centroid of the sub-zone, leading to measurement error, which is likely, not classical. Second, the limited variation in the distribution of effective wages at the sub-zone level and spatially restricted samples create a severe imbalance between treatment and control groups as sub-zones beyond 3 km are included in the sample. This limits our ability to conduct robustness tests using other proximity-based definitions of treatment and control groups.

2.5 Conclusion

In this paper, we examine the benefits of Mumbai's first subway line, Metro Line 1, by estimating its effects on property prices. Line 1 is 11.4 km in length and a marginal addition to Mumbai's expansive suburban railway network. It started operations in June 2014 and provided the first east-west rail link in the city. Effects on property prices surrounding Line 1 may reflect the reductions in travel times for commuters living near the line and improved access to jobs across the city, net of costs imposed by potentially increased congestion or crime. They may also reflect agglomeration economies that firms enjoy by locating near the line or other general equilibrium changes that occurred in response to Line 1.

Using a spatial difference-in-difference approach, we estimate the anticipatory effects and the effects of the opening of Line 1 on property prices for different floorspace-use categories including residential, commercial office, commercial shop, industrial, and open-use land. We compare areas within 1 km of Line 1 with those beyond 1 km but within 3 km of Line 1 and find an anticipatory increase in property prices by 5-6% and a price increase of 13-17% after the opening of Line 1. We find the maximum average increase for residential properties. Our results are robust to different treatment and control group definitions. If we assume that the entire area around Line 1 is covered by single-story residential properties, our estimate of the property price increases implies an increase in property values of $6.97 Billion PPP within 1 km of Line 1.

We corroborate our findings using evidence from two household surveys, conducted in

2004 and 2019, which contain data on housing costs, housing amenities, and worker commuting behavior. We use the information on commute flows from each survey to estimate an index of employment accessibility—an index, for each residential location, of commuting-costs-weighted employment opportunities in different employment locations in the city. We demonstrate that the benefits of Line 1 are similar or slightly higher in treatment areas that have experienced the largest increases in employment accessibility between 2004 and 2019, suggesting that improvements in employment accessibility are underlying factors behind the observed effects. We find that in particular, the improvements in employment location amenities for college-educated workers drove the property price appreciation for commercial office floorspace-use.

Since changes in location amenities may also be driven by factors unrelated to Line 1, our results underscore the importance of the strategic placement of transport investments. Our finding of no significant effects on property prices due to the monorail, which began operations four months before Line 1, suggests a poor placement choice for the monorail, also consistent with the small relative changes in employment accessibility around it.

2.6 Figures and Tables

Figure 2.1: Mumbai's Rail Network

This map shows the Suburban railway network of Mumbai in grey, along with Metro Line 1 (11.5 km) in blue, and Monorail Phase 1 (8.9 km) in green.

Figure 2.2: Population density
(People per sqkm)

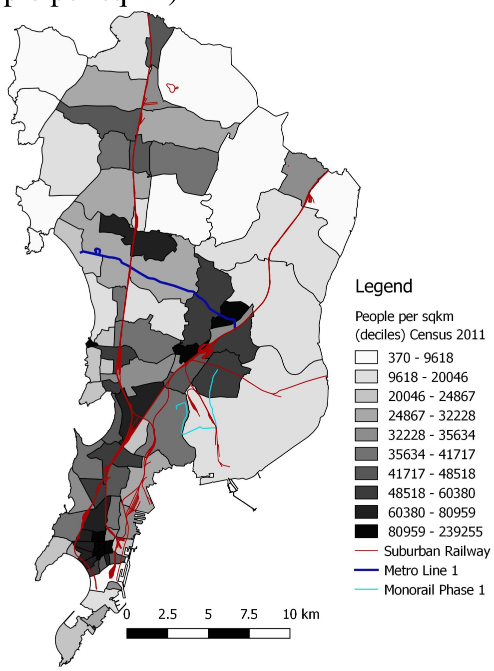
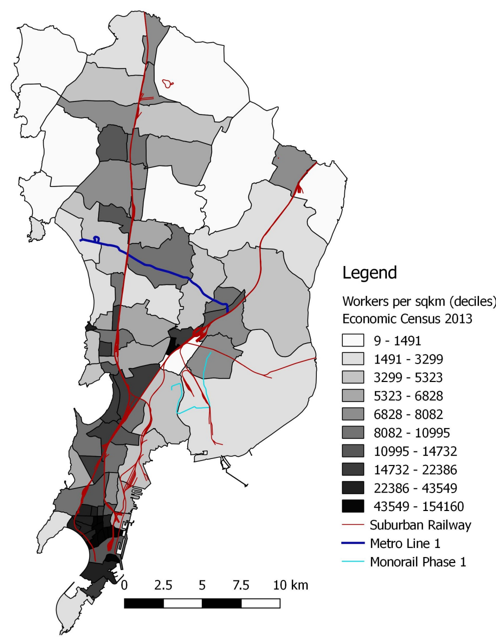
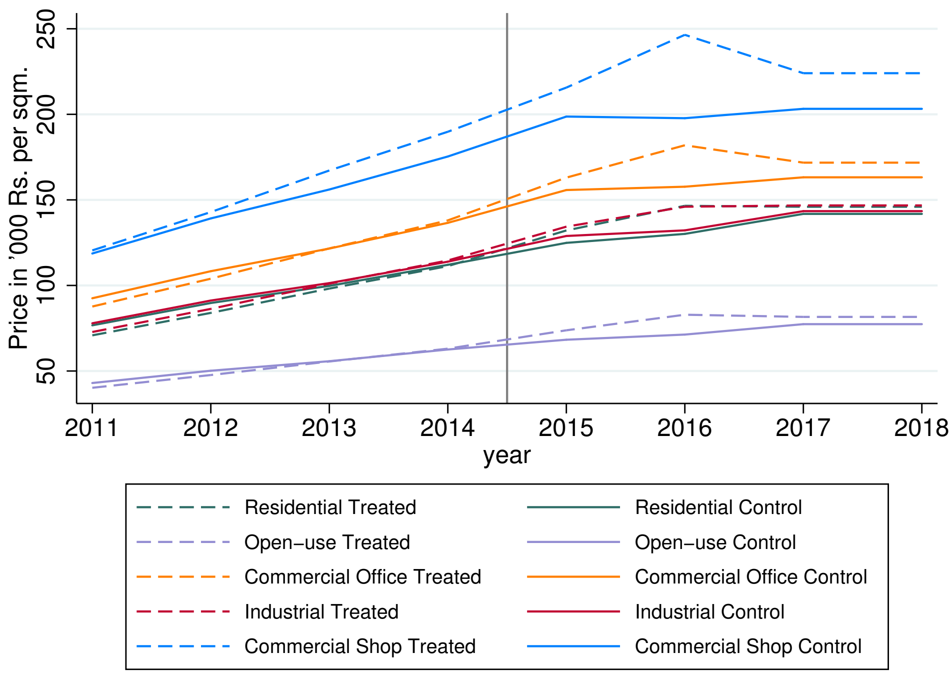


Figure 2.3: Employment density
(Workers per sqkm)



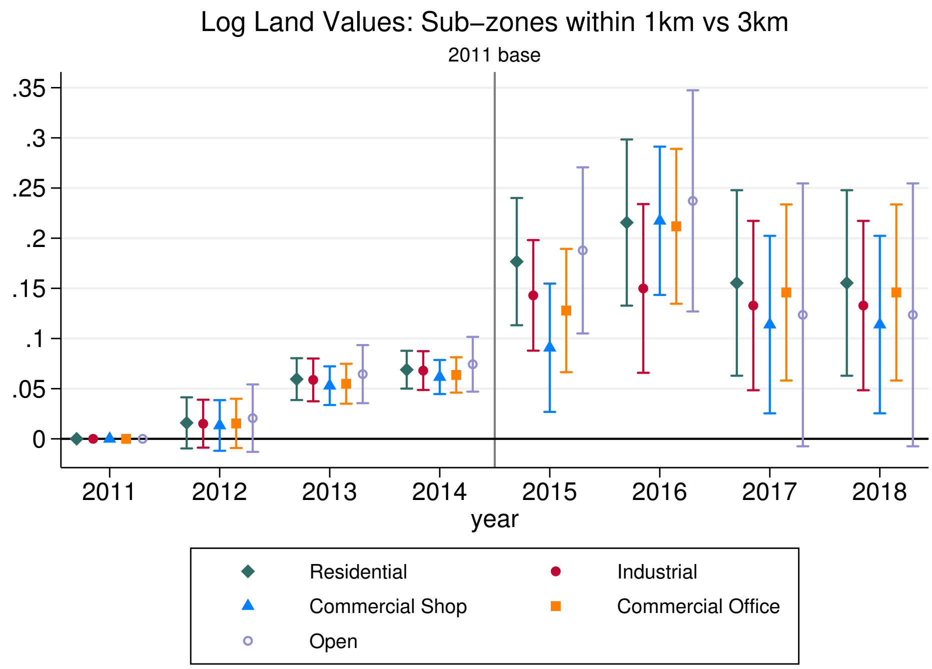
These maps show the 88 administrative sections in Mumbai shaded to reflect deciles of population per sqkm and workers per sqkm. Population density is calculated using population figures from Census 2011. Employment density is calculated using the number of workers employed in formal establishments from the Economic Census 2013. Area of each section is calculated using a digitized map of the city. The Suburban railway network of Mumbai is in red. Metro Line 1 is in royal blue. Monorail Phase 1 is in sky blue.

Figure 2.4: Trends in Property Prices in Sub-zones within 1 km and those beyond 1 km but within 3 km of Metro Line 1



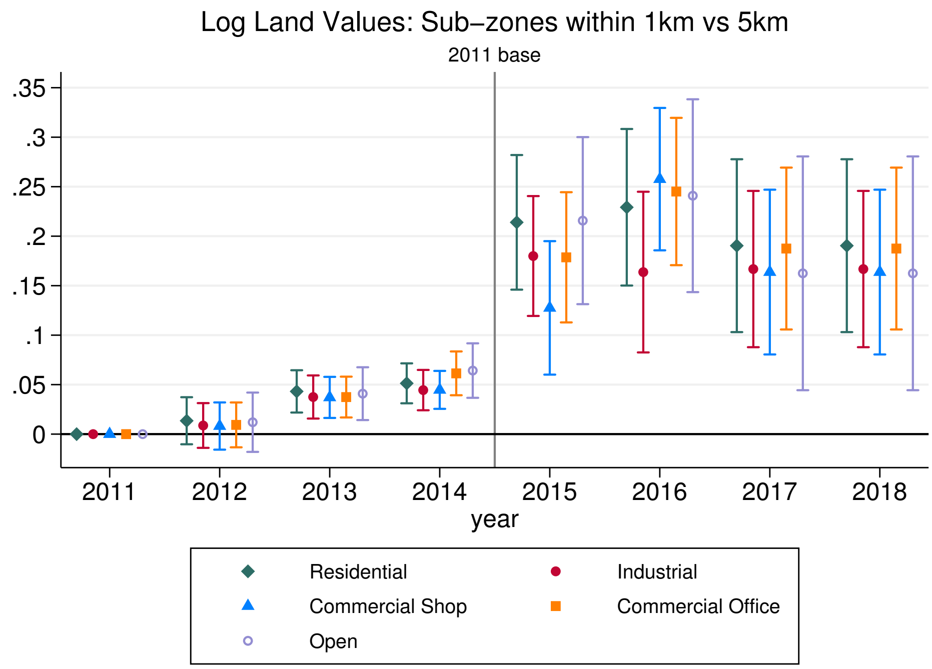
This figure shows trends in average property prices in sub-zones within 1 km of Line 1 (treated group) and those beyond 1 km but within 3 km of Line 1 (control group). Line 1 started in 2014 and the effects of its opening would show up 2015 onward, because assessed prices are based on previous year's market conditions.

Figure 2.5: Differences in Log Prices in Sub-zones within 1 km of Metro Line 1 vs Sub-zones within 1-3 km of Line 1



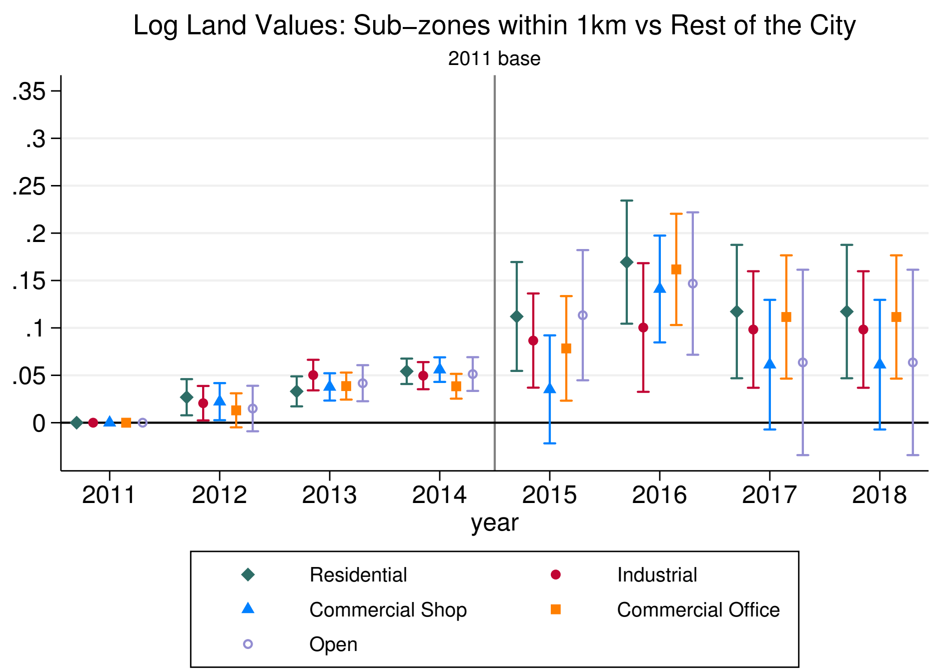
Each point on this graph shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following equation estimated using OLS. $\text{logP}_{st} = \alpha_s + \tau_t * \text{Year}_t + \beta_t * \text{Treated}_s * \text{Year}_t + \epsilon_{st}$. logP_{st} is the property price in sub-zone s and year t, α_s and τ_t represent sub-zone and year fixed effects, respectively. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). S.e. are clustered at the sub-zone and year levels.

