The Benefits of Public Transit to Households: Evidence from India*

Palak Suri† Maureen Cropper‡

August 9, 2024

Abstract

We measure benefits to households from Mumbai's new Metro rail system. We estimate a commute mode choice model to value commute time savings in the short run and a housing choice model to value the improved commuting utility that households experience due to spatial sorting. Aggregate benefits from Metro rail are approximately 10 times higher when spatial sorting occurs. In the short run women, college-educated workers, and workers with above median incomes experience higher benefits than their opposites. In the long run, households with lower incomes and assets benefit more than their wealthier counterparts.

JEL Codes: R2, R4, R1, O18

*We thank the World Bank for sharing the household survey data used in this paper, the Asian Development Bank for sharing relevant travel time data for Mumbai used to conduct sensitivity analyses in the paper, Sarath Guttikunda for sharing information on traffic speeds in Mumbai, and Jianguo Ma for facilitating access to ArcGIS Pro. We are grateful to Yi Jiang, Rana Hasan, Marjorie Villanueva Remolador and Jade Laranjo for their invaluable time and support during early phases of this project. We thank Prottoy Akbar, Mrinmoyee Chatterjee, Jessie Handbury, Misty Heggeness, Brad Humphreys, Sonakshi Jain, Vasudha Jain, Gaurav Khanna, Guido Kuersteiner, Aastha Malhotra, Alejandro Molnar, Victoria Perez-Zetune, Anusuya Sivaram and Sergio Urzúa for their helpful comments and discussions.

†palak.suri@mail.wvu.edu; Department of Economics, West Virginia University

‡mcropper@umd.edu; Department of Economics, University of Maryland and Resources for the Future

1 Introduction

The benefits of transport projects are often measured by the value of time savings they deliver. This approach underlies the World Bank's evaluation of transit projects and is also followed by cities in the United States (U.S. DOT (2016a)), the United Kingdom (U.K. DOT (2023)) and Canada (Alberta Transportation (2017)). The value of time savings can be estimated using a commute mode choice model, which assumes fixed residential and work locations. Due to this assumption, this approach is most appropriate for measuring short run benefits. Long run benefits are typically higher as mobility constraints are relaxed, but since politicians invest in infrastructure projects based on the salience of short run costs and benefits to the target voter base, this can lead to sub-optimal spending (Glaeser and Ponzetto (2018)). Due to weak institutions, cities in developing countries are even more susceptible to this.

In this paper, we leverage detailed household-level information and discrete choice models to contrast the short run and the long run benefits to households from Mumbai's new Metro system. We estimate the annual aggregate short run benefits from Mumbai's first Metro line to be \$51 million 2019 PPP dollars and aggregate long run benefits to equal \$565 million 2019 PPP dollars. Long run benefits exceed the construction cost of the Metro, but not the short run benefits. We compare this with a cost-benefit analysis of a more widely accessible Metro project that competes with the existing railway network in the city and find that annualized long run benefits exceed the construction cost only at low discount rates. Our results validate the presence of discounted foreign bank loans for large infrastructure projects, without which such infrastructure projects may not exist. Lastly, we investigate the nature of spatial and demographic heterogeneity in benefits generated from these Metro projects, highlighting the labor market benefits along the intensive margin.

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 of the city's growing economy. Private vehicle ownership has increased four-fold in the last two decades, resulting in severe traffic congestion problems and constrained intracity commutes. To alleviate congestion and improve commuting conditions, over 300 km of Metro rail lines have been planned. The first part of the Metro network, an 11.4 km east-west link (Line 1), opened in 2014. An additional 92 km of the Metro network (Lines 2, 3 and 7) was scheduled to open in 2022. We study the benefits to households from these two infrastructure projects. In doing so, we also inform transit policy by comparing benefits for two very different types of Metro projects.

We obtain data from a representative household survey conducted by the World Bank in 2019 with household locations, work locations, commute modes chosen by household workers, vehicle ownership, and housing characteristics and costs. We supplement this with travel times from Google Maps, HERE, and a network algorithm and information on travel costs. We first estimate a discrete choice model in which workers select an optimal commute mode for residence-to-work commutes, assuming fixed residence and work locations. This yields estimates of commuter preferences for in-vehicle travel time, out-of-vehicle travel time, and income net of commuting costs. In doing so, we add to the literature on the value of time by providing preference estimates for travel time for a developing country. We use these estimates to compute expected compensating variation for commute time savings generated by Metro Line 1 and Lines 2, 3 and 7 (Small and Rosen (1981) and Kling and Thomson (1996)). This provides a measure of short run welfare.

In the long run, since households can move residential locations, welfare is likely higher. We therefore estimate a housing choice model: households select a house, assuming fixed work locations, based on the expected utility from the commute mode decision and other housing and neighborhood characteristics (Barwick et al. (2021)). We use preference estimates from this model to compute a longer run welfare measure–the expected compensating variation for counterfactual improvements in commuting utility due to Metro projects.

The short run benefits implied by the commute mode choice model accrue to only those workers whose commute time between their residence and work location is reduced. This could be due to newly accessible Metro stations or more efficient transit routes. The mean expected compensating variation conditional on benefits being positive implies that the average beneficiary would be willing to pay 12-16% more than their current out-of-pocket expenditure for these time savings benefits. Long run welfare accrues to households experiencing improved commuting utility due to Metro projects linking workers' job locations and affordable houses in the city. The mean expected compensating variation implied by this model is 2% of average monthly rent for Line 1 and 6% of average monthly rent for lines 2, 3 and 7. The value to an average beneficiary in the long run is 2.4 times the short run value for Line 1 and 5.5 times the short run value for Lines 2, 3 and 7.

How benefits are distributed among various consumers is an important policy consideration. We examine the spatial heterogeneity in short run benefits by residence and work locations, and demographic heterogeneity by estimating commute mode choice models by gender, education, and income levels. In the long run, certain households may benefit more from re-sorting due to the geography of jobs. We therefore examine how long run benefits vary by work location. We also condition preferences on vehicle ownership, income, and the education level of the primary worker of the household.

In the short run, women benefit more than men, workers with at least a college education benefit more than those with less than a college education, and workers with above-median incomes benefit more than those with below-median incomes. Travel time savings generated by the Metro projects are similar for different sub-groups of individuals in the sample. Therefore, the heterogeneity in benefits is due to differences in preferences. In the long run, households with lower incomes, less education, and no vehicles experience greater benefits. Their initial commuting constraint, relieved by Metro rail, is stronger than for economically advantaged households.

Despite its smaller length, Line 1 yields average short run benefits per beneficiary that are 80% of the short run benefits per beneficiary due to Lines 2, 3 and 7, thus highlighting the importance of its strategic location. Line 1 provided the first east-west rail link in the city, while most of Lines 2, 3 and 7 run north-south, parallel to the existing Suburban Railway network. Lines 2, 3 and 7 are, however, much greater in length, and thus yield greater aggregate benefits, both in the short and the long run, than Line 1.

Our paper contributes to the growing literature on benefits of urban transit projects in developing world cities. We focus on a first order benefit, improved commute times and utilities, which is likely to be the most salient factor in policymakers' decision. Other types of benefits of urban transport projects include the impact on air pollution (reviewed in Li et al. (2020) and Cropper and Suri (2023)), population changes (Glaeser et al. (2008), Pathak et al. (2017), Khanna et al. (2021)), employment (Kwon (2022), Tyndall (2021)), innovation (Koh et al. (2021)), vehicle ownership (Mulalic and Rouwendal (2020)), congestion (Anderson (2014), Gu et al. (2021)), trade and economic growth (Donaldson (2018), Banerjee et al. (2020)), and access to consumption amenities (Zheng et al. (2016), Lee and Tan (2024)). A recent strand of literature has also examined the general equilibrium welfare impacts of transit in London (Heblich et al. (2020)), Los Angeles (Severen (2021)), Bogotá (Tsivanidis (2019)), and Buenos Aires (Warnes (2020)). While ours is a partial equilibrium approach, we emphasize on heterogeneity in household preferences and benefits generated, which is less prevalent in this literature.

The paper is organized as follows. Section 2 describes stylized facts about transportation in Mumbai. Section 3 describes the mode choices and characteristics of commuters in the household survey. The commute mode choice and housing location choice models are presented in Section 4 and the data used to estimate them in Section 5. The results of model estimations are reported in Section 6 and the welfare estimates that they imply in Section 7. Section 8 concludes.

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 in the world. Mumbai is the core of the MMR, with a population of 12.5 million in the 2011 Census. 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.

Mumbai's 24 administrative wards are divided into 6 zones (Appendix Figure B1). 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 public transit and lower property prices in the suburbs. Generally, population and employment are concentrated along the main rail lines. In the 2013 Employment Census, 80% of workers in formal jobs were concentrated in Zones 1-4. About 70% of individuals lived in Zones 1-4 according to the 2011 Census.

The Mumbai Suburban Railway, shown in black in Figure 1, is the heart of the city's public transit network; however, Suburban trains have severe overcrowding problems.[1] An extensive network of public buses complements the rail system, but its importance has been declining. Average daily bus ridership in 2019 was 2 million passengers (DNA India (2019)), compared to 4.2 million in 1997-98 (Korde (2018)). This is likely due to traffic congestion in the city and poor upkeep of public buses. Rising household incomes have also led to a sharp increase in private vehicle ownership. The two-wheeler and car population in the city increased by 340% and 200%, respectively, from 2000 to 2017, further contributing to traffic congestion.[2,3]

The Metro rail project was planned to alleviate Mumbai's congestion problems. The existing and planned lines are shown in Figure 1. Line 1 became operational in 2014. Lines 2, 3 and 7 were expected to become operational in 2021-22. Parts of Lines 2a (North-South part of the red line in Figure 1) and 7 became operational in 2022-2023, but the rest are still under construction. Line 2b will provide another east-west rail link, while the remaining parts of the upcoming network will run parallel to the Mumbai local railway network.

[1] There are about 14-16 passengers per sqm of floor space (Hindustan Times (2017)).

[2] 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)

[3] In 2018, the average speed in the city during morning rush hours was 22 kmph, with a peak traffic speed of 7 kmph. 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)).

3 Mode Choices and Individual Characteristics

This paper uses information on the residential location of households, household members' workplace locations and residence-to-work commute patternsfrom a survey conducted by the World Bank in January-March 2019 (Alam et al. (2021)).[4] 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. The location of sampled households is shown in Appendix Figure B2. While we do not know the exact work location of individuals in the survey, we denote the work location by a randomly selected post office with the same pin code as the work location.[5] In this section, we describe the commuting patterns based on this survey, which form the basis for estimating the models in Sections 4.1 and 4.2.

The existing pattern of residential and employment locations in the city determines the extent of short run commute benefits from Metro rail. 72% of sampled workers work in Zones 1-4, while 68% live in Zones 1-4. Commute trips are generally short: 50% of commuters travel less than 2.5 km to get to their work location. 75% of all workers work in the same zone as their residence.[6]

The success of new transit projects depends not only on their placement but also on commute mode choices. In our sample, 8-10% of workers work from home. Table 1 shows the main commute modes for workers who commute to work. 33% travel by foot or bicycle. 24% use public transportation: 16% for train and 8% for bus. 10% use auto-rickshaws or taxis, and 34% use private two-wheelers or four-wheelers. Distances traveled by commuters are in the last row. These suggest that mobility is generally low in the city and that public transit facilitates longer commute distances, with train riders commuting over 10 km one-way (Table 2).