Figure 2.6: Differences in Log Prices in Sub-zones within 1 km of Metro Line 1 vs Sub-zones within 1-5 km of Line 1



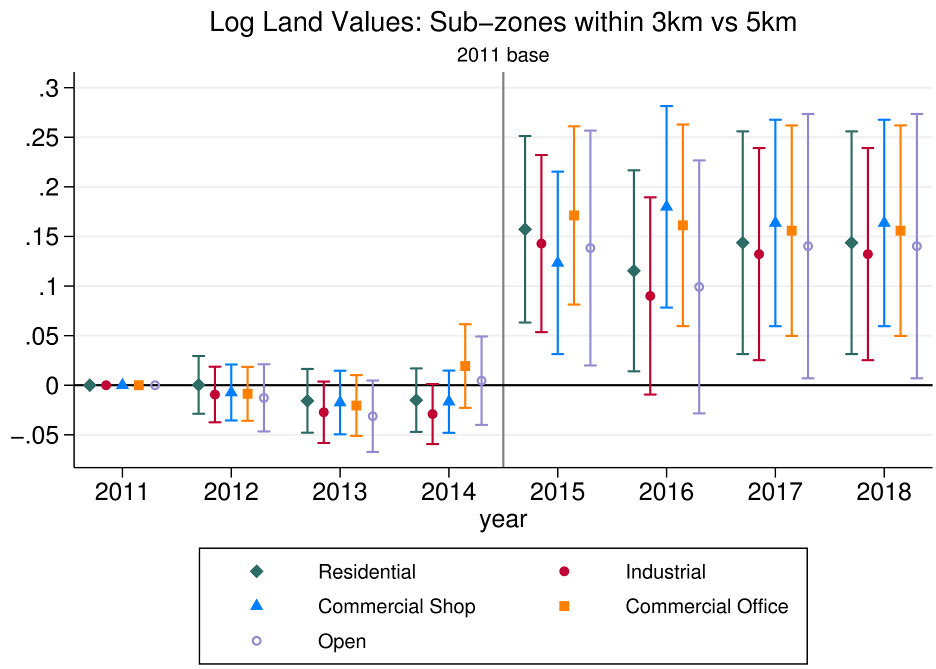
Each point on this graph shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following equation estimated using OLS. $\text{logP}_{st} = \alpha_s + \tau_t * \text{Year}_t + \beta_t * \text{Treated}_s * \text{Year}_t + \epsilon_{st}$. logP_{st} is the property price in sub-zone s and year t, α_s and τ_t represent sub-zone and year fixed effects, respectively. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). S.e. are clustered at the sub-zone and year levels.

Figure 2.7: Differences in Log Prices in Sub-zones within 1 km of Metro Line 1 vs Rest of the City



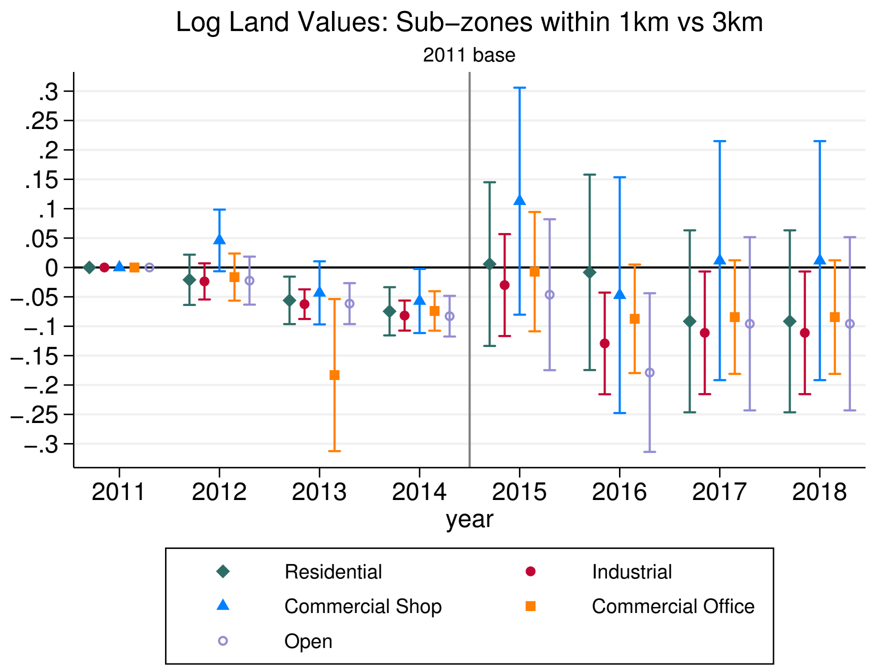
Each point on this graph shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following equation estimated using OLS. $\text{logP}_{st} = \alpha_s + \tau_t * \text{Year}_t + \beta_t * \text{Treated}_s * \text{Year}_t + \epsilon_{st}$. logP_{st} is the property price in sub-zone s and year t, α_s and τ_t represent sub-zone and year fixed effects, respectively. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). S.e. are clustered at the sub-zone and year levels.

Figure 2.8: Differences in Log Prices in Sub-zones within 3 km of Metro Line 1 vs Sub-zones within 3-5 km of Line 1



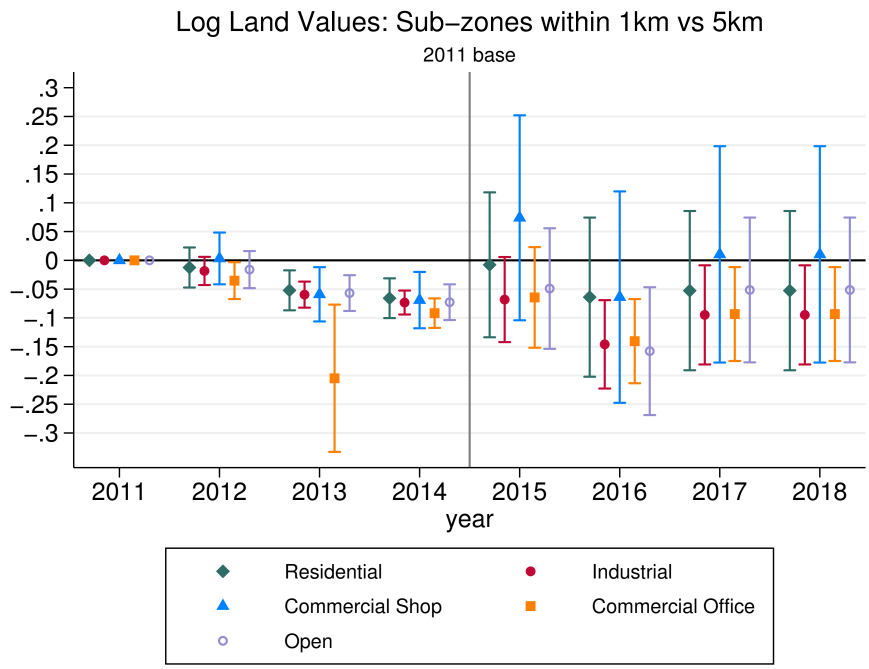
Each point on this graph shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following equation estimated using OLS. $\text{logP}_{st} = \alpha_s + \tau_t * \text{Year}_t + \beta_t * \text{Treated}_s * \text{Year}_t + \epsilon_{st}$. logP_{st} is the property price in sub-zone s and year t, α_s and τ_t represent sub-zone and year fixed effects, respectively. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). S.e. are clustered at the sub-zone and year levels.

Figure 2.9: Differences in Log Prices in Sub-zones within 1 km of the Monorail vs Sub-zones within 1-3 km of the Monorail



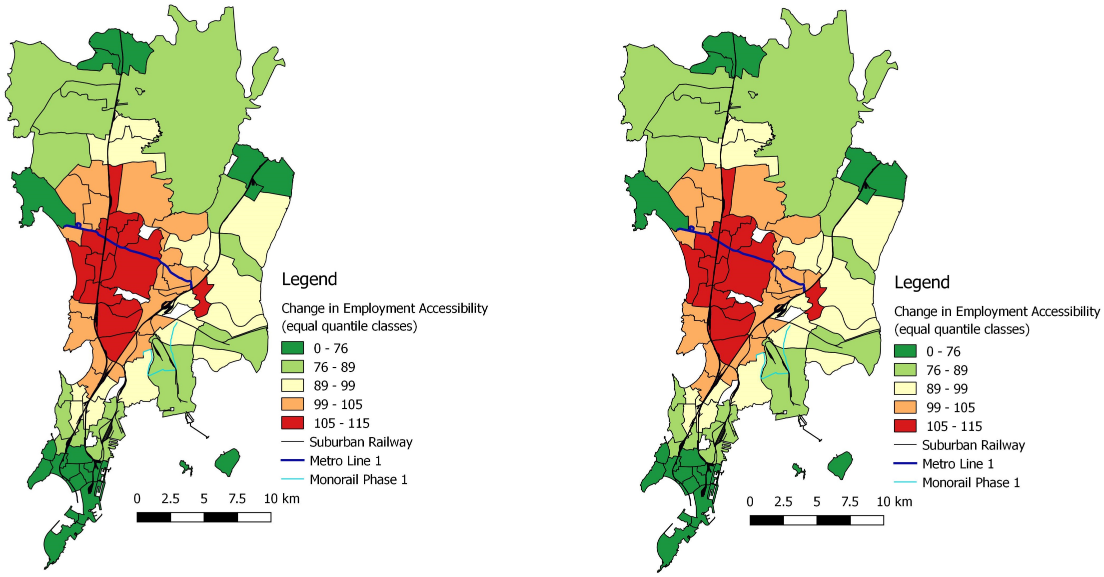
Each point on this graph shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following equation estimated using OLS. $\mathrm{logP}_{st} = \alpha_s + \tau_t * \mathrm{Year}_t + \beta_t * \mathrm{Treated}_s * \mathrm{Year}_t + \epsilon_{st}$. logP_{st} is the property price in sub-zone s and year t, α_s and τ_t represent sub-zone and year fixed effects, respectively. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). S.e. are clustered at the sub-zone and year levels.

Figure 2.10: Differences in Log Prices in Sub-zones within 1 km of the Monorail vs Sub-zones within 1-5 km of the Monorail



Each point on this graph shows the proportional change in the difference between average prices in treatment and control sub-zones relative to the difference in 2011. This is the coefficient vector β in the following equation estimated using OLS. $\mathrm{logP}_{st} = \alpha_s + \tau_t * \mathrm{Year}_t + \beta_t * \mathrm{Treated}_s * \mathrm{Year}_t + \epsilon_{st}$. logP_{st} is the property price in sub-zone s and year t, α_s and τ_t represent sub-zone and year fixed effects, respectively. Year_t is a vector of indicator variables for the years 2012-2018 (2011 is the base year). S.e. are clustered at the sub-zone and year levels.

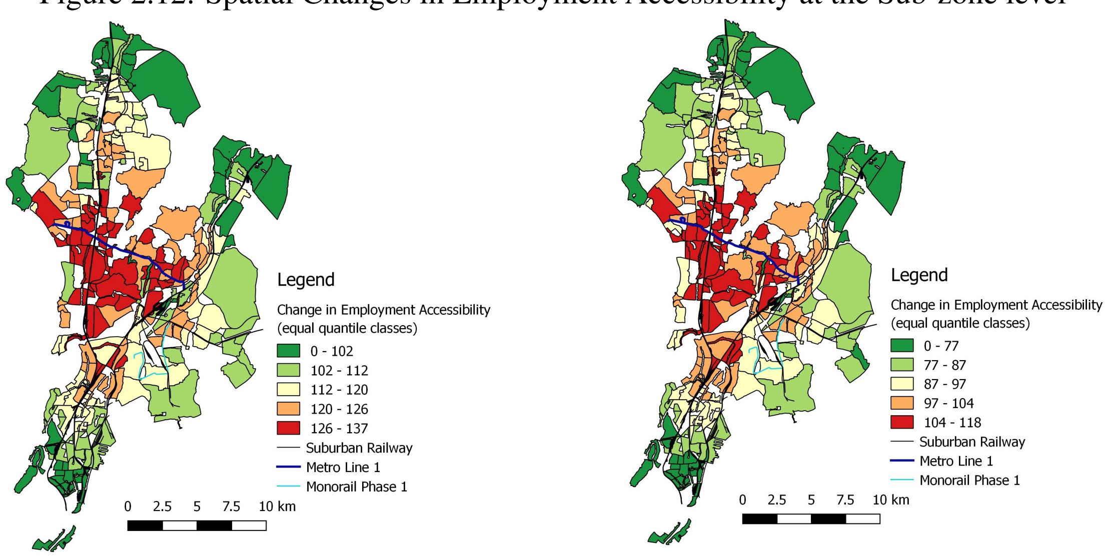
Figure 2.11: Spatial Changes in Employment Accessibility Index at the Pincode level



Main definition **Alternative definition**

These maps show changes in employment accessibility from 2004 to 2019 using standardized indices aggregated at the pincode level. Construction is discussed in Section 2.4. Main definition (Left map) uses commuting preference parameters estimated using 2019 data and alternative definition (Right map) uses parameters estimated using 2004 data.

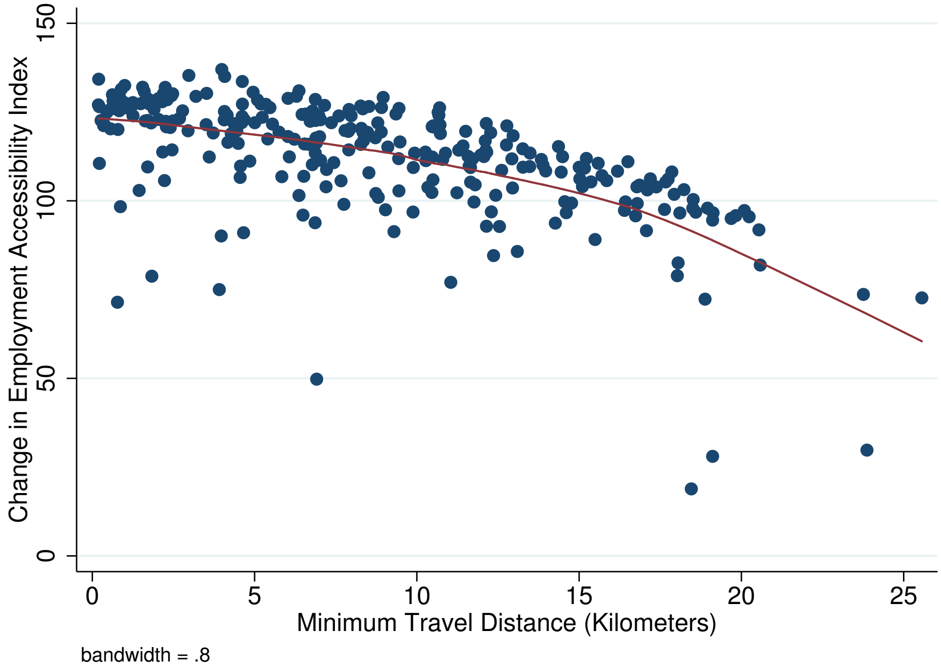
Figure 2.12: Spatial Changes in Employment Accessibility at the Sub-zone level



Main Definition **Alternative Definition**

These maps show changes in employment accessibility from 2004 to 2019 using standardized indices aggregated at the sub-zone level. Construction is discussed in Section 2.4. Main definition (Left map) uses commuting preference parameters estimated using 2019 data and alternative definition (Right map) uses parameters estimated using 2004 data.

Figure 2.13: Smooth Fitted Values from a Local Linear Regression of Change in Employment Accessibility on Distance from Metro Line 1



This figure shows the gradient of changes in employment accessibility with respect to distance from Line 1. The plot shows fitted values from a local linear regression of changes in employment accessibility index on distance from Line 1. Employment accessibility indices are calculated using 2019 preference parameters and changes are aggregated at the sub-zone level. Construction is discussed in Section 2.4.

Table 2.1: Commute Mode Shares for Workers who Commute to Work (Census 2011)

Mode	Share (in %)
On foot	31.07
Bicycle	1.50
Moped/Scooter/Motor Cycle	5.55
Car/Jeep/Van	5.96
Tempo/Autorickshaw/Taxi	3.91
Bus	20.41
Train	30.79
Water transport	0.21
Any other	0.61
Total	**100.00**
Workers who don't commute	19.62

This Table presents commute mode shares from Census 2011 for workers who are not in the agriculture and allied sectors.