Vehicle ownership and income are important determinants of commute mode choices: 50% of workers live in households with at least one vehicle, an increase from 20% in 2004 (Baker et al. (2005)). Columns 6 and 7 of Table 1 show modal shares by vehicle ownership. 70% of workers with a vehicle use it as their main commute mode, while the 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

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

[5] 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.

[6] In 2004, the median commuter traveled 2.9 km to get to work (Baker et al. (2005))

reflect potential heterogeneity in the distributional effects of Metro rail. The difference in average commute distance between men and women is not statistically significant, 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 (4 vs 5.7 km) and are more likely to walk to work than workers with a college education (41% vs 15%). College educated commuters are more likely to use train and private vehicles because of their higher incomes and greater commute distances. Workers earning above median incomes travel significantly greater distances (5km) than those earning less (4.3 km). They are more likely to travel via private vehicle than walk or use public transit. The models discussed in the following sections use these observations to estimate the underlying preference parameters.

4 Models

In this section, we develop models to value the time savings benefits of Metro rail in the short and long run. We begin with a model which characterizes commute mode choice as a function of in-vehicle travel time, out-of-vehicle travel time, and income net of travel costs. The commute mode choice model allows us to estimate individual preferences for time and costs, accounting for tastes for mode types such as public transit or private vehicles. Using these preferences, we estimate how much each individual would pay for the potential time savings due to Metro rail, holding their utility constant at the pre-policy level, i.e., we estimate their expected compensating variation. This is our measure of short run benefits due to Metro rail.

To measure long run benefits we estimate a model of residential location choice in which expected utility from the commute mode choice model enters the household's utility function, in addition to other housing amenities and housing cost (Barwick et al. (2021)). Households choose the house that maximizes their utility, holding work location fixed. We use this model to estimate households' compensating variation for Metro rail, which alters the maximum utility from commuting for each house in the household's choice set.

Our modeling approach allows us to separately measure the benefits that can be strictly attributed to commute time savings affecting individuals' optimal commute mode decisions (short run welfare) and those that arise due to household re-sorting in response to the potential time savings arising at different locations, conditional on their optimal commute mode decisions (long run welfare).

4.1 Commute Mode Choice

The motorized commute modes in Mumbai include bus and train, auto-rickshaw and taxi, and two-wheeler and car. Non-motorized modes (biking or walking) constitute one-third of commutes (see Table 1). We classify similar modes in categories to account for tastes for commute mode types. For example, individuals may have a preference for private vehicles or a distaste for public transit or walking.

Assuming fixed residence and work locations, individual i chooses a travel mode $m \in$ M={Walk, Bus, Train, Auto-rickshaw, Two-wheeler, Car} to maximize their utility. These modes can be classified into mutually exclusive categories or nests denoted by B_k. Our preferred nesting structure has B_1={Walk, Auto-rickshaw} , B_2 = {Bus, Train}, B_3={Two-wheeler}, B_4={Car}.

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, an unobserved nest-specific preference, and an individual-specific idiosyncratic random shock.[7] We assume a linear additive random utility function,

$$\begin{aligned} U_{imB_k} &= V_{imB_k} + \epsilon_{imB_k} \\ V_{imB_k} &= \alpha_1 * t^{ivt}_{imB_k} + \alpha_2 * t^{ovt}_{imB_k} + \alpha_3 * (w_i - c_{imB_k}) + \delta_{B_k} \end{aligned} \tag{1}$$

where $t^{ivt}_{imB_k}$ 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 utility. $w - c$ enters the model linearly for computational simplicity.[8] ϵ_{imB_k} is an i.i.d. random utility shock assumed to follow a generalized extreme value distribution.

$$\epsilon_{imB_k} \sim exp\Bigg(\sum_{k}^{K} \big(\sum_{m \in B_k} -e^{\epsilon_{imB_k}/\lambda_{B_k}} \big)^{\lambda_{B_k}} \Bigg) \tag{2}$$

In this specification, we assume that for any two alternatives m_1 and m_2 in nest B_k , $\epsilon_{im_1 B_k}$ is correlated with $\epsilon_{im_2 B_k}$.[9] Any two alternatives across nests are assumed to be uncorrelated,

[7]We 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.

[8]Allowing it to enter non-linearly as Cost/Wage lowers the estimated preferences for travel time slightly, but the model fit is similar. We discuss robustness in Section 6.1.

[9]For sensitivity analysis, we also consider a model where there is only one nest $K = 1$, and unobserved heterogeneity in preferences for in-vehicle and out-of-vehicle travel time that maybe correlated. We allow the parameters α_1 and α_2 to vary by individual and follow a joint Gaussian distribution. Nested logit is the

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

$$P_{im} = \frac{e^{V_{im}/\lambda_{B_k}} \left(\sum_{j \in B_k} e^{V_{ij}/\lambda_{B_k}} \right)^{\lambda_{B_k} - 1}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} e^{V_{ij}/\lambda_{B_l}} \right)^{\lambda_{B_l}}} \tag{3}$$

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 $\frac{\alpha_1}{\alpha_3}$ and the value of out-of-vehicle time is $\frac{\alpha_2}{\alpha_3}$. 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 measures the value of small changes in travel time reasonably well, but not large changes such as those brought about by new infrastructure projects since it does not allow modal shares to change in response to the policy (Train (2009)).

Welfare: To measure the value of changes in travel times in the commute model choice model we compute expected compensating variation, CV_i (Small and Rosen (1981), Varian (1992), Small et al. (2007)). CV_i measures willingness to pay for travel time changes induced by Metro rail, holding utility constant at the pre-policy level.

$$E\big[\max_m U(t^{ivt,0}_{imB_k}, t^{ovt,0}_{imB_k}, w^0_i - c^0_{imB_k}, \delta^0_{B_k})\big] = E\big[\max_m U(t^{ivt,1}_{imB_k}, t^{ovt,1}_{imB_k}, w^0_i - c^0_{imB_k} - CV_i, \delta^0_{B_k})\big] \tag{4}$$

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)),

$$\frac{1}{\alpha_3}\Big[ln\big[\sum_k \big(\sum_m e^{(V^1_{imB_k}/\lambda_{B_k})}\big)^{(\lambda_{B_k})}\big] - ln\big[\sum_k \big(\sum_m e^{(V^0_{imB_k}/\lambda_{B_k})}\big)^{(\lambda_{B_k})}\big]\Big] \tag{5}$$

The short run benefits implied by the commute mode choice model accrue to only those

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 6.1). 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.

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. In Section 7.1, we estimate the commute mode choice model and short run benefits for all sampled commuters and also for sub-groups of commuters (women, college-educated workers, workers with above-median incomes).

4.2 Housing Location Choice

Assuming a fixed work location, household i chooses a house h from the set of feasible housing alternatives, based on various housing amenities, including the expected utility of commuting, and housing cost.[10]

$$\begin{aligned} U_{ih} &= V_{ih} + \epsilon_{ih} \\ V_{ih} &= \beta_K * K_{ih} + \beta_Z * Z_{ih} + \alpha_p * \mathrm{P}_h + \alpha_x * X_h + \nu_h \end{aligned} \tag{6}$$

V_{ih} refers to the deterministic portion of the utility function. ϵ_{ih} is the household-specific idiosyncratic shock component assumed to follow an i.i.d. Type I extreme value distribution. K_{ih} represents the attractiveness of house h to household i in terms of the ease and comfort of its workers' commute. It is the expected utility from commuting between house h and the fixed work locations of each worker of household i via the optimal commuting mode. This expected maximized commuting utility is derived using preference parameters from equation 1 for each worker i_g and is combined to obtain a household-level expected commuting utility as given below.[11]

$$K_{ih} = \mathbb{E}_g\Bigg[ln \sum_k \Bigg(\sum_m e^{(\hat{V}_{i_g hmB_k}/\hat{\lambda}_{B_k})}\Bigg)^{\hat{\lambda}_{B_k}}\Bigg] \tag{7}$$
$$\text{where } \hat{V}_{i_g hmB_k} = \hat{\alpha}_1 * t^{ivt}_{i_g hmB_k} + \hat{\alpha}_2 * t^{ovt}_{i_g hmB_k} + \hat{\alpha}_3 * (w_{i_g} - c_{i_g hmB_k}) + \hat{\delta}_{B_k}$$

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 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.

[10] Each house is assumed to represent a housing type.

[11] The functional form is due to the idiosyncratic shock being GEV distributed. Expectation is taken across workers within the household to obtain household-level average expected commuting utility. Employment location information is available for two workers of the household.

β_Z and β_K capture average preferences for attributes in Z_{ih} and K_{ih}. In some specifications, we use household characteristics such as education of the primary worker, income level, and vehicle ownership as taste shifters. In those cases, β is composed of two components, one that is constant across all households, and one that is constant across all households 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.

Welfare: Long run welfare is given by the expected compensating variation for improved average commuting utility between a house and the household's workers' work locations due to the addition of Metro 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)).

$$\frac{1}{\alpha_p}\left[ln\Big(\sum_h e^{V_h(P_h^0, X_h^0, Z_{ih}^0, K_{ih}^1)}\Big) - ln\Big(\sum_h e^{V_h(P_h^0, X_h^0, Z_{ih}^0, K_{ih}^0)}\Big)\right] \tag{8}$$

Superscript 0 denotes baseline values of commuting utility under the existing rail network, while 1 denotes the utility from the optimal commute after the policy change.

In the housing choice model, time savings benefits accrue via improved commuting utility between the household's workers' work location and any house that the household has a positive probability of selecting. Commuting utility increases whenever the transit time is reduced between a house-work location pair.

5 Data

Commute Mode Choice Model: The 2019 World Bank household survey has information on up to three modes for a typical residence-work commute, along with time spent in each mode. The chosen travel mode in the commute model choice 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.

To estimate preferences for travel time, we need travel time and costs between residential and work locations for all feasible travel modes, in addition to the chosen mode. We compile this information using multiple sources. The computation of in-vehicle and out-of-vehicle travel times for each mode and each origin-destination pair is described in Appendix Section A.2.

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.[12] For two-wheeler and car, assuming a mileage of 26 kilometer per liter (kmpl) and 12 kmpl, we calculate the cost per trip km using the prevailing petrol price in Mumbai at the time (Rs. 86.16 per liter). We multiply the commute distances by the cost per km to calculate out-of-pocket costs.

Table 2 presents data on travel time, out-of-pocket cost, distance to work location, and the average monthly commuter's income by the main commute mode chosen. Both average travel time and distance 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 car have a higher income than the remaining sample. The distribution of average monthly incomes for train users indicates that train users include both low and high income commuters.

Housing Choice Model: The housing choice model is estimated using a sample of 2,170 households for whom complete information is available on relevant household characteristics and neighborhood and housing amenities. These are presented in Table 3. The average monthly household income is Rs. 30,939 ($1,454 PPP) with an average imputed monthly rent equal to 32% of household income. 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, we compute, for each house in the sample, the proportion of households within a 2 km radius with a given religion and language.

To obtain maximum expected commuting utility for each household-house pair, we compute the household average value of K_{ih} (equation 7) for all commutes between the work locations of workers of household i and houses in household i's choice set. Household expected commuting utility is the housing amenity through which we measure households' long run welfare due to Metro rail.