Table 2.2: Property Prices Before and After Line 1 (in Rs. per sqm)

	Pre-Line 1						Post-Line 1		
	Base 2011-2012			Anticipatory 2013-2014			Treatment 2015-2018		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Within 1 km of Line 1									
Residential	104	77,406	26,429	104	104,742	33,766	208	142,658	41,301
Commercial Office	104	95,815	29,985	104	129,839	38,645	208	172,110	51,200
Commercial Shop	104	131,694	39,949	104	178,505	52,524	208	227,540	71,593
Industrial	104	79,554	25,552	104	107,805	32,812	208	143,473	41,757
Open-use	104	43,923	18,894	104	59,307	24,473	208	79,976	29,955
Beyond 1 km but within 3 km of Line 1									
Residential	168	83,223	36,371	170	105,912	48,132	340	134,662	55,003
Commercial Office	168	100,430	42,238	170	129,077	53,357	340	159,961	63,643
Commercial Shop	168	128,917	52,925	170	165,713	66,143	340	200,708	78,109
Industrial	168	84,528	35,759	170	107,638	47,478	340	136,991	54,068
Open-use	168	46,579	24,490	170	59,149	31,977	340	73,584	38,370
Beyond 3 km but within 5 km of Line 1									
Residential	120	80,011	33,969	122	107,819	45,664	244	124,553	61,432
Commercial Office	120	100,025	43,809	122	133,087	57,935	244	146,788	72,331
Commercial Shop	120	128,661	55,255	122	173,019	73,109	244	183,890	89,629
Industrial	120	82,880	32,176	122	112,326	43,125	244	126,354	61,634
Open-use	120	42,290	22,967	122	55,749	28,225	244	64,588	40,300
Entire City									
Residential	1,434	102,291	85,999	1,452	1,610,652	56,400,000	2,903	165,393	107,045
Commercial Office	1,434	128,231	96,506	1,452	171,387	171,571	2,903	203,721	128,889
Commercial Shop	1,434	163,463	118,876	1,452	214,671	156,733	2,903	257,905	158,583
Industrial	1,434	103,304	78,297	1,452	135,667	108,864	2,903	168,035	107,211
Open-use	1,434	53,027	42,319	1,452	68,753	52,768	2,903	86,462	59,059

This Table presents summary statistics of prices in Rs. per sqm. for different property types. Categorization of years is consistent with the definitions of base period, anticipatory effects period, and post period in our analysis.

Table 2.3: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 1 km vs 1-3 km

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 1 km	0.056***	0.052***	0.051***	0.056***	0.059***
	(0.005)	(0.004)	(0.004)	(0.006)	(0.008)
Post-2014*Within 1 km	0.168**	0.150**	0.127**	0.132**	0.158*
	(0.048)	(0.047)	(0.052)	(0.042)	(0.067)
Observations	1,086	1,088	1,088	1,086	1,092
R^2	0.82	0.81	0.81	0.80	0.77

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 3 km of Metro Line 1. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 1 km vs 1-5 km

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 1 km	0.040***	0.045***	0.037***	0.037***	0.047***
	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)
Post-2014*Within 1 km	0.199***	0.195***	0.174**	0.165***	0.189**
	(0.046)	(0.045)	(0.051)	(0.041)	(0.060)
Observations	1,566	1,568	1,568	1,570	1,578
R^2	0.77	0.76	0.78	0.76	0.74

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 5 km of Metro Line 1. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 1 km vs Rest of the City

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 1 km	0.030***	0.032***	0.036***	0.040***	0.039***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Post-2014*Within 1 km	0.116**	0.109**	0.064	0.086**	0.089
	(0.040)	(0.038)	(0.042)	(0.034)	(0.050)
Observations	5,733	5,713	5,735	5,719	5,787
R^2	0.85	0.89	0.88	0.87	0.87

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample includes all sub-zones in the city. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 3 km vs 3-5 km

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 3 km	-0.016	0.004	-0.013	-0.023*	-0.007
	(0.013)	(0.016)	(0.013)	(0.012)	(0.017)
Post-2014*Within 3 km	0.140**	0.165**	0.161**	0.129*	0.136*
	(0.059)	(0.056)	(0.057)	(0.057)	(0.069)
Observations	1,566	1,568	1,568	1,570	1,578
R^2	0.77	0.75	0.78	0.76	0.74

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 5 km of Metro Line 1. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: DID estimates of the Effects of Monorail on Assessed Property Prices in Sub-zones within 1 km vs 1-3 km

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 1 km	-0.055***	-0.044***	-0.023***	-0.055***	-0.054***
	(0.000)	(0.007)	(0.002)	(0.001)	(0.001)
Post-2014*Within 1 km	-0.057	-0.032	0.039	-0.042	-0.030
	(0.045)	(0.070)	(0.068)	(0.044)	(0.047)
Observations	1,086	1,088	1,088	1,086	1,092
R^2	0.81	0.80	0.81	0.80	0.76

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 3 km of the Monorail. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: DID estimates of the Effects of Monorail on Assessed Property Prices in Sub-zones within 1 km vs 1-5 km

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 1 km	-0.060***	-0.043***	-0.027***	-0.063***	-0.056***
	(0.001)	(0.008)	(0.004)	(0.001)	(0.001)
Post-2014*Within 1 km	-0.013	0.020	0.089	-0.001	0.013
	(0.046)	(0.070)	(0.067)	(0.044)	(0.045)
Observations	1,566	1,568	1,568	1,570	1,578
R^2	0.76	0.75	0.77	0.76	0.73

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 5 km of the Monorail. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: DID estimates of the Effects of Monorail on Assessed Property Prices in Sub-zones within 3 km vs 3-5 km

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
Two years pre-2014*Within 3 km	0.005	0.011	0.011	0.017	0.024
	(0.020)	(0.021)	(0.021)	(0.016)	(0.018)
Post-2014*Within 3 km	0.019	0.004	0.000	0.016	0.063
	(0.110)	(0.104)	(0.105)	(0.092)	(0.124)
Observations	1,566	1,568	1,568	1,570	1,578
R^2	0.76	0.75	0.77	0.76	0.73

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 5 km of the Monorail. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.10: Gravity Equation Estimates using 2004 and 2019 Commute Flows, by Worker Education Level

	(1)	(2)	(3)	(4)	(5)	(6)
Travel time (minutes)	-.138***	-.101***	-.168***	-.117***	-.063***	-.139***
	(.006)	(.006)	(.009)	(.007)	(.005)	(.008)
Survey Round	2019	2019	2019	2004	2004	2004
Worker type	All	$\geq$College	$<$College	All	$\geq$College	$<$College
Observations	7480	7480	7480	7744	7744	7744
Pseudo-R^2	0.56	0.45	0.58	0.50	0.38	0.53

This Table shows estimates of Poisson regression of commute flows between pincode pairs on travel time and origin and destination pincode f.e. Robust s.e. are in parentheses. Travel time is calculated between household location and a randomly chosen post-office in the pincode of the work location using a network program in Python and then averaged for each pincode pair. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.11: Hedonic Regressions, Determinants of Rental and Sale Price Gradient

	(1) log(rent)	(2) log(sale price)
Housing Amenity Index	0.219***	0.002
	(0.007)	(0.005)
Distance to coast in km.	-0.031***	-0.081***
	(0.006)	(0.004)
Slum Area	-0.096***	-0.076***
	(0.023)	(0.016)
No. of reports of Crimes Against Women	0.001	-0.001*
	(0.001)	(0.000)
Distance to Station in km.	-0.047***	-0.056***
	(0.013)	(0.009)
Standardized Employment Accessibility Index	0.070***	0.040***
	(0.017)	(0.012)
Observations	2,742	3,001
R^2	0.33	0.19

Dependent variable in Column 1 is log(rental price in Rs.) and in Column 2 is log(assessed residential sale price in Rs. per sqm).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.12: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 1 km vs 1-3 km by Changes in Employment Accessibility

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
10th percentile					
Two years pre-2014*Within 1 km	0.060***	0.055***	0.054***	0.059***	0.063***
	(0.005)	(0.005)	(0.004)	(0.006)	(0.008)
Post-2014*Within 1 km	0.178***	0.160**	0.135**	0.141**	0.173**
	(0.049)	(0.048)	(0.053)	(0.043)	(0.067)
Observations	1,070	1,072	1,072	1,070	1,076
R^2	0.82	0.81	0.81	0.80	0.77
25th percentile					
Two years pre-2014*Within 1 km	0.057***	0.053***	0.052***	0.057***	0.060***
	(0.005)	(0.005)	(0.004)	(0.006)	(0.008)
Post-2014*Within 1 km	0.171***	0.153**	0.126*	0.140**	0.165**
	(0.049)	(0.049)	(0.054)	(0.043)	(0.068)
Observations	1,054	1,056	1,056	1,054	1,060
R^2	0.82	0.81	0.81	0.80	0.77
50th percentile					
Two years pre-2014*Within 1 km	0.059***	0.054***	0.053***	0.059***	0.062***
	(0.005)	(0.005)	(0.004)	(0.006)	(0.008)
Post-2014*Within 1 km	0.188***	0.168**	0.139**	0.156***	0.184**
	(0.049)	(0.049)	(0.054)	(0.043)	(0.069)
Observations	1,014	1,016	1,016	1,014	1,020
R^2	0.82	0.81	0.81	0.80	0.77
75th percentile					
Two years pre-2014*Within 1 km	0.064***	0.058***	0.057***	0.063***	0.068***
	(0.006)	(0.005)	(0.004)	(0.006)	(0.008)
Post-2014*Within 1 km	0.171**	0.157**	0.126*	0.136**	0.144*
	(0.052)	(0.050)	(0.056)	(0.045)	(0.073)
Observations	878	880	880	878	884
R^2	0.83	0.83	0.83	0.82	0.78
90th percentile					
Two years pre-2014*Within 1 km	0.068***	0.061***	0.059***	0.068***	0.071***
	(0.006)	(0.005)	(0.003)	(0.005)	(0.010)
Post-2014*Within 1 km	0.168**	0.160**	0.125*	0.143**	0.154*
	(0.053)	(0.051)	(0.058)	(0.047)	(0.077)
Observations	720	722	722	720	724
R^2	0.84	0.84	0.83	0.86	0.78

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 3 km of Line 1 where the change in employment accessibility was above the percentile indicated in each panel. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.13: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 3 km vs 3-5 km by Changes in Employment Accessibility

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
10th percentile					
Two years pre-2014*Within 3 km	-0.011	0.009	-0.009	-0.020	-0.002
	(0.014)	(0.016)	(0.013)	(0.013)	(0.017)
Post-2014*Within 3 km	0.136*	0.162**	0.151**	0.126*	0.124
	(0.061)	(0.059)	(0.059)	(0.059)	(0.071)
Observations	1,526	1,528	1,528	1,530	1,538
R^2	0.77	0.76	0.78	0.76	0.74
25th percentile					
Two years pre-2014*Within 3 km	-0.012	0.009	-0.010	-0.020	-0.003
	(0.014)	(0.016)	(0.013)	(0.013)	(0.017)
Post-2014*Within 3 km	0.130*	0.159**	0.148**	0.122*	0.118
	(0.061)	(0.059)	(0.059)	(0.059)	(0.071)
Observations	1,510	1,512	1,512	1,514	1,522
R^2	0.77	0.75	0.78	0.76	0.74
50th percentile					
Two years pre-2014*Within 3 km	-0.017	0.004	-0.015	-0.026*	-0.008
	(0.014)	(0.017)	(0.014)	(0.013)	(0.017)
Post-2014*Within 3 km	0.136*	0.157**	0.138*	0.120*	0.123
	(0.064)	(0.062)	(0.062)	(0.063)	(0.075)
Observations	1,438	1,440	1,440	1,442	1,450
R^2	0.77	0.75	0.78	0.76	0.73
75th percentile					
Two years pre-2014*Within 3 km	-0.015	0.012	-0.013	-0.027	-0.003
	(0.017)	(0.022)	(0.016)	(0.016)	(0.021)
Post-2014*Within 3 km	0.185**	0.210**	0.209**	0.160*	0.172*
	(0.071)	(0.068)	(0.068)	(0.070)	(0.077)
Observations	1,214	1,216	1,216	1,218	1,226
R^2	0.79	0.77	0.80	0.78	0.76
90th percentile					
Two years pre-2014*Within 3 km	-0.016	0.014	-0.013	-0.028	-0.002
	(0.018)	(0.024)	(0.017)	(0.017)	(0.023)
Post-2014*Within 3 km	0.172**	0.203**	0.179**	0.149*	0.157*
	(0.070)	(0.067)	(0.068)	(0.069)	(0.076)
Observations	1,032	1,034	1,034	1,036	1,042
R^2	0.80	0.77	0.80	0.80	0.77

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 5 km of Line 1 where the change in employment accessibility was above the percentile indicated in each panel. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.14: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 1 km vs Rest of the City by Changes in Employment Accessibility

	Residential	Comm. Office	Comm. Shop	Industrial	Open Land
10th percentile					
Two years pre-2014*Within 1 km	0.034***	0.038***	0.040***	0.044***	0.043***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)
Post-2014*Within 1 km	0.122**	0.118**	0.069	0.093**	0.103*
	(0.040)	(0.039)	(0.043)	(0.034)	(0.049)
Observations	5,501	5,481	5,503	5,487	5,555
R^2	0.85	0.89	0.88	0.87	0.87
25th percentile					
Two years pre-2014*Within 1 km	0.033***	0.037***	0.037***	0.043***	0.043***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.004)
Post-2014*Within 1 km	0.114**	0.113**	0.064	0.091**	0.096*
	(0.040)	(0.039)	(0.043)	(0.035)	(0.049)
Observations	5,173	5,157	5,175	5,163	5,227
R^2	0.85	0.89	0.88	0.87	0.87
50th percentile					
Two years pre-2014*Within 1 km	0.034***	0.039***	0.039***	0.039***	0.045***
	(0.000)	(0.002)	(0.000)	(0.002)	(0.004)
Post-2014*Within 1 km	0.125**	0.124**	0.068	0.101**	0.109*
	(0.041)	(0.039)	(0.044)	(0.036)	(0.050)
Observations	4,616	4,600	4,618	4,606	4,668
R^2	0.85	0.89	0.88	0.89	0.87
75th percentile					
Two years pre-2014*Within 1 km	0.039***	0.045***	0.047***	0.045***	0.052***
	(0.000)	(0.003)	(0.000)	(0.002)	(0.004)
Post-2014*Within 1 km	0.119**	0.125**	0.072	0.094**	0.093
	(0.045)	(0.043)	(0.048)	(0.039)	(0.054)
Observations	4,050	4,034	4,052	4,040	4,102
R^2	0.85	0.89	0.88	0.89	0.88
90th percentile					
Two years pre-2014*Within 1 km	0.045***	0.047***	0.052***	0.053***	0.058***
	(0.005)	(0.002)	(0.001)	(0.001)	(0.004)
Post-2014*Within 1 km	0.115**	0.126**	0.063	0.099*	0.103
	(0.048)	(0.045)	(0.052)	(0.044)	(0.057)
Observations	3,716	3,700	3,718	3,706	3,766
R^2	0.85	0.90	0.89	0.91	0.89

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones in the entire city where the change in employment accessibility was above the percentile indicated in each panel. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.15: DID estimates of the Effects of Metro Line 1 on Assessed Property Prices in Sub-zones within 1 km vs 1-3 km by Changes in Effective Wages for All Workers, College-educated Workers, and those Without a College Education

	(1) Residential	(2) Commercial office	(3) Commercial shop	(4) Industrial	(5) Open Land
50th percentile: Effective Wages for all Workers					
Two years pre-2014*Within 1 km	0.053***	0.047***	0.045***	0.052***	0.056***
	(0.006)	(0.005)	(0.004)	(0.006)	(0.009)
Post-2014*Within 1 km	0.160**	0.138**	0.099	0.127**	0.155*
	(0.054)	(0.053)	(0.057)	(0.047)	(0.076)
Observations	888	890	890	888	892
R^2	0.81	0.82	0.82	0.80	0.77
75th percentile: Effective Wages for all Workers					
Two years pre-2014*Within 1 km	0.034***	0.037***	0.041***	0.034***	0.035*
	(0.006)	(0.006)	(0.006)	(0.005)	(0.017)
Post-2014*Within 1 km	0.198*	0.164*	0.196*	0.151*	0.175
	(0.086)	(0.081)	(0.087)	(0.077)	(0.132)
Observations	348	348	348	348	352
R^2	0.80	0.81	0.79	0.81	0.74
50th percentile: Effective Wages for College-educated Workers					
Two years pre-2014*Within 1 km	0.053***	0.046***	0.044***	0.052***	0.055***
	(0.005)	(0.004)	(0.004)	(0.006)	(0.007)
Post-2014*Within 1 km	0.158**	0.149**	0.120*	0.124**	0.158*
	(0.056)	(0.057)	(0.061)	(0.046)	(0.074)
Observations	786	788	788	786	788
R^2	0.85	0.84	0.84	0.83	0.81
75th percentile: Effective Wages for College-educated Workers					
Two years pre-2014*Within 1 km	0.054***	0.046***	0.048***	0.054***	0.051***
	(0.009)	(0.008)	(0.007)	(0.009)	(0.012)
Post-2014*Within 1 km	0.190**	0.207**	0.217**	0.147*	0.211*
	(0.074)	(0.071)	(0.077)	(0.066)	(0.101)
Observations	458	460	460	458	460
R^2	0.83	0.83	0.83	0.84	0.81
50th percentile: Effective Wages for Workers Without a College Education					
Two years pre-2014*Within 1 km	0.045***	0.040***	0.041***	0.046***	0.046***
	(0.007)	(0.005)	(0.004)	(0.008)	(0.012)
Post-2014*Within 1 km	0.165**	0.140**	0.115	0.133**	0.157
	(0.059)	(0.058)	(0.062)	(0.051)	(0.084)
Observations	800	802	802	800	804
R^2	0.82	0.82	0.82	0.80	0.77
75th percentile: Effective Wages for Workers Without a College Education					
Two years pre-2014*Within 1 km	0.029***	0.022**	0.024***	0.029***	0.029*
	(0.007)	(0.006)	(0.005)	(0.006)	(0.014)
Post-2014*Within 1 km	0.188**	0.159**	0.168*	0.150**	0.198*
	(0.070)	(0.066)	(0.075)	(0.063)	(0.102)
Observations	512	514	514	512	516
R^2	0.81	0.81	0.80	0.82	0.77

This Table presents estimates of equation 2.1. Dependent variable is log of assessed property prices in Rs. per sqm. Sample restricted to sub-zones with centroid within 3 km of Line 1 where the change in effective wages was above the 50th and 75th percentiles for all workers, workers with a college education, and those without a college education as indicated in each panel. Each specification has sub-zone and year fixed effects. Robust s.e. clustered at the sub-zone and year levels are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3: Air Quality Impacts of Metro Rail in Mumbai

3.1 Introduction

India is among the countries with the worst air quality according to practically every measure. Pandey et al. (2021) estimate about 0.98 million deaths in India to be attributable to ambient air pollution in 2019, about 11% of all deaths. Increasing awareness and adverse health impacts among the population have pushed policymakers in certain states and, especially, cities to pay more heed to environment policies. For example, the Delhi government since the early 2000s has introduced CNG buses and banned diesel public buses, implemented strict pollution regulations, and introduced metro rail. However, environmental regulations are still not a priority for most states in the country. This is due to states not internalizing the true economic costs of pollution, perhaps led by a lack of political incentive.

One policy area where environmental and economic benefits are complementary is public transit infrastructure investment, especially mass rapid transit systems. For cities and towns that have added public transit infrastructure in the past two decades, technological advances have made the decision to add cleaner forms of transit easier. After the success of the Delhi metro, policymakers in other states in India have especially favored metro rail projects, citing environmental benefits as one of the major reasons for their choice. At least fourteen other metro systems have been proposed or are being constructed across the country. Mumbai's planned metro rail network (over 300 km) is one of the most extensive among these projects.

In this paper, we study the air quality impacts of Metro Line 1, the first metro line in Mumbai, operational since 2014. Even though Metro Line 1 uses cleaner technology, an improvement in pollution does not necessarily follow, because of two opposing forces. The presence of Metro Line 1 can create a substitution away from road transport (Mohring (1972)), thereby reducing pollution. The improvement in traffic conditions can, however, lead to additional drivers on the road (Duranton and Turner (2011)), thereby increasing pollution.

We use information on daily average levels of nitrogen dioxide (NO_2), sulfur dioxide (SO_2), respirable suspended particulate matter (equivalent to PM_{10}) from ground-level monitoring stations and information on aerosol optical depth (AOD) from satellite data from Google Earth Engine in an event study framework to estimate the impact of opening of Line 1 on air pollution. Metro Line 1 opened to the public on June 8th, 2014. We estimate the magnitude of the change in these pollutants on that day and interpret this as the effect of Line 1. We find a significant reduction in the level of NO_2 and AOD and weak evidence of reductions in SO_2 and PM_{10}. Our results are robust to inclusion of controls for weather, temperature, wind, humidity, day of the week, month, as well as a polynomial trend. We also use placebo dates to confirm that the observed effects are indeed due to the opening of Line 1.

We contribute to the literature on the effects of public transit on air quality, which finds mixed results. There is evidence of reduced pollution due to the Delhi metro (Goel and Gupta (2017)), Taipei's rail transit system (Chen and Whalley (2012)), BRT in Mexico city (Bel and Holst (2018)) and subway lines in Beijing (Guo and Chen (2019), Li et al. (2019)). There is also evidence of increased pollution due to public transit strikes (Bauernschuster et al. (2017)). In contrast, Gendron-Carrier et al. (2018) using AOD to measure air pollution in a sample of

34 cities, finds on an average no effect of subway openings. The study does find evidence of a reduction in pollution in some cities where pollution levels were high at the baseline and ridership was substantial. Our analysis provides an implicit connection between satellite measurements of AOD and ground monitor measurements of three pollutants, NO_2, SO_2, and PM_{10}.

All these papers rely on an event study approach to measure the change in air quality before and after metro rail opening. Some have shown persistent reductions in pollution. Since pollution data are generally high frequency, event study estimates can be interpreted as causal. In comparison to public transit, road pricing policies have been shown to improve air pollution in Milan (Gibson and Carnovale (2015)) where road pricing policies and public transit are likely substitutes for air quality improvements. Driving restriction policies are not always effective, however as has been shown in the contexts of Mexico City (Davis (2008)) and Delhi (Chowdhury et al. (2017)).

There is ample evidence that air pollution adversely affects health (Currie and Walker (2011), Schlenker and Walker (2016), Bauernschuster et al. (2017) and Simeonova et al. (2018)). The impacts of metro rail on air pollution could, therefore, constitute an important component of the benefits of a metro rail system.

3.2 Data

During the study period, Mumbai had three manual ground monitoring stations that recorded daily averages of nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and respirable suspended particulate matter (equivalent to particulate matter 10 μm or smaller, PM_{10}). Two of these were operational during 2010-15, while the third one was operational during 2011-15. Their locations are shown in Figure 3.1. These stations are not spatially close to Metro Line 1. However, since air pollutants typically travel, these stations should be able to capture Line 1's contribution to pollution in the city. Each station has daily averages between two to eleven times a month during 2013-15 with the median being 8 readings a month.

Levels of these pollutants during 2013-15 are summarized in Table 3.1. Average levels of NO_2, SO_2, and PM_{10} in the city were 18.31 μg/m^3, 3.66 μg/m^3, and 106.38 μg/m^3. In comparison, the recommended annual average emission of PM_{10} set by the World Health Organization was 20 μg/m^3 in 2005 and was revised to 15 μg/m^3 in 2021 (IQAir (May 1, 2022)). The recommendation for NO_2 was 40 μg/m^3 in 2005 and revised to 10 in 2021 μg/m^3.

It is difficult to predict the exact composition of pollutants that would be affected by Line 1. Therefore, we also collect information on aerosol optical depth (AOD) from Google Earth Engine at the 1 km grid level.[1] AOD is a measure of aerosols present in the atmosphere and is estimated based on the extinction rate of a ray of light passing through the atmosphere. It is often used to proxy the surface level of $PM_{2.5}$. Gendron-Carrier et al. (2018) use this data at the 3 km level, which may be less precise given the periods of missing information in the presence of cloud cover, especially during monsoon season. We collect AOD levels for 2,011 equally spaced points 500 m apart for each day during 2013-15.

Since measured pollution is affected by temperature, wind, and humidity, we collect daily level ERA5 satellite data on these from Google Earth Engine. Temperature and humidity are measured at 2 m while the u-component and v-component of wind are at 10 m. The u and v components of the wind can be used to calculate wind speed and direction; we simply use them

[1]This data is MODIS Terra and Aqua combined Multi-angle Implementation of Atmospheric Correction (MAIAC) Land AOD.

as controls in our analysis. We spatially match these data to monitoring station locations and the grid points for which we collect AOD levels.

3.3 Effects of the Metro on Air Quality

Did the existing metro rail in Mumbai lead to lower pollution levels and did these effects persist? Figures 3.2, 3.3, and 3.4 show the raw levels of these pollutants over time before and after the first day of operations of Metro Line 1 (June 8th, 2014) along with a flexible polynomial trend that is allowed to have a different intercept before and after the opening of Line 1. There is a striking discontinuity in the levels of these pollutants before and after the metro opening. Visually, the drop in SO_2 is smaller than NO_2 and PM_{10}. This is probably because transportation is not a major source of SO_2 pollution in Mumbai. In 2018, transport emissions accounted for 30% of the NO_x emissions, 12% of the PM_{10} emissions but only 4% of SO_2 emissions in Mumbai (Urban Emissions, Mumbai (2018)).[2,3] These effects, however, seem short-lived as these levels begin rising from 2015 onward.

June marks the start of the monsoon season in Mumbai. To show that the discontinuity in the levels of pollutants is specific to 2014, we plot the levels of these pollutants before and after June 8th, 2013, one year before the opening of Metro Line 1, in Figures 3.8, 3.9, and 3.10. There is no substantial difference in these pollutants before and after this placebo date, implying that the reductions after June 8th, 2014 are not due solely to the monsoon.

To quantify the magnitude of reductions in NO_2, SO_2 and PM_{10}, we use an event study specification, estimated using OLS. At the level of ground monitor or grid points, the estimation

[2] NO_x includes NO_2 (nitrogen dioxide) as well as NO (nitrogen monoxide).

[3] Estimates of the contribution of transport sector to NO_x emissions vary a lot by the location and time period of studies. But the general consensus is that transport sector is a significant contributor (Goel and Gupta (2017)).

equation is:

$$\log Y_t = \alpha_0 + \alpha_1 * \text{Post-metro}_t + \beta_x * X_t + f(t) + \epsilon_t \tag{3.1}$$

$\log Y_t$ represent the daily average of pollution readings in log form. Post-metro$_t$ is a dummy variable indicating the days after the opening of metro. X_t is a vector of controls including day of the week fixed effects and measurements of temperature, humidity, and u and v wind components at the same frequency as the outcome variable. $f(t)$ represents a polynomial time trend. Based on graphical evidence, we assume a cubic trend in our baseline specification.[4] Day of the week fixed effects are included to account for ridership differences across days. α_1 here represents the extent of the discontinuous jump observed in Figures 3.2, 3.3, and 3.4 and can, therefore, be interpreted as the causal effect of metro opening. To account for serial correlation in errors, we cluster standard errors at the monthly level for the basic specification because of sparse data. In specifications with additional time periods, we test for robustness by clustering at different levels. This empirical strategy is similar to a regression discontinuity design with time as the running variable.

There was a substantial decline in NO_2 levels due to the opening of Line 1 (Tables 3.2 and 3.3). The decline was about 60% (Column 2 of Table 3.3) or 8.9μg/m^3. This decline is even bigger when using data only for 2014, but 2013-15 is our preferred sample period because it allows us to control for seasonal month-specific effects, while allowing only limited exposure to other confounding events. Since the effects on pollution are identified based on temporal variation only, the identification argument becomes weaker as the sample is expanded to include

[4]Our results are robust to linear, quadratic, and fourth order polynomial trends. Our conclusions are also robust to a non-parametric specification.

measurements further away from the opening of Metro Line 1. The immediate declines in SO_2 and PM_{10} after the opening of Line 1 are not statistically significant (Columns 1 and 3 of Tables 3.2 and 3.3).

The extent of decline in NO_2 is sensitive to the presence of a cubic time trend. The decline in NO_2 after Line 1 is more likely to be attributable to the opening of Line 1 when time trends are included. This claim is supported by the various robustness checks we discuss below. The importance of accounting for secular trends in event study analyses such as this one has been emphasized in Davis (2008) and Chen and Whalley (2012).

Based on the 2014 sample, declines in both, NO_2 and SO_2 after Line 1 were significant. For SO_2, the effects are difficult to attribute to Line 1 since the initial drop in its level was not significant and because transport is not a significant source of SO_2 emissions in Mumbai. Results based on the 2013-15 sample show that relative to before Line 1, the magnitude of the decline in NO_2 due to the opening of Line 1, did not persist through the end of 2015. This is also seen in Figure 3.2. This is likely due to industrial activity or the lack of persistence in substitution away from private vehicles. It could also be that part of the initial decline was driven by the novelty of a new transit line.

We also estimate the effects of opening of Line 1 on levels of AOD, which is informative about particulate matter (Dey et al. (2012)). The relation between AOD and particulate matter, however, varies by season since cloud cover adversely affects AOD measurements. These models are in Table 3.4. The magnitudes of changes in AOD are sensitive to the exact specification used. Without a cubic trend, changes in AOD are positive (Columns 1, 2, 4), similar to the changes in NO_2 noted in Column 5 of Table 3.3. This is consistent with the presence of significant correlation between AOD and NO_2.