Housing characteristics common to all households are summarized in the second panel of Table 3. The average floorspace is 263 sqft, with the median house having only a single room. 59% of the houses have a separate kitchen space. Many houses do not have a toilet or bathroom inside the house; households living in these houses must rely on communal facilities. Access to public transit is good. Mean distance to the nearest railway station is

[12]Our conclusions are robust to using fares for commuters with a monthly or quarterly pass for bus and train.

1.5 km, which is an 18-minute walk assuming a walking speed of 5 kmph. nearest bus stop for most houses in the sample is within a 5-minute walk. We estimate preferences for a general employment accessibility index as well.

We control for employment accessibility as an amenity by constructing an index which is a commuting-cost-weighted average of the attractiveness of pin codes as employment locations. Let j index work locations in the city. The employment accessibility index for house h is defined as

$$\text{EA}_h = \sum_j \left(\frac{w_j}{d_{hj}}\right) \tag{9}$$

where w_j is the wage obtainable at location j and $d_{hj} = exp(\kappa * t_{hj})$ is the iceberg commuting cost from house h to location j. t_{hj} is the travel time between h and j. $\kappa > 0$ is the semi-elasticity of commuting costs d_{hj} to commuting times t_{hj}. In the absence of data on wages, we use the method in Kreindler and Miyauchi (2023) to construct a proxy for w_j. We estimate a reduced-form gravity equation of aggregate commute flows between pairs of pincodes as a function of the fastest travel time between the two pincodes and fixed effects for origin and destination pin codes. Parameter estimates of destination fixed effects represent the relative attractiveness of locations for employment and serve as a proxy for w_j. The elasticity of commute flows to commute time obtained as the parameter on travel time and the relationship between estimated destination fixed effects and aggregate wages by pincode is used to estimate κ. The estimation is discussed in Appendix Section A.3. We use the standardized values of EA_h as a housing amenity.[13]

6 Estimation and Results

6.1 Commute Mode Choice

We estimate the nested logit model in equations 1 and 2 using maximum likelihood estimation. We classify commute modes into nests based on 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 a two-wheeler or a car because of their logistical infeasibility in certain locations or for longer distances.[14] Commuters are also more likely to choose train for longer commute distances (Table 2)

[13] Results are robust to using an employment accessibility index which is a travel-time-weighted average of effective wages across the city, with time t_{hj} in the denominator instead of equation 13.

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

which could indicate an absence or unreliability of buses on certain routes.

Due to arbitrariness in classifications, we present estimated parameters for three different nesting structures in Table 4: Column 1 {(Walk, Auto-rickshaw), (Bus, Train), (Two-wheeler), (Car)}, Column 2 {(Walk, Auto-rickshaw), (Bus, Train), (Two-wheeler, Car)}, and Column 3 {(Walk, Auto-rickshaw), (Train), (Bus), (Car, Two-wheeler)}. Model 1 is our preferred specification: it is consistent with our beliefs about mode substitution behavior in Mumbai and fits the data well. Empirically, Model 2 also performs well, therefore, Models 1 and 2 are the focus of our analysis. The only difference between the two models is that Model 1 assumes independence between idiosyncratic preference shocks for Car and Two-wheeler, whereas Model 2 does not. Model 3 assumes that idiosyncratic preference shocks for bus and train are independent, which is less likely to be true. The modal share predictions generated by the three models are in Table 5.[15]

Results in Table 4 indicate a distaste for longer commute times. The distaste for out-of-vehicle time is greater than the distaste for in-vehicle time, a common finding in the literature (Small et al. (2007), Chapman et al. (2006), Koppelman and Bhat (2006), Buchholz et al. (2020)). Estimated nest-intercepts indicate that mean utility for private and flexible modes is higher than for bus or train. On average, the value of in-vehicle time is Rs 0.77-0.82 per minute, and the value of out-of-vehicle time Rs. 1.41-1.45 per minute, about 40-42% and 73-74% of the mean wage, respectively.

The estimated preference parameters lie in the range of estimates commonly used for transport policy analysis. The U.S. Department of Transportation recommends setting the value of time equal to the median gross wage (U.S. DOT (2016a), U.S. DOT (2016b)), or at 80-120% of the wage rate to allow for uncertainty. Small et al. (2007) and Concas and Kolpakov (2009) report a range of estimates from the literature ranging from 20% to over 100% of the average wage across countries. 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 equal the mean wage.

These estimates are robust to different measures of in-vehicle and out- of-vehicle travel time. We estimate Models 1 and 2 using transit time information from HERE so that transfer times for bus and rail are included in out-of-vehicle time. We also use out-of-vehicle time exclusively from the household survey assuming households behave according to their perceptions of travel time, and not necessarily the actual time. Results for Models 1 and 2 are in Appendix Tables B1 and Table B2, respectively. The values of time implied by

[15]Degenerate nests have a dissimilarity parameter of 1. To obtain predictions of modal shares consistent with equation 3, we constraint the dissimilarity parameters to 1 whenever they exceed 1 by a significant magnitude.

these alternative definitions are similar to Table 4.

For comparison, we estimate a nested logit model with income entering non-linearly as cost/wage and a mixed logit model allowing correlated heterogeneity in preferences for time. 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. Appendix Table B4 compares modal shares predicted by the nested logit models in Table 5 with the nested logit models with income entering non-linearly, and a mixed logit model.[16] The nested logit models perform better than the mixed logit specification in terms of predictions. Predictions from the preferred model specification with income entering linearly are similar to those from the model with income entering non-linearly.

6.2 Housing Choice

We estimate the housing choice model following the two-step approach of Berry et al. (1995), Bayer et al. (2004), and Bayer et al. (2007). In the first stage, we estimate the parameters of equation 10 using maximum likelihood estimation.

$$U_{ih} = \beta_K * K_{ih} + \beta_Z * Z_{ih} + \delta_h + \epsilon_{ih} \tag{10}$$

The housing-specific variables in equation 6 are subsumed in δ_h, the housing specific constant that captures the mean utility for house h. Each house observed in the survey is assumed to represent a housing type. House (housing type) 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. The estimation sample has 2,170 households choosing among 2,170 houses. Since the number of alternatives available per household is large, for computational reasons, we take a random sample of the feasible set in estimating the model (McFadden (1978)).[17]

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 household.

$$\hat{\delta}_h = \alpha_p * \mathrm{P}_h + \alpha_x * X_h + \nu_h \tag{11}$$

We use the estimated housing-specific intercepts from the first stage to estimate equation 11 using two-stage least squares. Unobserved housing attributes omitted from this equation

[16] Preference parameters from the model with income entering non-linearly are in Appendix Table B3. Parameter estimates for the mixed logit model are not shown for brevity.

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

contained in ν_h are likely correlated with rental price. Therefore, we instrument for rental price using the assessed value of properties in the neighborhood of h. Neighborhoods for the purposes of value assessments, called sub-zones, are defined by the municipal government based on historical boundaries, land regulations, market values, and market potential. Assessments are used for collecting transaction and property taxes, and may differ from the market price of houses. For house h in sub-zone s_h, rental price can be written as a function of assessed values in sub-zone s_h.

$$\mathrm{P}_h = \omega * \text{Assessed value}_{s_h} + \alpha_x * X_h + \zeta_h \tag{12}$$

This instrument is valid as long as it is not correlated with ν_h. This is likely to be satisfied given the heterogeneity in housing types within a sub-zone s. This is also likely to be true given the low correlation between observed amenities and assessed values (-0.2 - 0.09 for most amenities, and $\sim$ -0.3 for proximity to coast and employment accessibility). The same argument, however, suggests that the instrument may be weak. We show robustness to the weak instrument problem using the inference criteria suggested in Lee et al. (2022).

Preference parameters in equation 10 estimated using a conditional logit specification are presented in Table 6. Households have a preference for houses that offer a higher commuting utility and for living close to other households with the same religion and language. Parameter estimates are robust to whether expected commuting utility is computed using commute mode choice Model 1 or Model 2. To test the sensitivity of commuting utility preference parameters to the first-stage specification, we estimate specifications allowing for the preferences for commuting utility and proximity to households with similar language and religion to vary by income, education and vehicle ownership. Preferences for expected commuting utility from the first-stage of these models are in Table 7.

Preferences for commuting utility are heterogeneous, i.e., some household types place a higher value on the possible commuting utility when selecting a house (Table 7). Households with a primary worker without a college degree have a stronger preference for houses with higher commuting utility than households whose primary worker has a college degree. Households with below median incomes have a stronger preference for commuting utility than their higher-income counterparts. In contrast, households that own a vehicle have a stronger preference for commuting utility than their counterparts. This is likely due to the fact that most households with a vehicle own a two-wheeler, which is more convenient for shorter distances.[18]

[18]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.

In the second stage, estimated house-specific intercepts from Model 1 in Table 6 are regressed on housing-specific characteristics using two-stage least squares, with the log of assessed values for residential properties used as an instrument for monthly rental price. Table 8 presents these regressions for different sets of control variables. With standard errors clustered at the level of sub-zones, the first-stage F-statistic is ∼40.[19] Given the possibility of this instrument being weak, we use an adjusted critical t-value for inference at the 95% confidence level following Lee et al. (2022). These are reported in the last row of Table 8. 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 Appendix Table B7.[20]

Table 8 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 with a higher accessibility to potentially attractive work locations. While the consistency of the parameters estimated 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 contains a housing amenity index which is the first principal component of the housing amenities available in the survey.[21] Columns 2-5 add additional amenities: distance to coast, distance to the nearest railway station, slum classification of the residential area, number of reported crimes against women, an index of employment accessibility, and an index for proximity to doctor and hospitals.[22] The coefficient on distance to the nearest railway station captures the average disamenity associated with being close to a congested transit access point. The fact that access to a transit stop reduces the travel time needed to reach an employment location is captured by the employment accessibility index.[23] Column 4 is our preferred specification: it allows the greatest number of controls without the loss of sample size due to missing observations.

[19] According to Lee et al. (2022), first stage F-statistic below 100 may indicate a weak instrument problem.

[20] The second-stage results for Model 2 in Table 6 and the models in Table 7 are similar, therefore, we report only the results for the main specification.

[21] These include floorspace, number of rooms and dummy variables for good roof, separate kitchen, separate toilet, bathroom inside the house, and access to piped water. Factor loading of each of these variables in shown in Appendix Table B5.

[22] The index of proximity to health services is the first principal component of related variables from the survey, including categorical variables for walk time 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 B6.

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

7 Welfare Estimates

7.1 Short run Commuter Welfare

In 2019, the rail network in Mumbai consisted of the Suburban Railway and Metro Line 1. We compute counterfactual in-vehicle and out-of-vehicle travel times via rail by (i) removing Line 1 from the 2019 rail network, and (ii) adding Lines 2, 3 and 7 to the 2019 rail network. 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.

We compute expected compensating variation to value the time savings benefits due to Line 1 and the upcoming Lines, using parameters from Models 1 and 2 in Table 4. Since the models do not account for preferences for infrastructure quality, these benefits do not capture the improved utility from a more comfortable Metro rail infrastructure relative to the Suburban Railway. The estimates, therefore, likely understate benefits.

The monetary benefits of travel time savings are presented in Table 9. 25% of commuters have a positive willingness to pay for benefits due to Line 1, while 57% 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. 77 per month for Line 1 and Rs. 98 per month for Lines 2, 3 and 7. The value of benefits as a proportion of average out-of-pocket commuting cost is 12% for Line 1 and an 14% for Lines 2, 3 and 7. The corresponding values implied by Model 2 are similar (second panel of Table 9).

The spatial distribution of expected compensating variation highlights the nature of travel time benefits. Figure 2 shows benefits from the two Metro projects by household location. Many more commuters benefit from Lines 2, 3 and 7 due to the wider accessibility of the network (92 km). Commuters in the vicinity of Metro stations experience the highest benefits, mainly due to reductions in out-of- vehicle access times. But commuters in other parts of the city also experience benefits due to improved transit connections. This is especially so for Line 1, which provided the first east-west rail link in the city.

Benefits aggregated to the level of work locations highlight which parts of the city benefit in the short run due to improved transit availability. Figure 3 shows the share of short run welfare by pincode of work location. Roughly 70% of the benefits are concentrated in pincodes within 5 km of Line 1. Benefits from Lines 2, 3 and 7 are much more dispersed across the city. These patterns confirm that benefits from a widely distributed transit network are much more spatially dispersed than those from a small, geographically restricted network.

Equity is an important consideration in transit infrastructure planning. Depending on the demographic concentration of workers across residences and pincodes and the spatial concentration of jobs and housing types, 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, we estimate the commute mode choice model (equation 1) for various subgroups of individuals distinguished by gender, education and income level. Table 9 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 33% greater benefits than men due to Line 1 and 11% greater benefits due to Lines 2, 3 and 7.[24] 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. On average, women also have a lower distaste for public transit compared to men. This is reflected in the fact that a greater proportion of women use publicly available modes despite traveling similar distances as men (Table 1).[25]

Transport infrastructure projects have strategic importance in enabling certain sub-groups of the population to participate 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 (2022)). In the context of Mumbai, transport availability may not be the biggest factor constraining women's labor force participation (Alam et al. (2021)), but Table 9 indicates 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.

Reductions in travel times for workers with and without a college degree are similar. However, the former group experiences higher benefits under both Metro projects because workers with a college degree have a stronger distaste for both in-vehicle and out-of-vehicle travel times. They also commute longer distances compared to individuals with less than a

[24] The difference in benefits due to Line 1 is significant at the 95% confidence level, while the difference for Lines 2, 3 and 7 is significant at 80-85% confidence level.

[25] Difference in the value of time are due to differences in preferences for time and for mode category. Gender differences in preferences for mode categories might reflect access to household vehicles, preference for traveling in groups, or safety concerns. Women's preferences for public transit may reflect the availability of women-only coaches in trains and reserved seating for women on buses.

college education (Table 1). Their monthly expected compensating variation, conditional on positive benefits, is 67-96% higher than for workers below college education for Line 1 and 47-72% higher for Lines 2, 3 and 7.[26] College educated workers in the sample are also more likely to have above median incomes.

Commuters with above median-incomes experience greater benefits than commuters with below-median incomes. The value of time savings due to Line 1 is 41% higher for above-median income commuters, while the value of time savings is 14% higher for Lines 2, 3 and 7.[27] Both groups experience similar time savings, indicating that this pattern is due to differences in preferences. We 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.[28] The dispersion in monetized benefits accruing to different sub-groups is higher under the upcoming network than under Line 1, which accords with the greater length of the upcoming lines.[29]

7.2 Long run Household Welfare

The above estimates capture benefits to commuters if their residence and work locations are fixed. We also compute long run benefits, allowing households to choose a house that offers higher utility for its workers' commute.

We find that average value of improved commuting welfare from Line 1 if households are allowed to re-sort is about 1.4-2% of monthly rent or about Rs. 135-190 per month. These results, shown in Table 10, are stable across the various second-stage specifications. In comparison, the corresponding value of improved commuting utility due to Lines 2, 3 and 7 is 4-5.5% of monthly rent or Rs. 380-535 per month.[30] The larger mean expected compensating variation for lines 2,3 and 7 is not surprising, given the much larger extent of these lines relative to Line 1.[31]

In Figure 4, we examine the distribution of long run benefits by residential location. The greatest benefits from Metro Line 1 accrue to households located in the center and northern parts of the city, whereas benefits from Lines 2, 3 and 7 also extend to the southern tip of the

[26]Differences for both projects are significant at the 99% confidence level.

[27]The differences are significant only for Model 2: for Line 1 at 95% confidence levels and for Lines 2, 3 and 7 at 90% confidence levels.

[28]There is limited statistical power to separately estimate the model for these subsamples.

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

[30]Previous versions of this paper featured a model where households chose commute mode and housing location simultaneously, producing similar results (Suri (2022)).

[31]These welfare estimates are also robust to assumptions of preference heterogeneity in the first-stage (Appendix Table B8).

city due to the length of the network. Almost every household experiences benefits in the long run, whereas short run benefits accrue only to those whose current residence-work commute would benefit from the Metro.

Given the geography of existing jobs, neighborhood sorting in response to the Metro means that work locations accessible via Metro rail should experience the greatest benefits. This is seen in Figure 5, which shows the share of long run welfare by pincode of work location. The share of long run benefits to workers from Line 1 is higher for workers in the middle of the city. Long run benefits from Lines 2, 3 and 7 are more pronounced for workers employed along the upcoming network, although large shares of benefits accrue to workers in the very north and south of the city.

Long run benefits are more likely to accrue to those demographic groups who can move to a better house to take advantage of the Metro network. This depends on households' workers' work locations as well as the distribution of housing amenities and rents across locations. To examine the distribution of benefits across groups, we estimate the main specification for the entire sample and summarize the expected compensating variation for different sub-samples. Results are in Table 11.

There is no significant difference in the benefits of Line 1 based on education but households without a vehicle, and those below the median income experience significantly greater benefits from Line 1 than households owning a vehicle or households with above-median incomes. Similarly, households with below-college educations, lower incomes, or who do not own a vehicle benefit more from Lines 2, 3 and 7 in the long run than their opposites. Households with greater incomes and assets have fewer commuting constraints to begin with, and most of the benefits accrue to households for whom Metro rail relaxed a significant constraint.

Few other papers in the literature have examined heterogeneity in the benefits from transit project. In the context of Bogotá, Tsivanidis (2019) finds that the introduction of Bus Rapid Transit (BRT) generated greater general equilibrium welfare gains for high-skilled workers. This is consistent with our short run results, however, we find that households with less educated workers experience greater long run benefits. In Buenos Aires, Warnes (2020) finds that the BRT benefited high-skilled and low-skilled workers similarly. In the context of U.S., Akbar (2020) finds that rail transit speed 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. This is weakly consistent with our short run results but not with our long run results. 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 (Tan and Lee (2020)).

7.3 Benefits Cost Analysis

The short run benefits of Metro rail accrue to only a fraction of commuters– 25% for Line 1 and 57% for Lines 2, 3 and 7. When households are allowed to adjust their housing location, practically every household experiences positive benefits. Conditional on benefits being positive, the average long run benefits are more than twice the short run benefits for Line 1 and more than five times the short run benefits for Lines 2, 3 and 7. Relaxing the assumption of fixed housing location has a larger impact on the benefits of Lines 2, 3 and 7 because they extend the Metro network more than Line 1.

To estimate aggregate benefits of each Metro project, we scale household-level mean expected compensating variation using ward-level sample weights to obtain annual aggregate benefits at the city level. Total annual short run benefits of Line 1, which accrue to only one-quarter of the city's commuters are \$51 million (PPP), while the long run benefits, which accrue in expectation to all households, are 11 times this value (\$565 million PPP).[32] Total annual short run benefits due to Lines 2, 3 and 7 are \$170 million (PPP), and their long run benefits are 9 times the short run value (\$1,560 million PPP). Despite the much smaller length of Line 1, its aggregate benefits are 36% 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.

In Table 12, we compare both short run and long run aggregate benefits of each Metro project with the equivalent annualized capital cost of construction (EACC) based on various assumptions about the discount 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)).[33] Interest rates on loans for other parts of the Mumbai Metro network have been more favorable at 1-2% (CareEdge Ratings (2023)). Another option is to use an interest rate that is closer to the long run yield on government bonds, rates of 10% and 8%. 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). The equivalent annualized capital cost (EACC) ranges from \$ 1.9-3 billion (PPP).

[32]Households in the sample were chosen such that there was at least one working member, and one male and one female respondent available. Since each chosen household has at least one working member, we scale household-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. We compute these benefits at the annual level. It is assumed that households that drop out of the sample due to missing information are randomly spatially distributed.

[33]Metro Line 1 is operated by a Public Private Partnership.

Under no set of assumptions do the short run benefits of either Line 1 or Lines 2, 3 and 7 cover the construction costs of the lines; however, the long run benefits of Line 1 do exceed the annualized construction cost. This is true of Lines 2, 3 and 7 only at a discount rate of 2 percent or lower. Table 12 illustrates the difficult choices facing decisionmakers: short run benefits rarely cover costs (even construction costs) and certainly not construction and operating costs. We note, however, these benefits are only due to improvements in commute time and commuting utility and therefore, capture only the first-order direct benefits to households. Other papers in the literature have considered agglomeration benefits (Heblich et al. (2020), Severen (2021), Tsivanidis (2019), Warnes (2020)), air pollution benefits (reviewed in Li et al. (2020) and Cropper and Suri (2023)), and impacts on poverty (Pathak et al. (2017)), crime (Khanna et al. (2021)), employment (Kwon (2022), Tyndall (2021)), innovation (Koh et al. (2021)), vehicle ownership (Mulalic and Rouwendal (2020)), congestion (Anderson (2014), Gu et al. (2021)), trade and economic growth (Donaldson (2018), Banerjee et al. (2020)) and access to consumption amenities (Zheng et al. (2016), Lee and Tan (2024)).

8 Conclusion

In this paper, we estimate structural models of commute mode choice and housing choice using data from a 2019 household survey in Mumbai. We use these estimates to compute short run and long run benefits to households from an existing and an upcoming Metro rail project in Mumbai. 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), which was scheduled to open in 2022, but is only partly operational. The reason for building Metro rail was to alleviate overcrowding problems facing the historic Suburban Railway network and road traffic congestion by moving commuters towards Metro rail. This is expected to improve intracity commuting accessibility. The benefits of Lines 2, 3 and 7 are greater than those of Line 1 because of the greater extent of the network. But the benefits per km are higher for Line 1.

To compute short run benefits, we use estimated preferences of commuters for in-vehicle travel time, out-of-vehicle travel time, travel costs and commute modes while assuming fixed residence and work locations. We compute long run benefits using estimated preferences of households for commuting utility in addition to other housing amenities and housing cost while assuming fixed work location. Household sorting leads to the aggregate value of long run benefits being substantially higher than short run benefits.

There are important spatial and demographic heterogeneities in benefits. Consistent with findings in the limited literature evaluating heterogeneity, we find that workers living close to Metro stations, women, workers with a college education or with above-median incomes receive disproportionately greater short run benefits than their opposites. This is due to difference in preferences as each Metro project generates similar travel time reductions for each group. In the long run, however, households with lower incomes and assets experience greater benefits.