Our preferred specification includes a cubic time trend, day of the week fixed effect, year fixed effects, and weather controls. Given the sparse AOD measurements during monsoon (June-September), we do not include month fixed effects in these regressions. We test the sensitivity of our estimates to including quarter fixed effects instead. In Columns 6 and 7 of Table 3.4, we note a decline in AOD levels by 23-37%. Gendron-Carrier et al. (2018) finds no effect of the average subway opening on AOD. While the authors find impacts for some cities at the top of ridership and baseline pollution level distribution, it is unclear where Mumbai lies in this distribution.

In our context, it is difficult to determine whether these effect sizes are reasonable. For an 11.4 km long metro line with a daily ridership of 260,000 passengers, the decline may be too big to be attributable to Line 1. The effects could be explained by a major substitution from road transportation towards the metro, but there isn't enough information at present to test this channel. In future work, we aim to ground truth these estimates using information from emissions inventory in Mumbai and vehicular pollution regulations in the city during the study period, which is not currently publicly available.

In the context of Delhi Metro, Goel and Gupta (2017) estimates a 31% reduction in NO_2 due to the first Blue Line extension but not for the other extensions studied. In the context of Beijing Metro, Guo and Chen (2019) finds a reduction in NO_2 by 37% in Tianjin, 60-104% in Shijiazhuang, and 66-74% in Shanghai, depending on the length of the sample window around the opening. In the context of BRT in Mexico City, Bel and Holst (2018) finds a 4.7-6.5% reduction in NO_x in the short run, depending on the city area. In the context of Taipei metro, Chen and Whalley (2012) finds a 14% reduction in NO_x but these effects are not robust across the different model specifications.

While it is difficult to confirm the appropriate magnitude of decline in NO_2 due to Line 1, we perform various robustness checks, which suggest that the declines were indeed due to Line 1. Our results could only be confounded by an event that occurred exactly on June 8th, 2014, the likelihood of which is small.

3.4 Robustness Checks

The main threat to identification in our analysis is the presence of confounding events that may have simultaneously affected pollution levels in the city alongside metro opening. We use various falsification tests to check for this possibility.

First, it is possible that the estimated decline in NO_2 is driven by the weather in the month of June, which is roughly the onset of the monsoon in Mumbai. Figures 3.8, 3.9, and 3.10 show that there was no discontinuous jump in the levels of any of the pollutants on June 8th, 2013. We estimate equation 3.1 with a cubic time trend using data for the year 2013. Results are in Table 3.5. There is no significant difference in the level of pollutants before and after June 8th, 2013. Results are not shown here, but without the cubic time trend, the coefficients for NO_2 and SO_2 are positive and significant. If the monsoon were the driving factor behind our results, we would have seen a negative effect, suggesting that our results are not driven by the monsoon.

Second, we test for a discontinuous jump 10 and 17 days prior to the opening of Line 1 to further investigate the possibility of confounding events. Results are in Table 3.6. The coefficient on change in NO_2 is negative but not significant. This suggests that the significant decline in NO_2 after the opening of Line 1 observed in Table 3.3 is indeed driven by the opening of Line 1.

Third, it is possible that the results are driven by some event that occurred when Metro Line 1 opened for the public. To test for this we estimate equation 3.1 using 2013-15 sample but dropping observations for 7, 15, and 30 days before and after the day Metro Line 1 opened to the public. This is akin to the donut hole approach in a regression discontinuity design to assess the influence of observations around the cutoff. Results are in Table 3.7. The magnitude of the decline in NO_2 in these samples is about 45-49%, only slightly smaller than the decline of 60% observed in Column 2 in Table 3.3. This suggests that the effects observed earlier are indeed attributable to the opening of Line 1. They become somewhat weaker as we move away from that time period, but the they are not entirely driven by some temporary event. These results also suggest that the effects are not entirely driven by the novelty factor of Line 1.

The Eastern Freeway opened in Mumbai in three phases during our sample period. The first two phases opened on June 14, 2013 and April 12, 2014. Treating these dates as placebo events, we find no effects on pollution. The third phase opened on June 16, 2014, 8 days after the opening of Line 1. When we drop 7 days before and after the opening of Line 1, the magnitude of the effects on NO_2 are slightly lower than our main estimates, which means that these effects were not driven by the opening of the third phase of the freeway. We cannot rule out that there was a compounding effect but, it seems unlikely that the effects are due only to the freeway.

Lastly, it is possible that the effects are driven by the opening of the monorail, a light rail system that opened to the public on February 2nd, 2014. Figures 3.5, 3.6, and 3.7 plot the raw levels of NO_2, SO_2 and PM_{10} with a polynomial trend before and after the monorail opening. These figures show a small increase in the levels of NO_2 and SO_2, and a decline in the level of PM_{10} after the monorail. Estimates of equation 3.1 using February 2nd, 2014, the day monorail became operational as the treatment date are in Table 3.8. With controls for weather and secular trends, we obtain a significant increase in SO_2 and NO_2 and no significant change in PM_{10}. It is not surprising that we observe a decline in raw levels in Figure 3.7 but not in Table 3.8 since about 50% of PM_{10} emissions in Mumbai are due to dust, and their measurements are heavily affected by weather conditions.

The lack of effects of the monorail on air pollution relative to Line 1 is likely a result of difference in use. The daily ridership of the monorail in 2014 was about 20,000 people, while ridership of Line 1 was about 260,000 people per day. It also then suggests that the effects observed for Line 1 are likely due to individuals substituting away from private vehicles to public transit.

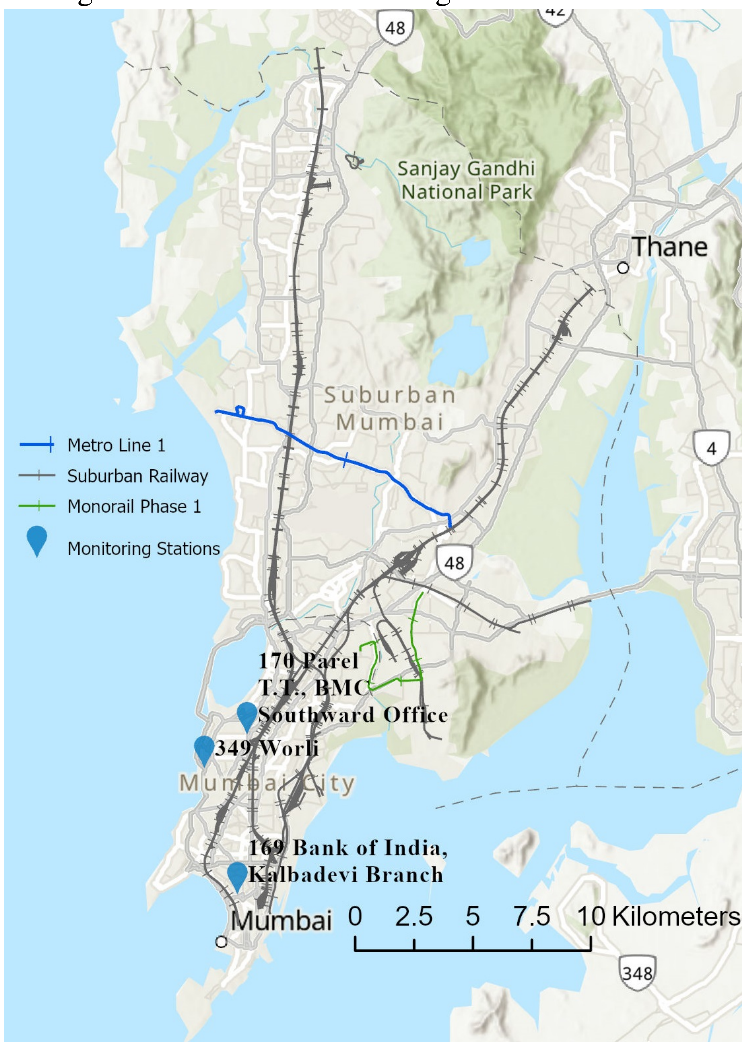
3.5 Conclusion

In this paper, we study the effect of opening of Line 1 on air pollution using an event study approach. We use information on ambient nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and respirable suspended particulate matter (equivalent to PM_{10}) from ground monitoring stations along with information on aerosol optical depth (AOD), wind, temperature, and humidity from satellite data collected using Google Earth Engine. We find that the opening of Line 1 led to a reduction of about 9 $\mu g/m^3$ or 60% in levels of NO_2 and of 24-38% in AOD levels. We find no evidence of changes in SO_2 or PM_{10}.

Transport emissions account for 30% of NO_x emissions in Mumbai. While other studies have found similar magnitudes of effects of subways on NO_2, it is possible that the magnitude of the decline we find in Mumbai is too big to have been caused only by Line 1. We test for some possible confounding events; however, these cannot explain the effects we attribute to Line 1. Additional information about pollution regulations in Mumbai and pollution inventories is needed to further ground truth these results and inform this literature. We leave this for future work.

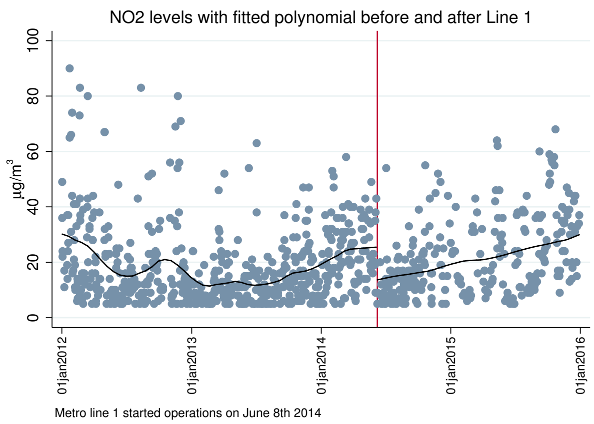
3.6 Figures and Tables

Figure 3.1: Ground Monitoring Stations Locations



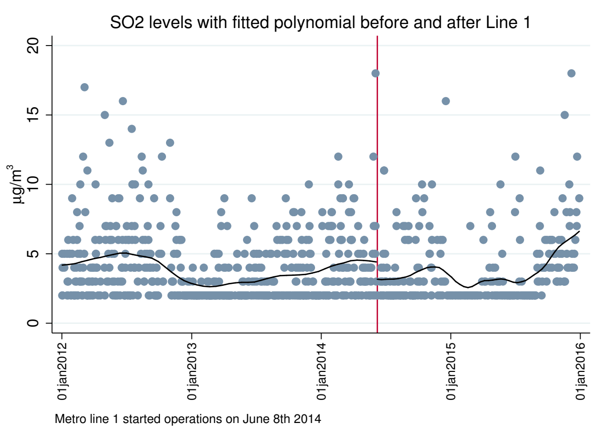
This map shows the three manual ground monitoring stations that were operational in 2014. Suburban railway network of Mumbai is in grey, Metro Line 1 in blue, and Monorail Phase 1 in green.

Figure 3.2: NO_2 Levels with Fitted Polynomial before and after Metro Line 1



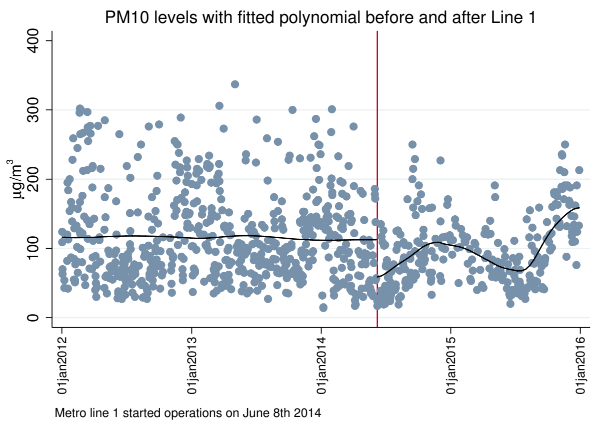
This map shows levels of NO_2 during 2012-16 with fitted polynomial trends before and after the opening of Line 1. Red Line indicates June 8th, 2014, the day Line 1 became operational.

Figure 3.3: SO_2 Levels with Fitted Polynomial before and after Metro Line 1



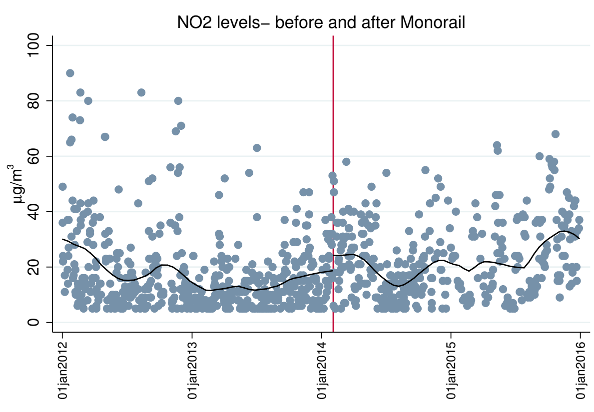
This map shows levels of SO_2 during 2012-16 with fitted polynomial trends before and after the opening of Line 1. Red Line indicates June 8th, 2014, the day Line 1 became operational.

Figure 3.4: PM_{10} Levels with Fitted Polynomial before and after Metro Line 1



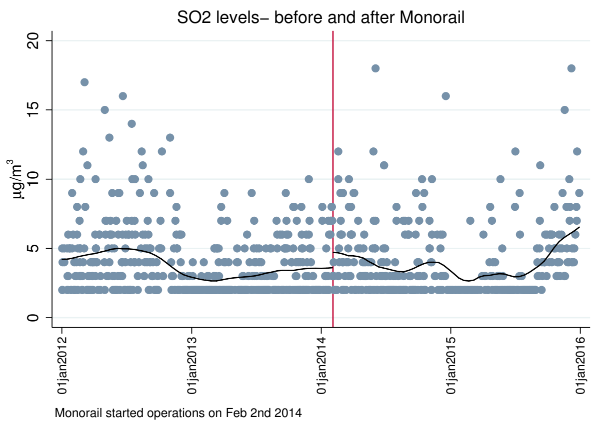
This map shows levels of PM_{10} during 2012-16 with fitted polynomial trends before and after the opening of Line 1. Red Line indicates June 8th, 2014, the day Line 1 became operational.

Figure 3.5: NO_2 Levels with Fitted Polynomial before and after Monorail



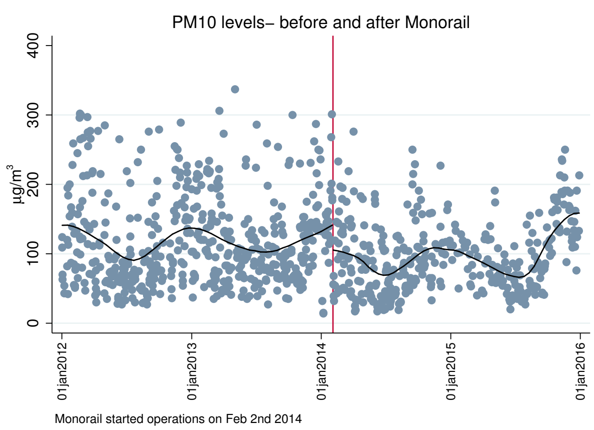
This map shows levels of NO_2 during 2012-16 with fitted polynomial trends before and after the opening of Phase 1 of Monorail. Red Line indicates Feb 2nd, 2014, the day Monorail became operational.