References

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.

Alberta Transportation (2017), "Benefit Cost Model." Technical report, URL https://www.alberta.ca/benefit-cost-model-and-user-guide.

Anderson, Michael L (2014), "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion." *American Economic Review*, 104, 2763–96.

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.

Barwick, Panle Jia, Shanjun Li, Andrew R Waxman, Jing Wu, and Tianli Xia (2021), "Efficiency and equity impacts of urban transportation policies with equilibrium sorting." Technical report, National Bureau of Economic Research.

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.

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

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.

CareEdge Ratings (2023), "Mumbai Metro Rail Corporation Limited." URL https://www.careratings.com/upload/CompanyFiles/PR/202310131058_Mumbai_Metro_Rail_Corporation_Limited.pdf.

Chapman, Bruce, Hiroyuki Iseki, Brian D Taylor, and Mark Miller (2006), "The effects of out-of-vehicle time on travel behavior: Implications for transit transfers (deliverable# 1)." *CA: California Department of Transportation*.

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 (2023), "Measuring the air pollution benefits of public transport projects." *Regional Science and Urban Economics*, 103976.

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.

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.

Glaeser, Edward L and Giacomo AM Ponzetto (2018), "The Political Economy of Transportation Investment." *Economics of Transportation*, 13, 4–26.

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

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.

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 (2023), "Measuring commuting and economic activity inside cities with cell phone records." *Review of Economics and Statistics*, 105, 899–909.

Kwon, Eunjee (2022), "Why Do Improvements in Transportation Infrastructure Reduce the Gender Gap in South Korea?" *Working paper*, URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4081288.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter (2022), "Valid t-ratio Inference for IV." *American Economic Review*, 112, 3260–3290.

Lee, Kwok Hao and Brandon Joel Tan (2024), "Urban transit infrastructure and inequality." *Review of Economics and Statistics*, 1–46.

Li, Shanjun, Jianwei Xing, Lin Yang, and Fan Zhang (2020), "Transportation and the Environment in Developing Countries." *Annual Review of Resource Economics*, 12, 389–409.

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

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.

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).

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.

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), *The Benefits of Metro Rail in Mumbai, India: Reduced Form and Structural Approaches*. Ph.D. thesis, University of Maryland, College Park.

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.

Train, Kenneth E (2009), *Discrete Choice Methods with Simulation*. Cambridge university press.

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.

Tyndall, Justin (2021), "The local labour market effects of light rail transit." *Journal of Urban Economics*, 124, 103350.

U.K. DOT (2023), "Transport analysis guidance unit a1.1."

U.S. DOT (2016a), "Benefit-Cost Analysis Guidance for Rail Projects." Technical report, U.S. Department of Transportation, Federal Railroad Administration, URL https://www.dot.ny.gov/divisions/operating/opdm/passenger-rail/passenger-rail-repository/FRA%20Benefit-Cost%20Analysis%20Guidance%20for%20Rail%20Projects.pdf.

U.S. DOT (2016b), "The Value of Travel Time Savings: Departmental Guidance for Conducting Economic Evaluations." Technical report, U.S. Department of Transportation, URL https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-valuation-travel-time-economic.

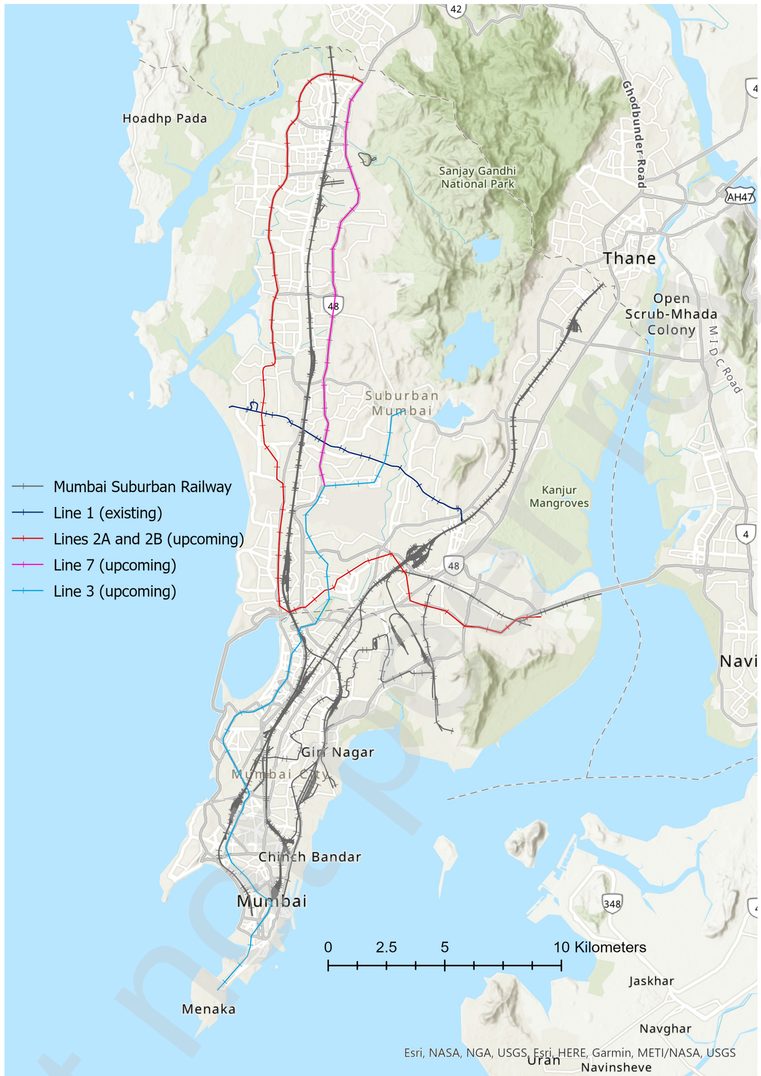
Varian, Hal R (1992), *Microeconomic Analysis*. WW Norton.

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.

Zheng, Siqi, Yangfei Xu, Xiaonan Zhang, and Rui Wang (2016), "Transit development, consumer amenities and home values: Evidence from beijing's subway neighborhoods." *Journal of Housing Economics*, 33, 22–33.

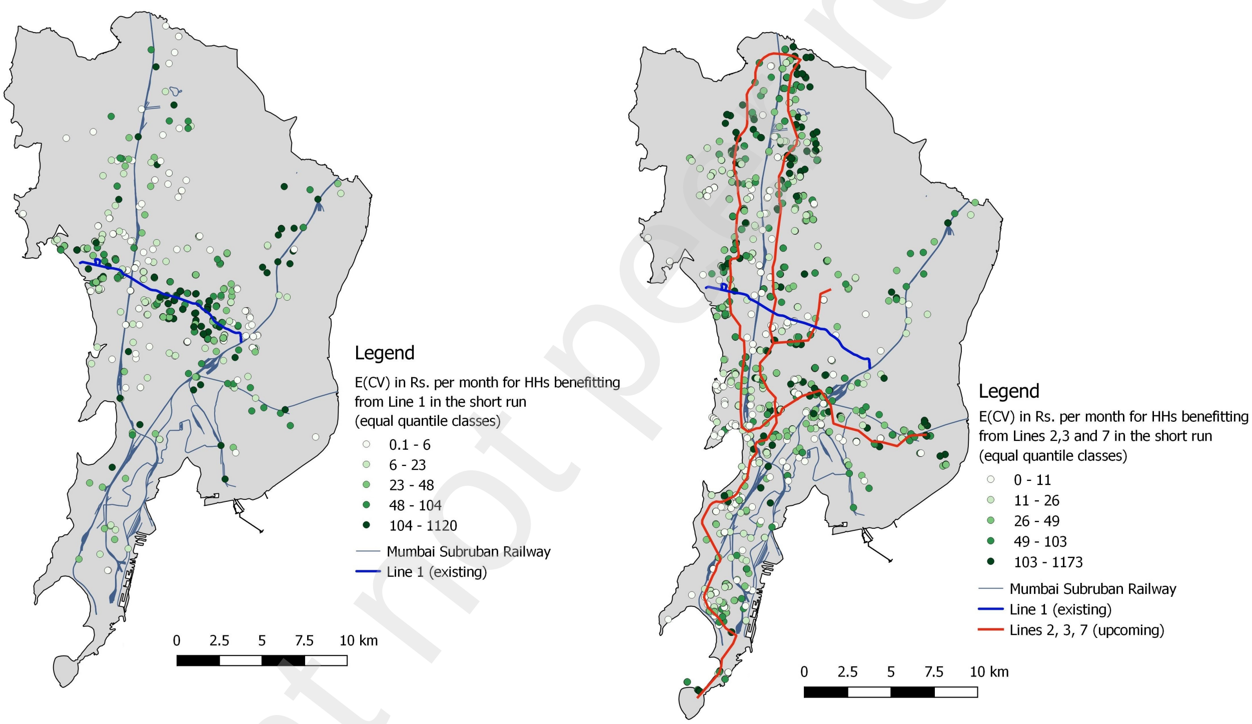
9 Figures and Tables

Figure 1: Existing and Planned Rail Transit Network in Mumbai



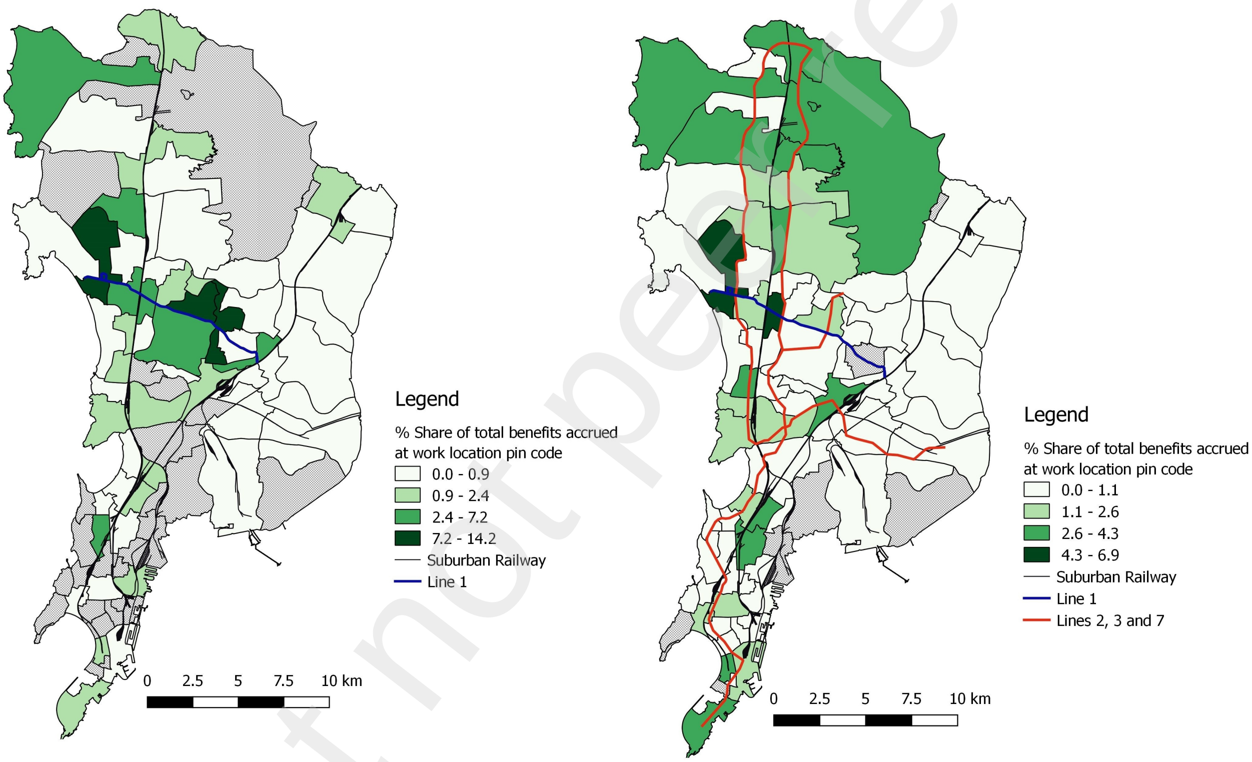
This map shows the existing Suburban railway network of Mumbai in grey, along with the existing Metro Line 1 in blue, and the three upcoming Metro Lines 2, 3 and 7 that are the focus of this paper in red, aqua and magenta, respectively.