Figure 3.6: SO_2 Levels with Fitted Polynomial before and after Monorail



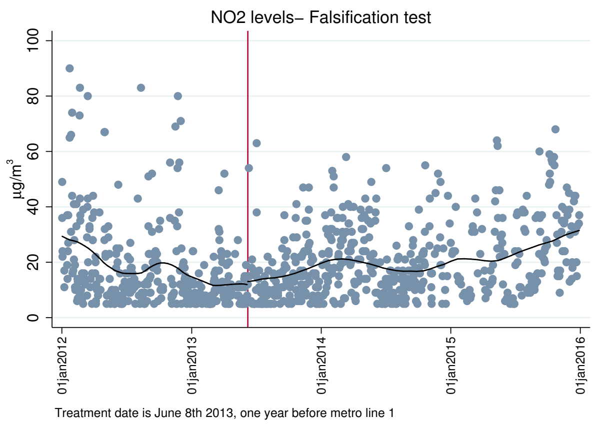
This map shows levels of SO_2 during 2012-16 with fitted polynomial trends before and after the opening of Phase 1 of Monorail. Red Line indicates Feb 2nd, 2014, the day Monorail became operational.

Figure 3.7: PM_{10}Levels with Fitted Polynomial before and after Monorail



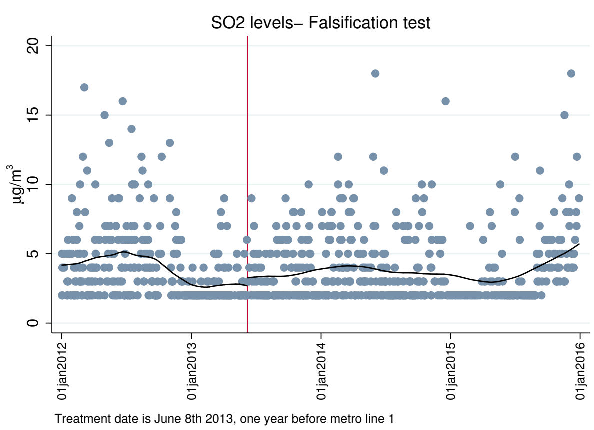
This map shows levels of PM_{10} during 2012-16 with fitted polynomial trends before and after the opening of Phase 1 of Monorail. Red Line indicates Feb 2nd, 2014, the day Monorail became operational.

Figure 3.8: NO_2 Levels with Fitted Polynomial before and after a Placebo Date



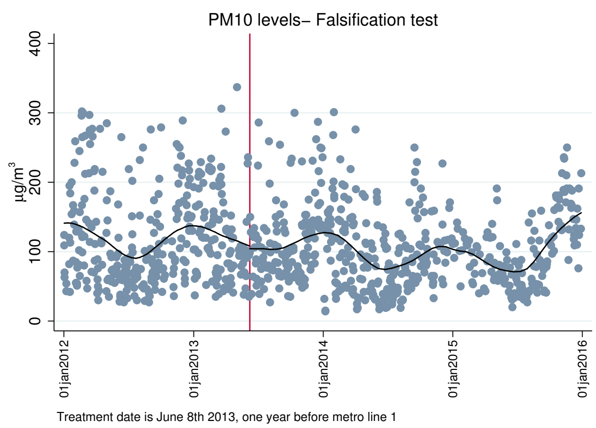
This map shows levels of NO_2 during 2012-16 with fitted polynomial trends before and after June 8th, 2013. Red Line indicates June 8th, 2013, one year before the day Line 1 became operational.

Figure 3.9: SO_2 Levels with Fitted Polynomial before and after a Placebo Date



This map shows levels of SO_2 during 2012-16 with fitted polynomial trends before and after June 8th, 2013. Red Line indicates June 8th, 2013, one year before the day Line 1 became operational.

Figure 3.10: PM_{10} Levels with Fitted Polynomial before and after a Placebo Date



This map shows levels of PM_{10} during 2012-16 with fitted polynomial trends before and after June 8th, 2013. Red Line indicates June 8th, 2013, one year before the day Line 1 became operational.

Table 3.1: Summary Statistics of Pollutants from Ground Monitoring Stations 2013-15 in μg/m^3

	N	Mean	Std. Dev.	Min	Max
SO_2	665	3.66	2.72	2	26
NO_2	664	18.31	12.89	5	68
PM_{10}	665	106.38	61.36	14	660

Table 3.2: Effects of the Opening of Metro Line 1 on Air Pollution (2014 sample)

	(1)	(2)	(3)	(4)	(5)	(6)
	logSO_2	logNO_2	logPM_{10}	logSO_2	logNO_2	logPM_{10}
Post-metro	-0.094	-1.064**	-0.428	-0.197**	-0.469***	-0.025
	(0.171)	(0.355)	(0.391)	(0.075)	(0.089)	(0.103)
Cubic time trend	✓	✓	✓	✗	✗	✗
Observations	236	235	236	236	235	236
R^2	0.12	0.31	0.25	0.11	0.27	0.24

Robust standard errors clustered at the monthly level are in parentheses. Each specification has a day of the week fixed effect and controls for air temperature, humidity, u and v components of the wind. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Effects of the Opening of Metro Line 1 on Air Pollution (2013-15 sample)

	(1)	(2)	(3)	(4)	(5)	(6)
	logSO_2	logNO_2	logPM_{10}	logSO_2	logNO_2	logPM_{10}
Post-metro	-0.050	-0.599**	-0.122	-0.005	0.242**	-0.215**
	(0.197)	(0.240)	(0.249)	(0.082)	(0.112)	(0.081)
Cubic time trend	✓	✓	✓	✗	✗	✗
Observations	664	663	664	664	663	664
R^2	0.21	0.37	0.25	0.10	0.22	0.23

Robust standard errors clustered at the month-year level are in parentheses. Each specification has day of the week and month fixed effects and controls for air temperature, humidity, u and v components of the wind. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Effects of the Opening of Metro Line 1 on log of Aerosol Optical Depth (2013-15 sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post-metro	0.171***	0.071***	-0.181***	0.072***	0.106***	-0.238***	-0.372***
	(0.002)	(0.002)	(0.011)	(0.002)	(0.005)	(0.012)	(0.011)
Cubic time trend	✗	✗	✓	✗	✗	✓	✓
Day of the week f.e.	✗	✗	✗	✓	✗	✓	✓
Year f.e.	✗	✗	✗	✗	✓	✓	✓
Quarter f.e.	✗	✗	✗	✗	✗	✗	✓
Weather Controls	✗	✓	✓	✓	✓	✓	✓
Observations	829,177	821,768	821,768	821,768	821,768	821,768	821,768
R^2	0.03	0.26	0.27	0.26	0.26	0.28	0.31

Weather controls include air temperature, u and v components of the wind, and water vapor levels. Robust standard errors clustered at the level of 2km grids are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Effects of a Placebo Event One Year before the Opening of Metro Line 1 on Air Pollution (2013 sample)

	(1) $\log SO_2$	(2) $\log NO_2$	(3) $\log PM_{10}$
Post-placebo	0.233	0.569	0.288
	(0.205)	(0.357)	(0.293)
Observations	266	266	266
R^2	0.15	0.15	0.16

Post-placebo is an indicator for the period after June 8th, 2013. Robust standard errors clustered at the monthly level are in parentheses. Each specification has a day of the week fixed effect, weather controls, and a cubic time trend. Weather controls include air temperature, u and v components of the wind, and water vapor levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Effects of Placebo Events 10 and 17 Days before the Opening of Metro Line 1 on Air Pollution (2013-15 sample)

	(1) $\log SO_2$	(2) $\log NO_2$	(3) $\log PM_{10}$	(4) $\log SO_2$	(5) $\log NO_2$	(6) $\log PM_{10}$
Post-placebo	0.194	-0.314	0.122	0.306	-0.229	-0.021
	(0.207)	(0.267)	(0.139)	(0.187)	(0.233)	(0.148)
Observations	664	663	664	664	663	664
R^2	0.21	0.35	0.25	0.21	0.35	0.25

Post-placebo dummy is the period after May 29th, 2014 in Columns 1-3 and the period after May 25th, 2014 in Columns 4-6. Robust standard errors clustered at the month-year level are in parentheses. Each specification has day of the week and month fixed effects, weather controls, and a cubic time trend. Weather controls include air temperature, u and v components of the wind, and water vapor levels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Effects of the Opening of Metro Line 1 estimated using 2013-15 sample excluding 7, 15, and 30 days before and after the Opening of Line 1

	(1) $\log SO_2$ 7 days	(2) $\log NO_2$ 7 days	(3) $\log PM_{10}$ 7 days	(4) $\log SO_2$ 15 days	(5) $\log NO_2$ 15 days	(6) $\log PM_{10}$ 15 days	(7) $\log SO_2$ 30 days	(8) $\log NO_2$ 30 days	(9) $\log PM_{10}$ 30 days
Post-metro	0.106	-0.452*	0.085	0.204	-0.491*	0.058	0.098	-0.479*	0.076
	(0.229)	(0.256)	(0.153)	(0.234)	(0.251)	(0.164)	(0.271)	(0.274)	(0.200)
Observations	656	655	656	645	644	645	622	621	622
R^2	0.21	0.37	0.25	0.22	0.37	0.25	0.22	0.37	0.23

The sample period is 2013-15. Columns 1-3 drop 7 days before and after the opening of Line 1. Columns 4-6 drop 15 days before and after the opening of Line 1. Columns 7-9 drop 30 days before and after the opening of Line 1. Robust standard errors clustered at the month-year level are in parentheses. Each specification has day of the week and month fixed effects, a cubic time trend, and controls for air temperature, humidity, u and v components of the wind. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Effects of the Opening of Monorail on Air Pollution (2013-15 sample)

	(1) logSO_2	(2) logNO_2	(3) logPM_{10}
Post-monorail	0.573***	0.321*	-0.061
	(0.133)	(0.182)	(0.138)
Observations	664	663	664
R^2	0.24	0.36	0.25

Robust standard errors clustered at the month-year level are in parentheses. Each specification has day of the week and month fixed effects, a cubic time trend, and controls for air temperature, humidity, u and v components of the wind. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix A: Public Transit Infrastructure and Employment Accessibility: The Benefits of the Mumbai Metro

A.1 Household Survey Description

Individual and household level data used in this paper are from a survey of 3,024 households representative of the Greater Mumbai Region (GMR) conducted by the World Bank in Jan-Feb 2019. Two members were interviewed in each household, an adult male and female (ages 18-45) with priority given to primary earners and/or decision makers of the household. The survey contains information on household members' education, occupation, income, household demographic composition, housing condition, household assets, pin code of the work location and commute trips from residence to workplace. Additionally, a travel diary was filled out by each of the main respondents for a 24-hour period with the following information for all trips taken on the chosen day: origin, destination, purpose, duration, time of day trip originated, distance traveled, mode(s) chosen, and the out-of-pocket cost. These data are described in Alam et al. (2021).

Definition of main mode: For the mode choice model discussed in Section 1.4, household location is treated as the origin, and a randomly selected post office that has the same pin code as the individual's workplace becomes the destination. Number of post offices per pin code in Mumbai ranges from 1 to 9, with the median being 4. Chosen travel mode in the model is the main mode that is defined using information available in the household survey. For individual workers who commute to work, the survey records up to three modes of transportation that workers usually use while commuting from residence to their usual workplace, time spent on each mode and the out-of-pocket cost for a one-way trip. When a mix of motorized and non-motorized transportation is used, the main mode is defined as the motorized mode of transportation on which maximum amount of time is spent. For example, if a person spends 15 minutes walking, 5 minutes on a two-wheeler, and 10 minutes on a train, then train is the main mode. If two modes are being used for the same duration, then the underrepresented mode is defined as the main mode. This is, however, a rare instance in the data. Main mode is non-motorized (walk or bicycle) only if that is the only reported travel mode. This definition is adopted from Takeuchi et al. (2007) which uses data from a similar survey conducted by the World Bank in Mumbai in 2004.

Definition of religion and language variables in Section 1.5: For each house, a 2 km neighborhood was defined using Euclidean distance. The median house has 117 neighbors. Proportion of neighbors with the same language and religion as the chooser household in the housing and mode choice model is calculated by matching the chooser's religion and language with that of the households in the neighborhood of a given house in the feasible set of the chooser household.

A.2 Commute Trips and Durations

I implement a program in Python using the packages, GOSTnets and NetworkX to construct travel time via drive, train, and walking.[1]. The program uses maps of the existing road and rail networks to find the shortest route via the specified network between origin-destination pairs using Dijkstra's algorithm. It converts distance along a path into travel time by dividing paths into smaller segments of equal lengths, computing travel time for each segment using user-specified speed information and adding together travel times for each segment along a path. For travel time by train, I assume that individuals can walk along the road network at a speed of 5 kmph, or cover the train route at a speed of 40 kmph for Mumbai suburban railway and 35 kmph for Metro rail. For walking, I assume a flat speed of 5 kmph along the road network. For drive times, I assume a flat speed of 20 kmph which is the median and modal speed in Mumbai in a 2015 dataset of traffic speeds in the city constructed using Google Directions API by Sarath Guttikunda. The reason for assuming flat speeds is the low variation in speeds observed in this dataset.

For a more accurate representation of the variation in traffic speeds, in the current set of results in Section 1.4, I construct travel time for motorized road trips using a Google Maps dataset of over 500,000 shortest duration driving trips, and 250,000 shortest duration transit trips between randomly selected origin-destination pairs in Mumbai that was compiled and generously shared by researchers at the Asian Development Bank for a more recent time period. I use randomly matched origin-destination pairs from this dataset that are within 1 km of the survey households' origin-destination points. The median distance between survey households and the origin point of a matched trip in this dataset is 148 meters.[2] In the data on transit trips, it is

[1] link for GOSTnets: *https://github.com/worldbank/GOST_PublicGoods/*

[2] The median distance between the post office and the destination of a matched trip is 717 meters; but, since the

not possible to distinguish between train and bus trips since this dataset only has total travel time information. Google Maps API can be used to obtain step-by-step detailed information for a trip that would allow this distinction, but rail transit information has not been available since the pandemic restrictions were implemented in Mumbai. To improve on this, I collect additional transit information using HERE API that has a similar data format, and provides historical information on transit travel time. Using detailed step-by-step trip details, I also defined separate measures of access time and transfer time using these data, allowing a test of sensitivity of commuting preferences to the definition of out-of-vehicle, and in-vehicle travel time.

In constructing the in-vehicle and out-of-vehicle time variables used in the main analysis of short-term commute mode choice in Section 1.4, travel time by train is always from the network program. Travel time by bus is from HERE data, whenever the information was available. In the absence of valid data from HERE, the maximum of Google Maps transit and Google Maps drive time is used.[3] If a suitable match was not available in Google Maps data, drive time calculated via network program was used. For the combined housing and mode choice analysis in Section 1.5, in the current set of results, all three travel times– via walking, rail and drive– are from the network program.

A.3 Employment Accessibility Index

Let j index the possible number of work locations in the city. The employment accessibility index for a house h is a travel time weighted average of the potential attractiveness of j as

post office is not the exact work location, this is simply classical measurement error.

[3] Sometimes HERE queries resulted in valid trips but missing travel times, while sometimes they returned completely empty results.

employment locations.

$$\text{EA}_h = \sum_j \left(\frac{w_j}{d_{hj}}\right)$$

w_j is the wage obtainable at location j. $d_{hj} = exp(\kappa * t_{hj})$ is the iceberg commuting cost from house h to location j. t_{hj} is the travel time between h and j. κ is a decay parameter specifying the semi-elasticity of commuting costs d_{hj} to commuting times t_{hj}. I use the methodology in Kreindler and Miyauchi (2021) to obtain a proxy for w_j and estimate κ.

The underlying model that allows identification of this proxy for wages is one where commuters choose an origin and destination for commute based on the characteristics of each location, and iceberg commuting costs. Under parametric assumptions on preferences, this model implies a probability of observing flows between a given origin-destination pair. This probability has a reduced-form representation in the gravity equation.

The probability of a worker living in location i and working in location j, π_{ij}, is given by a gravity equation of aggregate commuter flows,

$$\pi_{ij} = -\beta * t_{ij} + \gamma_i + \psi_j + \nu_{ij} \tag{A.1}$$

β captures the sensitivity of commuting decisions to commuting time. γ_i and ψ_j are origin and destination fixed effects that reflect residence and workplace amenities, respectively. ν_{ij} is the random error. I use data on commute flows from the household survey, considering flows between residence and work location pin codes. There are 85 unique residential pin codes and 88 unique work location pin codes in the data, implying a possible 7480 unique flows. I estimate this equation using Poisson pseudo-maximum likelihood estimator.[4] Estimates of work location

[4] I use aggregate commute flows instead of share of flows as the outcome variable because it provides a better

fixed effects, $\hat{\psi}_j$ are assumed to proxy the effective wage obtainable w_j. The correlation between $\hat{\psi}_j$ and average (maximum) income from the survey data at the level of work location pin code is 0.24 (0.48).

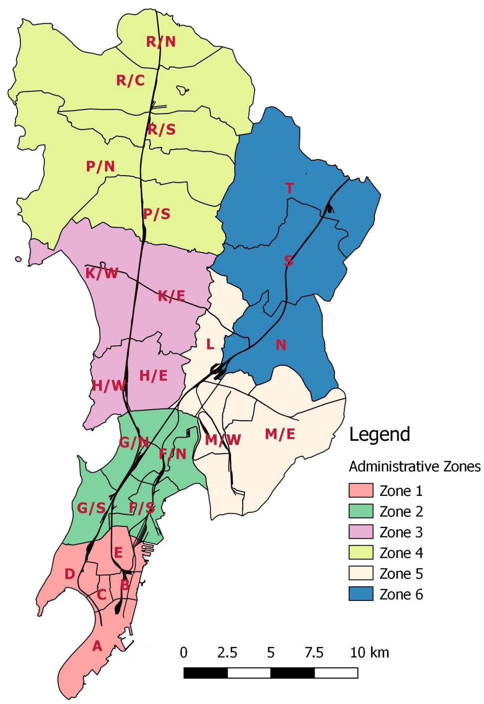
The sensitivity of commuting decisions to commute time, β is composed of two components: the semi-elasticity of commuting shares to commute costs (say, θ), and the semi-elasticity of commuting costs to commuting time (κ). Kreindler and Miyauchi (2021) argues that θ is inversely proportion to the semi-elasticity of actual work location wages to $\hat{\psi}_j$. I obtain θ by inverting the coefficient from an OLS regression of average incomes across work locations on $\hat{\psi}_j$. I then obtain $\kappa = \frac{\hat{\beta}}{\hat{\theta}}$. I find κ=0.029.[5] Note that $\hat{\psi}_j$ does not have a fixed scale, so I standardize EA$_h$ to be mean 0 with variance 1 for the second stage of the combined housing and mode choice model.

A.4 Appendix Figures and Tables

model fit without changing the results. The current parameterization uses geodetic distances in kilometer instead of travel time t_{ij}. But the results are robust to using travel times from the network program (as in Chapter 2) or Google Maps.

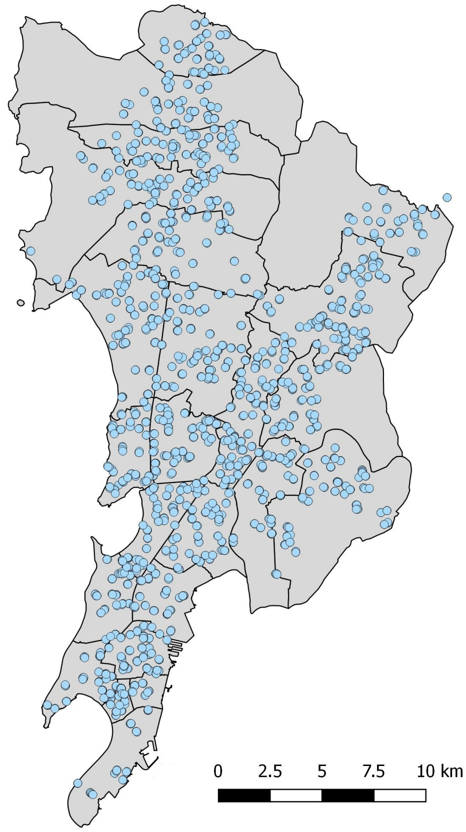
[5] Ahlfeldt et al. (2015) and Tsivanidis (2019) estimate this to be 0.001. Severen (2021) discusses the consequences of mismeasurement of this parameter.

Figure A.1: Administrative wards and zones in Mumbai



This map shows the 24 administrative wards in the Greater Mumbai Region. These wards are divided into six zones by the City for jurisdictional purposes indicated by the six colors. The existing rail lines (including Metro Line 1) are in black.

Figure A.2: Sample of Households in the World Bank 2019 Survey



This map shows the locations of households sampled for the World Bank Survey. Sampling was done in proportion to population at the ward level. Sample is representative at the ward and city levels.

Table A.1: Sensitivity of the commute mode choice model parameters to different definitions of IVT and OVT for Model 3 in Table 1.4

	(1)	(2)	(3)
OVT definition	Survey+NetworkX	Survey+HERE (includes transfer time)	Survey
IVT definition	GM+HERE +NetworkX	GM+HERE (excluding transfer time)	GM+HERE +NetworkX
Income-Cost	0.026***	0.024***	0.025***
	(0.001)	(0.001)	(0.001)
IVT	-0.011**	-0.016***	-0.007**
	(0.003)	(0.003)	(0.002)
OVT	-0.036***	-0.029***	-0.028***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Walk, Auto-rickshaw)	Omitted	Omitted	Omitted
(Car, Two-wheeler)	0.719***	0.312***	0.260***
	(0.077)	(0.078)	(0.078)
Train	-1.353***	-1.690***	-1.836***
	(0.086)	(0.083)	(0.081)
Bus	-2.038***	-2.283***	-2.426***
	(0.112)	(0.086)	(0.096)
Dissimilarity parameters:			
(Car, Two-wheeler)	1	1	1
	(constrained)	(constrained)	(constrained)
Train	1.000	1.000	0.872
	(64730.965)	(105822.833)	(85929.134)
Bus	1.000	1.000	1.064
	(159231.971)	(168960.780)	(132476.652)
(Walk ,Auto-rickshaw)	0.604***	0.154***	0.125***
	(0.058)	(0.018)	(0.013)
Individuals	2767	2754	2725
LR chi2	353.749	348.891	341.923
Log likelihood	-2866.638	-2841.136	-2796.132
IVT value (Rs. per minute)	0.414	0.671	0.272
OVT value (Rs. per minute)	1.383	1.235	1.135
Value of IVT (% wage)	33	35	14
Value of OVT (% wage)	70	64	59
Mean IVT (minutes)	16.15	14.37	16.24
Mean OVT (minutes)	10.34	16.60	15.47

This Table presents estimated preference parameters for the nested logit model in equation 1.1 for different definitions of in-vehicle time and out-of-vehicle time. In-vehicle time in Columns (1) and (3) for train is from the network program; for bus, it is from HERE Transit API and Google Maps API; and for the remaining options, it is from Google Maps API. In these two columns, out-of-vehicle time measures the first mile access. In Column (1), out-of-vehicle time for walk, train and bus are from the network program; for auto-rickshaw, it is from the survey. In Column (3), out-of-vehicle time is from the survey. Column (2) is the same as Column (1) except that out-of-vehicle time includes transfer time for bus and train from HERE API, and the same is excluded from in-vehicle time. Std. errors are in parentheses. Estimated parameters are based on the nesting structure in Model 3 of Table 1.4: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus). Estimated parameters based on Model 1 are in Table 1.6. Dissimilarity parameter is constrained to be ≤ 1 so that predictions are consistent with equation 1.3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Comparison of predicted mode shares of nested logit models from Table 1.4 with a mixed logit model (correlated random coefficients)

Travel modes	True shares	Model 1	Model 2	Model 3	Mixed Logit
Walk	32.2	28.1	34.3	27.4	29.8
Train	15.6	12.3	10.7	15.6	22.7
Bus	8.5	11.9	14.5	8.5	16
Auto-rickshaw	9.4	13.4	8.3	14.1	11.9
Two-wheeler	29.9	30.8	28.3	30.8	19.8
Car	4.4	3.5	3.9	3.5	1.8

This Table presents a comparison between the mode shares predicted under a mixed logit specification of the mode choice model, the three nested logit specifications, and true mode shares. The utility function specification for mixed logit model is $U_{im} = \alpha_1^i * t_{im}^{ivt} + \alpha_2^i * t_{im}^{ovt} + \alpha_3^i * (w_i - c_{im}) + \epsilon_{im}$ with ϵ_{im} i.i.d. random error following Type I extreme value distribution $\epsilon_{im} \sim f(\epsilon_{im}) = e^{-\epsilon_{im}} e^{-e^{-\epsilon_{im}}}$ and $\alpha_1^i, \alpha_2^i, \alpha_3^i \sim N(\alpha, \mathbf{\Sigma})$. Nesting structure in Model 1: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus); Model 2: (Walk, Car, Two-wheeler), (Train, Bus, Auto-rickshaw); Model 3: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus).

Table A.3: Commute mode choice models estimated under nested logit assumption with income entering non-linearly for three nesting structures

	(1)	(2)	(3)
Cost/Wage	-0.061***	-0.016***	-0.046***
	(0.005)	(0.002)	(0.004)
IVT	-0.012***	-0.015***	-0.012**
	(0.003)	(0.003)	(0.004)
OVT	-0.030***	-0.048***	-0.038***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Train, Bus)	Omitted		
(Train, Bus, Auto-rickshaw)		Omitted	
(Car, Two-wheeler)	2.258***		1.094***
	(0.096)		(0.118)
(Walk, Auto-rickshaw)	0.466*		Omitted
	(0.186)		
(Walk, Car, Two-wheeler)		1.875***	
		(0.063)	
(Train)			-0.950***
			(0.103)
(Bus)			-1.620***
			(0.134)
Dissimilarity parameters:			
(Car, Two-wheeler)	1		1
	(constrained)		(constrained)
(Bus, Train)	0.653***		
	(0.087)		
(Bus, Train, Auto-rickshaw)		1	
		(constrained)	
(Walk,Auto-rickshaw)	2.139***		1.241***
	(0.300)		(0.156)
(Walk, Car, Two-wheeler)		0.989***	
		(0.081)	
(Train)			1.000
			(152916.587)
(Bus)			1.000
			(525400.527)
Individuals	2792	2792	2792
LR chi2	328.043	489.042	359.843
Log likelihood	-2934.117	-2861.960	-2912.182
IVT value (% wage)	19.2	92.4	26.8
OVT value (% wage)	49.4	291.4	82.3

This Table presents estimated preference parameters for in-vehicle time in minutes (IVT), out-of-vehicle time in minutes (OVT) and cost/wage, the ratio of out-of-pocket cost per trip (in Rs.) to wage (in Rs. per minute). Wage per minute is calculated by scaling the monthly income with number of working days (22), working hours per day (9), and minutes in an hour (60). Nesting structure in Model 1: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus); Model 2: (Walk, Car, Two-wheeler), (Train, Bus, Auto-rickshaw); Model 3: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus). Dissimilarity parameter has been constrained to 1 so that predictions are consistent with equation 1.3. Walk also includes bicycle. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.4: Predicted mode shares of nested logit models from Tables 1.4 (income entering linearly) and A.3 (income entering as c/w)

		Income entering linearly			Income entering non-linearly		
Travel modes	True shares	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Walk	32.2	28.1	34.3	27.4	26.5	34.0	25.8
Train	15.6	12.3	10.7	15.6	14.2	10.8	16.3
Bus	8.5	11.9	14.5	8.5	10.5	14.7	8.5
Auto-rickshaw	9.4	13.4	8.3	14.1	14.7	8.5	15.4
Two-wheeler	29.9	30.8	28.3	30.8	30.6	27.9	30.3
Car	4.4	3.5	3.9	3.5	3.5	4.1	3.7

This Table compares sample commute mode shares with predicted mode shares from nested logit models in Table 1.4 where income-cost enters linearly and models in Table A.3 where income enters non-linearly to compare model fit. Nesting structure in Model 1: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus); Model 2: (Walk, Car, Two-wheeler), (Train, Bus, Auto-rickshaw); Model 3: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train), (Bus).

Table A.5: First stage for 2SLS regression in Table 1.12

	(1)	(2)	(3)	(4)	(5)
Log(Annual assessed sale value)	3616.928***	3530.585***	3541.733***	3543.660***	3102.384***
	(567.637)	(564.386)	(576.327)	(571.947)	(498.264)
Housing amenities index	2245.691***	2235.200***	2236.554***	2244.203***	2223.889***
	(140.494)	(145.748)	(146.707)	(146.266)	(145.336)
Distance to coast in km		31.351	28.316	-114.140	-59.242
		(77.192)	(81.422)	(100.283)	(93.945)
Slum classification dummy		-968.576***	-966.551***	-980.839***	-826.536**
		(331.921)	(330.823)	(324.195)	(335.226)
Crimes against women (No. of reports)		-2.782	-3.141	-4.157	10.826
		(13.200)	(13.466)	(13.493)	(14.113)
Distance to nearest station in km			37.504	-336.331	-380.073
			(209.158)	(277.820)	(282.361)
Standardized employment accessibility index				899.101**	686.476**
				(366.833)	(343.294)
Index for proximity to doctor/hospital					405.597***
					(134.658)
R-squared	0.378	0.382	0.382	0.389	0.406
Observations	2,209	2,209	2,209	2,209	2,024

This Table presents the first-stage estimates of the 2SLS specifications in Table 1.12. These parameters are in equation 1.9. Dependent variable is the monthly rental price of houses in Rs. Robust std errors clustered at the sub-zone level are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Factor loading for variables in Housing amenities index

Variable	Factor Loading
Good roof	0.247
Floorspace (in sqft.)	0.3786
Number of rooms	0.3719
Separate Kitchen	0.4493
Toilet inside the house	0.4332
Bathroom inside the house	0.3725
Piped water	0.3531
Well-connected to the city	0.0656
Footpath in the neighborhood	-0.007

This Table presents factor loadings for the Housing amenities index variable used in the second stage of the combined housing and mode choice model (Table A.5). These variables are summarized in Table 1.10.

Table A.7: Factor loading for variables in Index for proximity to doctor/hospital

Variable	Factor Loading
Pvt. Doctor/Clinic nearby	0.4071
Municipal Hospital nearby	0.6476
Pvt. Hospital/Nursing Home nearby	0.6441

This Table presents factor loadings for the proximity to doctor/hospital index variable used in the second stage of the combined housing and mode choice model (Column (5), Table A.5). Each of these variables have four categories increasing in order of proximity.