Figure 2: Spatial Variation in the Value of Short Run 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)



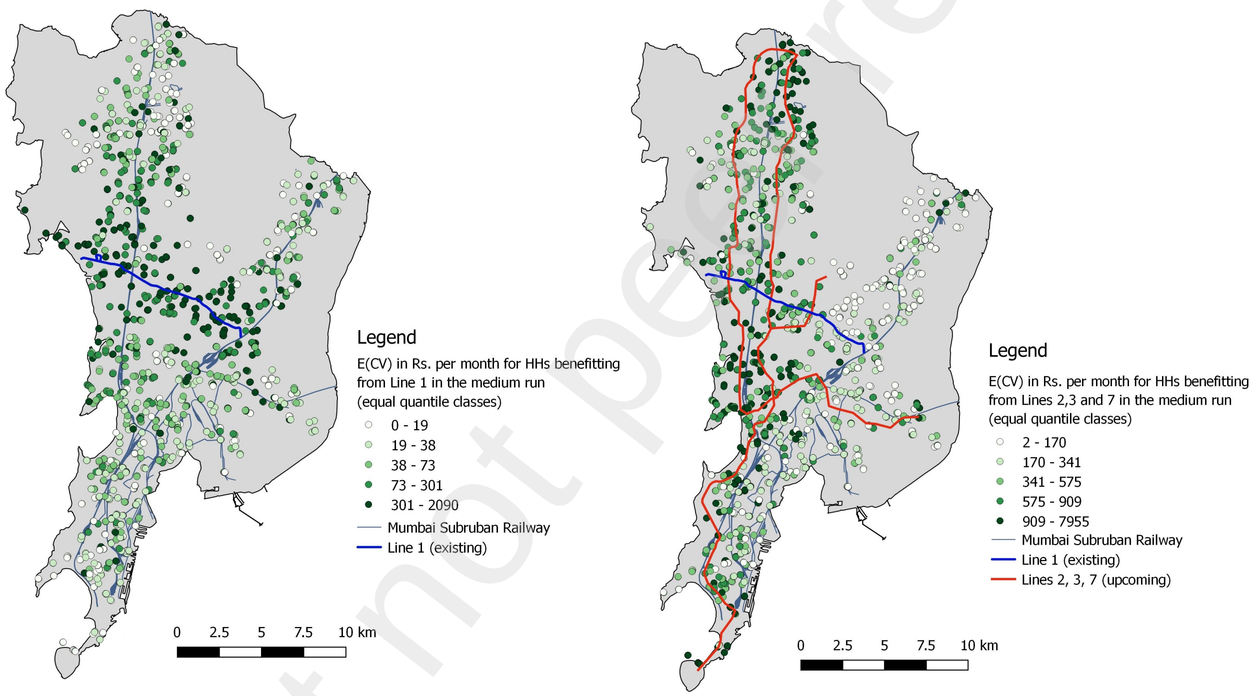
These maps show the sampled households with positive expected compensating variation computed using the formula in equation 5.

Figure 3: Spatial Variation in the Share of Short Run Benefits (in %) by Work Location Pincode– Line 1 (left) and Lines 2, 3 and 7 (right)



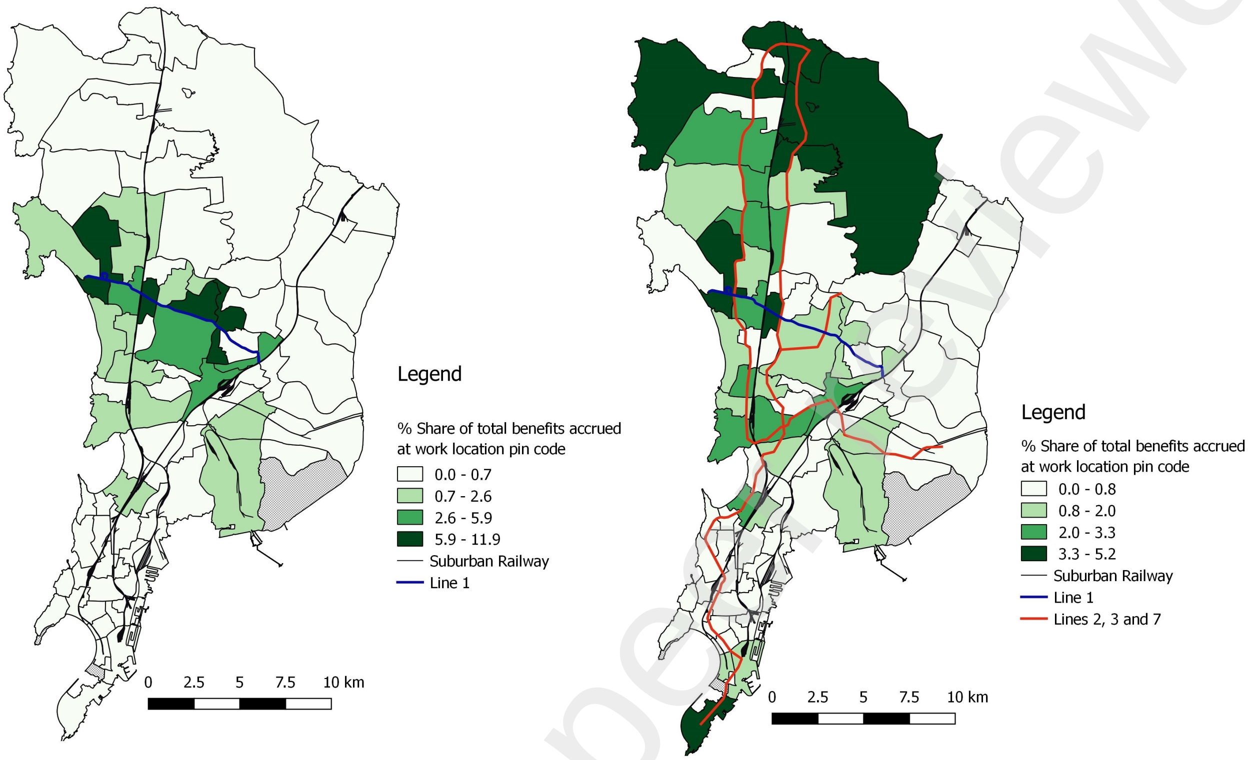
These maps show the share of total benefits accruing to different work location pin codes. Individual benefits were calculated using the formula in equation 5. White spaces indicate pin codes where no individual had positive benefits, including pin codes where no individual in the sample worked.

Figure 4: Spatial Variation in the Value of Long Run Benefits from the Housing 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 8.

Figure 5: Spatial Variation in the Share of Long Run Benefits (in %) by Primary Worker's Work Location Pincode– Line 1 (left) and Lines 2, 3 and 7 (right)



These maps show the share of total benefits accruing to different work location pin codes. Household benefits were calculated using the formula in equation 8. White spaces indicate pin codes where no primary worker of households in our sample worked.

Table 1: Main Mode Chosen for Work Commutes– 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	32.6	30.8	42.6	40.6	14.5	49.0	15.3	38.2	16.1
Train	16.0	15.7	17.8	14.2	20.0	24.1	7.5	16.8	13.7
Bus	8.4	7.7	12.3	9.3	6.3	14.1	2.4	10.5	2.5
Auto-rickshaw	9.1	8.1	15.3	9.3	8.7	12.8	5.3	10.7	4.6
Own two-wheeler	29.5	33.0	9.3	25.1	39.4		60.3	22.7	49.2
Car	4.5	4.8	2.8	1.5	11.1		9.1	1.2	13.9
Observations	2,876	2,444	432	1,992	884	1,472	1,404	2,144	732
Mean Distance (in km)	4.5	4.5	4.2	4.0	5.7	4.4	4.6	4.3	5.0

This Table shows the commute mode shares for different sub-groups of individuals in the commute mode choice estimation 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 2: Summary Statistics by Chosen Travel Mode for Commute Mode Choice Models' Estimation Sample

Chosen travel mode	Mean	Std. Dev.	Min	Max
Road distance from residence to work location in km.				
Walk	1.78	1.18	0.07	7.82
Train	10.36	7.69	0.23	38.27
Bus	4.79	4.06	0.16	31.20
Auto-rickshaw	3.67	4.65	0.07	30.36
Own two-wheeler	4.25	4.63	0.00	31.32
Car	6.08	5.73	0.12	36.10
Full Sample	**4.50**	**5.39**	**0.00**	**38.27**
In-vehicle time (IVT) in minutes				
Walk	0	0	0	0
Train	34.76	20.02	0.70	110
Bus	37.27	22.39	5.00	134
Auto-rickshaw	16.24	14.05	2.62	95.2
Own two-wheeler	17.95	13.13	2.50	83.35
Car	21.31	16.46	2.95	74.55
Full Sample	**16.42**	**19.15**	**0**	**134**
Out-of-vehicle time (OVT) in minutes				
Walk	21.31	14.14	0.88	93.85
Train	15.55	9.23	0.63	49.22
Bus	4.12	2.73	0.30	14.44
Auto-rickshaw	6.86	3.32	5	20
Own two-wheeler	0	0	0	0
Car	0	0	0	0
Full Sample	**10.4**	**12.82**	**0**	**93.85**
Cost of one-way trip (c) in Rs.				
Walk	0	0	0	0
Train	6	2	5	10
Bus	14	5	8	42
Auto-rickshaw	60	62	18	395
Own two-wheeler	14	13	2	85
Car	53.57	47.09	6.92	259.20
Full Sample	**14.13**	**28.88**	**0**	**394.7**
Average monthly income in Rs.				
Walk	18,630	10,475	2,500	75,000
Train	21,310	11,937	2,500	75,000
Bus	17,417	6,801	2,500	37,500
Auto-rickshaw	18,973	8,568	7,500	75,000
Own two-wheeler	28,288	16,032	7,500	125,000
Car	47,129	26,823	7,500	125,000
Full Sample	**23,101**	**14,878**	**2,500**	**125,000**
Cost per minute wage (c/w)				
Walk	0	0	0	0
Train	4.18	2.55	0.79	23.76
Bus	11.74	7.75	2.53	66.53
Auto-rickshaw	43.30	44.74	5.70	267.90
Own two-wheeler	7.35	7.80	0.38	67.57
Car	17.84	19.94	1.15	147.10
Full Sample	**8.58**	**19.14**	**0.00**	**267.90**

This Table presents summary statistics of variables used in the estimation of mode choice model in Section 4.1 for the estimation sample with 2,876 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 and out-of-vehicle times are described in Appendix Section A.2. 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 3: Summary Statistics for Housing Choice Models' Estimation Sample

Variables	Mean	Std. Dev.
Household characteristics		
Income in Rs.	30,939	19,207
Monthly rental price in Rs.	9,757	7,155
College-educated Primary Worker	0.30	
Vehicle Ownership	0.51	
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		
Distance to nearest railway station in km	1.50	1.17
Standardized employment accessibility index	0.01	0.98
Floorspace in sqft.	262.76	165.91
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	0.75	
Piped water	0.76	
Footpath in the neighborhood	0.75	
Slum Classification	0.44	
Distance to coast (in km)	4.67	3.17
Reports of crimes against women (Median)	38.5	18.28
Walk time to the nearest pvt. doctor	8.05	6.00
Walk time to the nearest govt. hospital	19.96	8.86
Walk time to the nearest pvt. hospital	17.30	8.86

This Table presents summary statistics of variables used in the estimation of housing choice model in Section 4.2. 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 30,939 = $1,454 (PPP); mean monthly rent Rs. 9,757 = $458.45 (PPP).

Table 4: Preference Parameters from Nested Logit Models of Commute Mode Choice for Different Nesting Structures

	Model 1	Model 2	Model 3
Income-Cost	0.025***	0.023***	0.024***
	(0.002)	(0.001)	(0.001)
IVT	-0.019***	-0.019***	-0.014***
	(0.003)	(0.003)	(0.003)
OVT	-0.035***	-0.034***	-0.035***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Walk, Auto-rickshaw)	Omitted	Omitted	Omitted
(Two-wheeler)	0.734***		
	(0.079)		
(Car)	1.190***		
	(0.147)		
(Train, Bus)	-1.289***	-1.270***	
	(0.095)	(0.093)	
(Car, Two-wheeler)		0.796***	0.710***
		(0.075)	(0.075)
(Train)			-1.296***
			(0.084)
(Bus)			-1.991***
			(0.109)
Dissimilarity Parameters:			
(Two-wheeler)	1		
(Car)	1		
(Train, Bus)	0.687***	0.681***	
	(0.104)	(0.103)	
(Walk, Auto-rickshaw)	0.626***	0.596***	0.555***
	(0.063)	(0.058)	(0.051)
(Car, Two-wheeler)		1	1
		(constrained)	(constrained)
(Train)			1
(Bus)			1
Individuals	2876.000	2876.000	2876.000
LR chi2	314.955	318.729	358.420
Log likelihood	-2982.274	-2987.271	-2955.539
IVT value (Rs. per minute)	0.772	0.824	0.575
OVT value (Rs. per minute)	1.412	1.446	1.460
Value of IVT (% wage)	39.7	42.4	29.6
Value of OVT (% wage)	72.6	74.4	75.1

This Table presents estimated preference parameters for the nested logit model in equation 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: (Walk, Auto-rickshaw), (Car), (Two-wheeler), (Train, Bus); Model 2: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train, Bus); Model 3: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train), (Bus). Dissimilarity parameter is constrained to be $\leq$1 so that predictions are consistent with equation 3. 'Walk' also includes 'bicycle'.

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

Table 5: Predicted Mode Shares of Nested Logit Models from Table 4

Travel modes	True shares	Model 1	Model 2	Model 3
Walk	32.55	29.08	28.99	28.39
Train	15.99	12.54	12.52	15.99
Bus	8.41	11.87	11.89	8.41
Auto-rickshaw	9.14	12.61	12.70	13.30
Own two-wheeler	29.45	29.45	30.21	30.23
Car	4.45	4.45	3.69	3.67

This Table compares the predicted mode shares under the three nested logit models in Table 4 with the true sample shares. Nesting structure in Model 1: (Walk, Auto-rickshaw), (Car), (Two-wheeler), (Train, Bus); Model 2: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train, Bus); Model 3: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train), (Bus).

Table 6: Preference Parameters from the First Stage of the Housing Choice Model

	Model 1 (1)	Model 2 (2)
Expected Commuting Utility	2.362***	2.409***
	(0.043)	(0.044)
Proportion of HHs with same language	2.182***	2.186***
	(0.250)	(0.250)
Proportion of HHs with same religion	2.668***	2.666***
	(0.393)	(0.393)
Households	2,170	2,170
Wald chi2	3556	3527
Log likelihood	-6910	-6918

This Table presents estimated preference parameters for the first-stage (equation 10) of the housing location choice model in Section 4.2. Estimates of $\hat{\delta}_h$ are not shown. Standard errors are in parentheses. Expected commuting utility is computed using equation 7. Proportion of HHs with the same language and religion as the chooser are defined within a 2 km neighborhood around the house. Columns 1 and 2 have expected commuting utility estimated using Models 1 and 2 in Table 4, respectively. Nesting structure in Model 1: (Walk, Auto-rickshaw), (Car), (Two-wheeler), (Train, Bus); Model 2: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train, Bus).

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

Table 7: Preferences for Commuting Utility from the Housing Choice Models with Taste Shifters

Taste shifter	(1) None	(2) Educ	(3) Income	(4) Vehicle ownership
Model 1:				
Base category	2.362***	2.542***	2.415***	2.207***
	(0.043)	(0.052)	(0.055)	(0.049)
$\geq$ college educ		1.996***		
		(0.065)		
Median HH income			2.385***	
			(0.068)	
$>$ Median HH income			2.038***	
			(0.109)	
HH owns vehicle				2.649***
				(0.069)
Households	2170	2170	2170	2170
Log Likelihood	-6910	-6883	-6901	-6891
Model 2:				
Base category	2.409***	2.587***	2.452***	2.235***
	(0.044)	(0.053)	(0.056)	(0.050)
$\geq$ college educ		2.045***		
		(0.066)		
Median HH income			2.438***	
			(0.070)	
$>$ Median HH income			2.108***	
			(0.113)	
HH owns vehicle				2.735***
				(0.071)
Households	2170	2170	2170	2170
Log Likelihood	-6918	-6893	-6910	-6896

This Table shows preferences for travel duration estimated in the first-stage of the housing choice model using specifications with taste-shifters. 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). These models also account for heterogeneous preferences for the proportion of HHs with the same language and religion within a 2 km radius of the feasible house.

Table 8: Mean Preferences for Housing Amenities from Second Stage Regressions

	(1)	(2)	(3)	(4)	(5)
Rental Price	-0.000169***	-0.000157***	-0.000123***	-0.000122***	-0.000137***
	(0.00004)	(0.00004)	(0.00003)	(0.00003)	(0.00003)
	[-4.330]	[-3.949]	[-3.693]	[-4.075]	[-4.304]
Housing Amenity Index	0.456600***	0.441438***	0.375659***	0.381520***	0.412724***
	(0.08352)	(0.08199)	(0.07028)	(0.06494)	(0.07055)
	[5.467]	[5.384]	[5.345]	[5.875]	[5.850]
Distance to coast (in km)		0.032555**	0.009907	-0.051829***	-0.040053***
		(0.01626)	(0.01242)	(0.01406)	(0.01407)
		[2.003]	[0.798]	[-3.686]	[-2.846]
Slum Classification Dummy		-0.177700**	-0.113486	-0.129261**	-0.121339*
		(0.09039)	(0.06997)	(0.06347)	(0.06699)
		[-1.966]	[-1.622]	[-2.037]	[-1.811]
No. of Reported Crimes Against Women		-0.004437	-0.008021***	-0.009455***	-0.005905**
		(0.00381)	(0.00283)	(0.00258)	(0.00289)
		[-1.166]	[-2.830]	[-3.669]	[-2.045]
Distance to the nearest railway station (in km)			0.342368***	0.151881***	0.119253**
			(0.03663)	(0.04914)	(0.05425)
			[9.346]	[3.091]	[2.198]
Standardized Employment Accessibility Index				0.435875***	0.421044***
				(0.05762)	(0.05935)
				[7.564]	[7.094]
Proximity to Health Services Index					0.007399
					(0.03130)
					[0.236]
F(excluded IV)	40.44	41.01	38.82	39.33	39.83
Observations	2,170	2,170	2,170	2,170	1,989
Critical Value for |t| at 95% level (Lee et al. (2022))	2.247	2.247	2.247	2.247	2.247

This Table presents 2SLS parameter estimates of the second stage of the housing choice model. Dependent variable is the vector of estimated intercepts from the first stage conditional logit model presented in Table 6). Log of assessed property 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. (2022) for valid inference at the 95% level. Robust std. errors clustered at the sub-zone level are in parentheses. t-statistics are in brackets. Star marks reflect conventional inference values.

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

Table 9: The Value of Time Savings from the Commute Mode Choice Model

	Full sample	Men	Women	< College education	≥ College education	≤median income	> median income
Nested logit Model 1:							
Individuals	2876	2444	432	1992	884	2144	732
IVT Value (Rs./min)	0.77	0.83	0.59	0.66	1.15	0.59	1.60
IVT Value (% of wage)	39.72	40.46	45.14	40.53	43.26	44.17	43.37
OVT Value (Rs./min)	1.41	1.39	1.55	1.31	1.83	1.28	1.75
OVT Value (% of wage)	72.62	67.68	117.89	80.68	68.69	94.99	47.24
% sample with positive E(CV) Line 1	24.58	24.67	24.07	22.09	30.43	24.35	25.27
Mean E(CV) Line 1|Positive E(CV) (Rs./month)	77.27	74.56	98.30	66.17	110.47	70.84	84.07
% sample with positive E(CV) Lines 2,3,7	56.71	57.12	54.40	55.47	60.18	56.62	56.83
Mean E(CV) Lines 2,3,7|Positive E(CV) (Rs./month)	97.63	98.32	109.24	88.43	130.10	96.47	90.47
Nested logit Model 2:							
Individuals	2876	2444	432	1992	884	2144	732
IVT Value (Rs./min)	0.82	0.88	0.66	0.64	1.42	0.56	2.00
IVT Value (% of wage)	42.37	42.61	50.29	39.15	53.24	41.97	54.12
OVT Value (Rs./min)	1.45	1.42	1.59	1.30	2.02	1.25	1.92
OVT Value (% of wage)	74.36	69.09	121.32	79.74	75.75	92.72	51.88
% sample with positive E(CV) Line 1	24.58	24.67	24.07	22.09	30.54	24.35	25.41
Mean E(CV) Line 1|Positive E(CV) (Rs./month)	80.35	77.09	102.57	65.00	127.76	68.69	97.00
% sample with positive E(CV) Lines 2,3,7	56.71	57.12	54.40	55.47	60.41	56.62	56.83
Mean E(CV) Lines 2,3,7|Positive E(CV) (Rs./month)	101.32	101.39	115.65	86.88	149.63	93.76	106.78

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) estimated separately for the subsamples indicated in the columns. Preference parameters for the full sample are in Table 4. Individuals indicate the number of individuals in each of these estimation samples. Nesting structure in Model 1: (Walk, Auto-rickshaw), (Car), (Two-wheeler), (Train, Bus); Model 2: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train, Bus).