References

Ahlfeldt, Gabriel (2011), "If Alonso Was Right: Modeling Accessibility and Explaining the Residential Land Gradient." *Journal of Regional Science*, 51, 318–338. Wiley Online Library.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf (2015), "The Economics of Density: Evidence from the Berlin Wall." *Econometrica*, 83, 2127–2189. Wiley Online Library.

Akbar, Prottoy (2020), "Who Benefits from Faster Public Transit?" *Working paper*, URL `https://drive.google.com/file/d/1zexOoCxNZThfXl9wzk5DGFlvuIb4gJv7/view`.

Akbar, Prottoy A, Victor Couture, Gilles Duranton, Ejaz Ghani, and Adam Storeygard (2018), "Mobility and Congestion in Urban India." Working paper, World Bank, Washington, DC, URL `http://hdl.handle.net/10986/30236`.

Alam, Muneeza Mehmood, Maureen Cropper, Matias Herrera Dappe, and Palak Suri (2021), "Closing the Gap : Gender, Transport, and Employment in Mumbai." Working paper, World Bank, Washington, DC, URL `http://hdl.handle.net/10986/35248`.

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager (2021), "Synthetic difference-in-differences." *American Economic Review*, 111, 4088–4118.

Baker, Judy, Rakhi Basu, Maureen Cropper, Somik Lall, and Akie Takeuchi (2005), "Urban Poverty and Transport: The Case of Mumbai." Working paper, World Bank, Washington, DC.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2020), "On the road: Access to transportation infrastructure and economic growth in China." *Journal of Development Economics*, 102442. Elsevier.

Bauernschuster, Stefan, Timo Hener, and Helmut Rainer (2017), "When Labor Disputes Bring Cities to a Standstill: The Impact of Public Transit Strikes on Traffic, Accidents, Air Pollution, and Health." *American Economic Journal: Economic Policy*, 9, 1–37.

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

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007), "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115, 588–638. The University of Chicago Press.

Bayer, Patrick, Robert McMillan, and Kim Rueben (2004), "An Equilibrium Model of Sorting in an Urban Housing Market." Working Paper 10865, National Bureau of Economic Research, URL http://www.nber.org/papers/w10865.

Bel, Germà and Maximilian Holst (2018), "Evaluation of the impact of Bus Rapid Transit on air pollution in Mexico City." *Transport Policy*, 63, 209–220. Elsevier.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium." *Econometrica: Journal of the Econometric Society*, 841–890. JSTOR.

Billings, Stephen B (2011), "Estimating the Value of a New Transit Option." *Regional Science and Urban Economics*, 41, 525–536.

Borusyak, Kirill and Peter Hull (2020), "Non-Random Exposure to Exogenous Shocks: Theory and Applications." Technical report, National Bureau of Economic Research.

Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matějka, and Tobias Salz (2020), "The Value of Time: Evidence From Auctioned Cab Rides." Working Paper 27087, National Bureau of Economic Research, URL http://www.nber.org/papers/w27087.

Chacko, Benita (2018), "'Mumbai to have 300-km Metro in 5-6 years', says metropolitan commissioner." URL https://indianexpress.com/article/cities/mumbai/mumbai-to-have-300-km-metro-in-5-6-years-says-metropolitan-commissioner-5094492. Indian Express (March 12, 2018).

Chen, Yihsu and Alexander Whalley (2012), "Green Infrastructure: The Effects of Urban Rail Transit on Air Quality." *American Economic Journal: Economic Policy*, 4, 58–97.

Chowdhury, Sourangsu, Sagnik Dey, Sachchida Nand Tripathi, Gufran Beig, Amit Kumar Mishra, and Sumit Sharma (2017), ""Traffic Intervention" Policy Fails to Mitigate Air Pollution in Megacity Delhi." *Environmental science & policy*, 74, 8–13.

Concas, Sisinnio and Alexander Kolpakov (2009), "Synthesis of Research on Value of Time and Value of Reliability." Technical report, URL https://www.nctr.usf.edu/pdf/77806.pdf.

Craig, Andrea (2019), "Commute Mode and Residential Location Choice." Working paper, URL http://web2.uwindsor.ca/economics/RePEc/wis/pdf/1904.pdf.

Cropper, Maureen and Palak Suri (2022), "Air Quality Impacts of Metro Rail in Mumbai." *Working paper*.

Currie, Janet and Reed Walker (2011), "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics*, 3, 65–90.

Davis, Lucas W (2008), "The Effect of Driving Restrictions on Air Quality in Mexico City." *Journal of Political Economy*, 116, 38–81. The University of Chicago Press.

Dey, Sagnik, Larry Di Girolamo, Aaron van Donkelaar, SN Tripathi, Tarun Gupta, and Manju Mohan (2012), "Variability of Outdoor Fine Particulate ($PM_{2.5}$) Concentration in the Indian Subcontinent: A Remote Sensing Approach." *Remote sensing of environment*, 127, 153–161.

Diao, Mi, Delon Leonard, and Tien Foo Sing (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. Elsevier.

DNA India (2019), "Mumbai: Big boost for BEST, 36 lakh more board the bus (Jul 16, 2019)." URL `https://www.dnaindia.com/mumbai/report-mumbai-big-boost-for-best-36-lakh-more-board-the-bus-2772348`.

Donaldson, Dave (2018), "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review*, 108, 899–934.

Dubé, Jean, Marius Thériault, and François Des Rosiers (2013), "Commuter Rail Accessibility and House Values: The Case of the Montreal South Shore, Canada, 1992–2009." *Transportation Research Part A: Policy and Practice*, 54, 49–66.

Duranton, Gilles and Matthew A Turner (2011), "The Fundamental Law of Road Congestion: Evidence from US Cities." *American Economic Review*, 101, 2616–52.

Gangan, Surendra P (2017), "Mumbai monorail a waste of public money, says Maharashtra public accounts committee." URL `https://www.hindustantimes.com/mumbai-news/mumbai-monorail-a-waste-of-public-money-says-maharashtra-public-accounts-committee/story-jUwiiq9P2ZQPspULcdoDXI.html`. Hindustan Times (Dec 16, 2017).

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner (2018), "Subways and Urban Air Pollution." Working Paper 24183, National Bureau of Economic Research, URL `http://www.nber.org/papers/w24183`.

Gibson, Matthew and Maria Carnovale (2015), "The Effects of Road Pricing on Driver Behavior and Air Pollution." *Journal of Urban Economics*, 89, 62–73.

Glaeser, Edward L, Matthew E Kahn, and Jordan Rappaport (2008), "Why Do The Poor Live In Cities? The Role of Public Transportation." *Journal of Urban Economics*, 63, 1–24. Elsevier.

Goel, Deepti and Sonam Gupta (2017), "The Effect of Metro Expansions on Air Pollution in Delhi." *The World Bank Economic Review*, 31, 271–294. Oxford University Press.

Gu, Yizhen, Chang Jiang, Junfu Zhang, and Ben Zou (2021), "Subways and Road Congestion." *American Economic Journal: Applied Economics*, 13, 83–115.

Guo, Shihong and Liqiang Chen (2019), "Can urban rail transit systems alleviate air pollution? empirical evidence from Beijing." *Growth and Change*, 50, 130–144. Wiley Online Library.

Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine E Kontokosta (2020), "Take the Q Train: Value Capture of Public Infrastructure Projects." *NBER Working Paper*.

Hansen, Walter G (1959), "How Accessibility Shapes Land Use." *Journal of the American Institute of Planners*, 25, 73–76. Taylor & Francis.

Harari, Mariaflavia (2020), "Cities in Bad Shape: Urban Geometry in India." *American Economic Review*, 110, 2377–2421.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm (2020), "The Making of the Modern Metropolis: Evidence from London." *The Quarterly Journal of Economics*, 135, 2059–2133. Oxford University Press.

Hindustan Times (2017), "Mumbai locals: World's busiest urban rail system is also the deadliest." URL `https://www.hindustantimes.com/india-news/mumbai-locals-world-s-busiest-urban-rail-system-is-also-the-deadliest/story-zqbW39tWfd5yzB18DzZePI.html`.

IQAir (May 1, 2022), "New WHO air quality guidelines will save lives." *Blog*, URL `https://www.iqair.com/blog/air-quality/2021a-WHO-air-quality-guidelines`.

Khanna, Gaurav, Carlos Medina, Anant Nyshadham, Daniel Ramos, Jorge Tamayo, and Audrey Tiew (2021), "Spatial Mobility, Economic Opportunity, and Crime." *Working paper*, URL `https://612bbf4b-3210-4927-957e-9e9090dd882c.filesusr.com/ugd/f85d25_c4edf93a599944a3a840699e3d8f8595.pdf`.

Kling, Catherine L and Cynthia J Thomson (1996), "The Implications of Model Specification for Welfare Estimation in Nested Logit Models." *American Journal of Agricultural Economics*, 78, 103–114. Oxford University Press.

Koh, Yumi, Jing Li, and Xu Jianhuan (2021), "Subway, Collaborative Matching, and Innovation." *Working paper*, URL `https://www.dropbox.com/s/sc7gyx8wobqyyab/BJSubway%20%281%29.pdf?dl=0`.

Koppelman, Frank S and Chandra Bhat (2006), "A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models." Technical report, FTA US Department of Transportation.

Korde, Kailash (2018), "In reverse gear: Fleet of Mumbai public buses reduces, but number of pvt vehicles increases." URL `https://www.hindustantimes.com/mumbai-news/in-reverse-gear-fleet-of-mumbai-public-buses-reduces-but-number-of-pvt-vehicles-increases/story-vSpQRb8o2eQE7SwiY4tjmO.html`. Hindustan Times (Apr 29, 2018).

Kreindler, Gabriel E and Yuhei Miyauchi (2021), "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records." Working Paper 28516, National Bureau of Economic Research, URL `http://www.nber.org/papers/w28516`.

Kwon, Eunjee (2020), "Why Do Improvements in Transportation Infrastructure Reduce the Gender Gap in South Korea?" *Working paper*, URL `https://www.dropbox.com/s/g5llvksna4onr8t/120220_JMP_EJK.pdf?dl=0`.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack R Porter (2021), "Valid t-ratio Inference for IV." Working Paper 29124, National Bureau of Economic Research, URL http://www.nber.org/papers/w29124.

Li, Shanjun, Yanyan Liu, Avralt-Od Purevjav, and Lin Yang (2019), "Does subway expansion improve air quality?" *Journal of Environmental Economics and Management*, 96, 213–235. Elsevier.

McFadden, Daniel (1973), "Conditional Logit Analysis of Qualitative Choice Behavior." *Frontiers in Econometrics, Ed. by Zarembka, Paul*. Academic Press: New York.

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

Mohring, Herbert (1972), "Optimization and scale economies in urban bus transportation." *The American Economic Review*, 62, 591–604.

Mulalic, Ismir and Jan Rouwendal (2020), "Does improving public transport decrease car ownership? Evidence from a residential sorting model for the Copenhagen metropolitan area." *Regional Science and Urban Economics*, 83, 103543. Elsevier.

Osland, Liv and Inge Thorsen (2008), "Effects on Housing Prices of Urban Attraction and Labor-Market Accessibility." *Environment and Planning A*, 40, 2490–2509. SAGE Publications Sage UK: London, England.

Pandey, Anamika, Michael Brauer, Maureen L Cropper, Kalpana Balakrishnan, Prashant Mathur, Sagnik Dey, Burak Turkgulu, G Anil Kumar, Mukesh Khare, Gufran Beig, et al. (2021), "Health and Economic Impact of Air Pollution in the States of India: the Global Burden of Disease Study 2019." *The Lancet Planetary Health*, 5, e25–e38.

Pathak, Rahul, Christopher K Wyczalkowski, and Xi Huang (2017), "Public Transit Access and the Changing Spatial Distribution of Poverty." *Regional Science and Urban Economics*, 66, 198–212. Elsevier.

Prasad, Rachita (2015), "Reliance Infrastructure completes refinancing of Rs 1,650 crore Mumbai Metro borrowing." URL https://economictimes.indiatimes.com/industry/indl-goods/svs/engineering/reliance-infrastructure-completes-refinancing-of-rs-1650-crore-mumbai-metro-borrowing/articleshow/47629986.cms?from=mdr. The Economic Times (June 11, 2015).

Redding, Stephen J and Esteban Rossi-Hansberg (2017), "Quantitative Spatial Economics." *Annual Review of Economics*, 9, 21–58. Annual Reviews.

Redding, Stephen J and Matthew A Turner (2015), "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*, volume 5, 1339–1398. Elsevier.

Reddy, Alla V, Alex Lu, and Ted Wang (2010), "Subway Productivity, Profitability, and Performance A Tale of Five Cities." *Transportation research record*, 2143, 48–58. SAGE Publications Sage CA: Los Angeles, CA.

Schlenker, Wolfram and W Reed Walker (2016), "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies*, 83, 768–809. Oxford University Press.

Severen, Christopher (2021), "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification." Working paper, URL https://cseveren.github.io/files/Severen_LAMetro_Pretty.pdf.

Simeonova, Emilia, Janet Currie, Peter Nilsson, and Reed Walker (2018), "Congestion Pricing, Air Pollution and Children's Health." Technical report, National Bureau of Economic Research.

Small, Kenneth A (2012), "Valuation of travel time." *Economics of transportation*, 1, 2–14. Elsevier.

Small, Kenneth A and Harvey S Rosen (1981), "Applied Welfare Economics with Discrete Choice Models." *Econometrica: Journal of the Econometric Society*, 105–130. JSTOR.

Small, Kenneth A, Erik T Verhoef, and Robin Lindsey (2007), *The Economics of Urban Transportation*. Routledge.

Suri, Palak (2022), "Public Transit Infrastructure and Employment Accessibility: The Benefits of the Mumbai Metro." *Working paper*.

Takeuchi, Akie, Maureen Cropper, and Antonio Bento (2007), "The Impact Of Policies To Control Motor Vehicle Emissions In Mumbai, India." *Journal of Regional Science*, 47, 27–46.

Tan, Brandon and Kwok-Hao Lee (2020), "Urban Transit Infrastructure and Inequality: The Role of Access to Non-Tradable Goods and Services." *Available at SSRN 3750438*, URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3750438.

Tsivanidis, Nick (2019), "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." *Working paper*, URL https://static1.squarespace.com/static/55bb98e2e4b0ba843f39599e/t/5d039b14655edc0001d608a9/1560517406526/TsivanidisTransMilenio_6.2019.pdf.

Urban Emissions, Mumbai (2018), "Urban Emissions report for Mumbai, Maharashtra." URL https://urbanemissions.info/india-apna/mumbai-india/.

Varian, Hal R (1992), *Microeconomic Analysis*. WW Norton.

Warnes, Pablo Ernesto (2020), "Transport Infrastructure Improvements and Spatial Sorting: Evidence from Buenos Aires." *Working paper*, URL https://pewarnes.github.io/files/warnes_pablo_jmp.pdf.

Zhou, Zhengyi, Hong Chen, Lu Han, and Anming Zhang (2019), "The Effect of a Subway on House Prices: Evidence from Shanghai." *Real Estate Economics*.