Table 10: The Value of Time Savings from the Housing Choice Model

	(1)	(2)	(3)	(4)	(5)
Mean E(CV) Line 1 in Rs. per month:					
Model 1 (in Rs. per month)	134.08	144.68	185.15	186.68	165.93
Model 1 (% Monthly Rent)	1.37	1.48	1.90	1.91	1.70
Model 2 (in Rs. per month)	135.03	146.32	187.38	188.94	167.90
Model 2 (% Monthly Rent)	1.38	1.50	1.92	1.94	1.72
Mean E(CV) Lines 2, 3 and 7 in Rs. per month:					
Model 1 (in Rs. per month)	384.45	414.83	530.89	535.26	475.78
Model 1 (% Monthly Rent)	3.94	4.25	5.44	5.49	4.88
Model 2 (in Rs. per month)	383.32	415.35	531.92	536.33	476.61
Model 2 (% Monthly Rent)	3.93	4.26	5.45	5.50	4.88
Controls:					
Housing amenities index	✓	✓	✓	✓	✓
Distance to coast in km	✗	✓	✓	✓	✓
Slum classification dummy	✗	✓	✓	✓	✓
Crimes against women (Reports)	✗	✓	✓	✓	✓
Distance to nearest station	✗	✗	✓	✓	✓
Employment accessibility index	✗	✗	✗	✓	✓
Proximity to Health Services	✗	✗	✗	✗	✓

This Table presents the mean expected compensating variation for Line 1 and Lines 2, 3 and 7 (equation 8) computed using the first stage estimates in Table 6 and various second-stage specifications (Table 8), the controls for which are indicated. The average monthly rent is Rs. 9757.147.

Table 11: Heterogeneity in the Value of Time Savings from the Housing Choice Model

	Line 1		Lines 2, 3, 7	
	in Rs.	% of Monthly Rent	in Rs.	% of Monthly Rent
Full sample	186.68	1.91	535.26	5.49
Primary worker has < college education	189.25	2.38	580.29	7.29
Primary worker has ≥ college education	180.57	1.29	428.54	3.06
Below median income HH	211.50	3.15	640.35	9.53
Median income HH	181.25	1.57	474.69	4.11
Above median income HH	81.33	0.43	214.72	1.14
HH does not own vehicle	300.21	4.10	833.97	11.39
HH Owns vehicle	75.42	0.62	242.54	2.00

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 housing choice model without taste-shifters (Model 1 in Table 6) and using the preferred second-stage specification (Column (4) of Table 8) for the full sample of households and averages are calculated for each sub-group.

Table 12: Aggregate Benefits of Metro lines vs. the Equivalent Annualized Capital Costs (EACC) in $ Million (PPP)

	EACC				Short Run	Long Run
	(1)	(2)	(3)	(4)	Benefits	Benefits
Line 1	300	200	200	124	51.2	565.2
Lines 2, 3 and 7	3,000	2,300	1,900	1,344	169.6	1,563.4
EACC assumptions:						
Life of asset (years)	20	30	35	20		
Interest rate (in %)	12	10	8	2		

This Table presents the equivalent annualized capital costs under three assumptions indicated in the second panel; and the annual aggregate short run and long run benefits of Metro Line 1 and Lines 2, 3 and 7 computed by scaling the individual-level estimates for Model 1. 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

A Appendix: Data

A.1 Household Survey Description

The individual and household 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 January-March 2019. Two members were interviewed in each household, an adult male and female (ages 18-45) with priority given to primary earners and decision makers of the household. The survey contains information on household members' education, occupation, income, household demographic composition, housing condition, household assets, pin codes of work locations and commute trips from residence to workplace. A travel diary was also 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 out-of-pocket cost. These data are described in Alam et al. (2021).

Definition of main mode: For the commute mode choice model in the paper, household location is treated as the origin, and a randomly selected post office that has the same pin code as the individual's workplace as the destination. The number of post offices per pin code in Mumbai ranges from 1 to 9, with the median being 4. For workers who commute, the survey records up to three modes of transportation used and the time spent in each mode for a one-way trip. The chosen travel mode in the commute mode choice model is the main mode, defined as follows. When a mix of motorized and non-motorized transportation is used, the main mode is defined as the motorized mode on which the most time is spent. 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. The main mode is non-motorized (walk or bicycle) when 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 4.2: For each house, a 2 km neighborhood was defined using Euclidean distance. The median house has 117 neighbors. The proportion of neighbors with the same language and religion as household i in the housing choice model is calculated by matching household i's religion and language with that of the households in the neighborhood of each house in household i's choice set.

A.2 In-vehicle and Out-of-vehicle Time Variables

We 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.[34] 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. 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. To compute walk times, we assume a speed of 5 kmph along the road network. For travel times by train, we assume a speed of 40 kmph for the Mumbai Suburban Railway network segments, 35 kmph for Metro rail network segments, and 5 kmph (walking speed) for the road segments connecting gaps in the train network.

In addition to computing travel times under the rail network that includes Suburban Railway and Metro Line 1, we also calculate travel times (a) with Lines 2, 3 and 7 added to the existing network and (b) without any Metro line, for short run welfare calculations.

Second, we 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.[35] We match commute trips in our sample to a randomly chosen origin-destination pair from the set of trips in 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.[36] 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 us to accurately model the tradeoff between rail and road transport, which is critical because of the traffic problems in Mumbai. For residence-work pairs for which Google Maps data is missing, we use the network program with road network to compute drive times at 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.

[34] The program is implemented in Python using packages GOSTnets and NetworkX. Link for GOSTnets: https://github.com/worldbank/GOST_PublicGoods

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

[36] 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.

Third, we 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 us 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 us 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.[37] 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.[38]

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

A.3 Employment Accessibility Index

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

$$\text{EA}_h = \sum_j \left(\frac{w_j}{d_{hj}}\right) \tag{13}$$

w_j is the effective 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}. We use the

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

[38] Since the exact work location is not known and the destination of a 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.

methodology in Kreindler and Miyauchi (2023) to obtain a proxy for w_j and estimate κ. The underlying model that allows identification of these parameters is one where commuters choose an origin and destination for commutes based on the characteristics of each location and commuting costs.

The utility that a worker living at location h receives from working at employment location j is given by

$$U_{hj}(\omega) = \frac{w_j * \epsilon_{hj}(\omega)}{d_{hj}} \tag{14}$$

w_j is the effective wage obtainable at j and each worker gets the same wage. $d_{hj} = exp(\kappa * t_{hj})$ is the iceberg commuting cost between h and j represented by an exponential function of commuting time t_{hj} times the semi-elasticity of commuting costs to time κ. $\epsilon_{hj}(\omega)$ is an idiosyncratic utility shock assumed to follow an i.i.d. Fréchet distribution with shape parameter θ and scale parameter normalized to one.[39] A higher value of θ implies lower dispersion in random shocks across individuals that lead to the observed pattern of commute flows. That is, the higher the θ, the more likely that the pattern of commute flows came about as a result of individuals responding to the spatial distribution of wages, amenities, and commuting costs.

Equation 14 implies that the probability of a worker working in j conditional on living in h is given by

$$\pi_{hj|h} = \frac{(w_j/d_{hj})^\theta}{\sum_j (w_j/d_{hj})^\theta} \tag{15}$$

Equation 15 implies the following gravity equation of commute flow shares.

$$\log\pi_{hj|h} = -\kappa * \theta * t_{hj} + \theta * \log w_j - \log\big(\sum_j (w_j/exp(\kappa * t_{hj}))^\theta\big) \tag{16}$$

We estimate the following reduced-form gravity equation of commuter flows derived from 16 using a Poisson pseudo-maximum likelihood estimator.

$$N_{hj} = -\beta * t_{hj} + \psi_j + \gamma_h + \nu_{hj} \tag{17}$$

[39] The random shock encompasses many different unaccounted for reasons that could be behind the observed spatial distribution, for example, proximity to family members or a cultural center. Kreindler and Miyauchi (2023) shows that their model is robust to alternate assumptions, for example, in (Tsivanidis (2019)), θ represents the inverse of dispersion in worker productivity across locations.

N_{hj} represent aggregate commute flows between h and j.[40] β captures the sensitivity of commuting decisions to commuting time. γ_h 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_{hj} is the random error.

To calculate the employment accessibility index in equation 13, we first estimate equation 17 using data on commute flows between residence and work location pincodes from each household survey. 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. Estimates of work location fixed effects, $\hat{\psi}_j$ are assumed to proxy the effective wage at each location, w_j.[41]

The parameter measuring 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). We find β=0.138 from estimating the gravity equation.

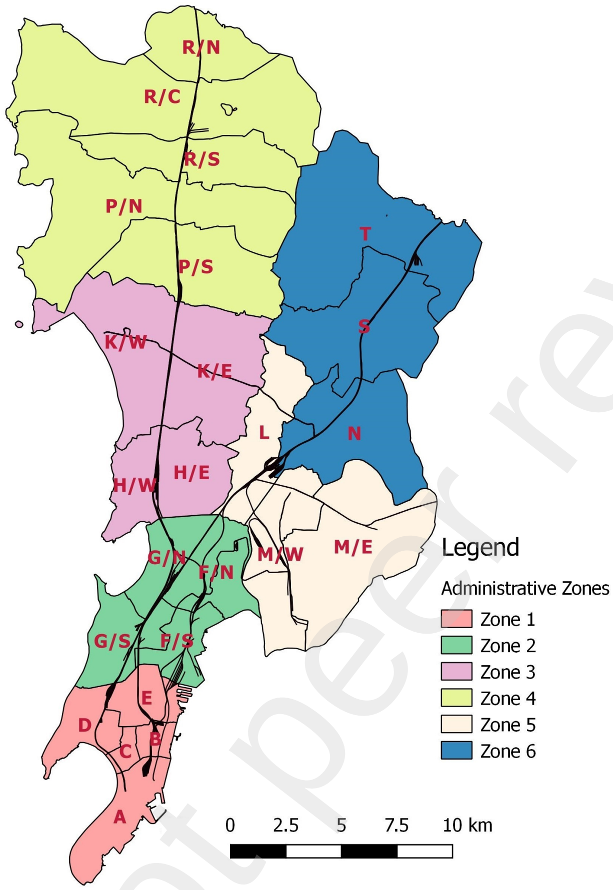
Following Kreindler and Miyauchi (2023), we obtain $\hat{\theta}$ by inverting the coefficient from an OLS regression of log of average incomes aggregated at the pin code level on $\hat{\psi}_j$. We then obtain $\hat{\kappa} = \frac{\hat{\beta}}{\hat{\theta}}$. Intuitively, $\hat{\psi}_j$ are model predicted wages and they deviate from actual wages in proportion to the variation in idiosyncratic shocks. We find $\hat{\theta}$=12.85 and therefore, $\hat{\kappa} = 0.0107$. Ahlfeldt et al. (2015) estimates $\kappa = 0.01$ and Tsivanidis (2019) estimates $\kappa = 0.012$. Note that $\hat{\psi}_j$ does not have a fixed scale, so we standardize EA_h to be mean 0 with variance 1 for the second stage of the housing choice model.

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

[41] The correlation between $\hat{\psi}_j$ and average income from the 2019 survey data at the level of work location pincode is 0.24.

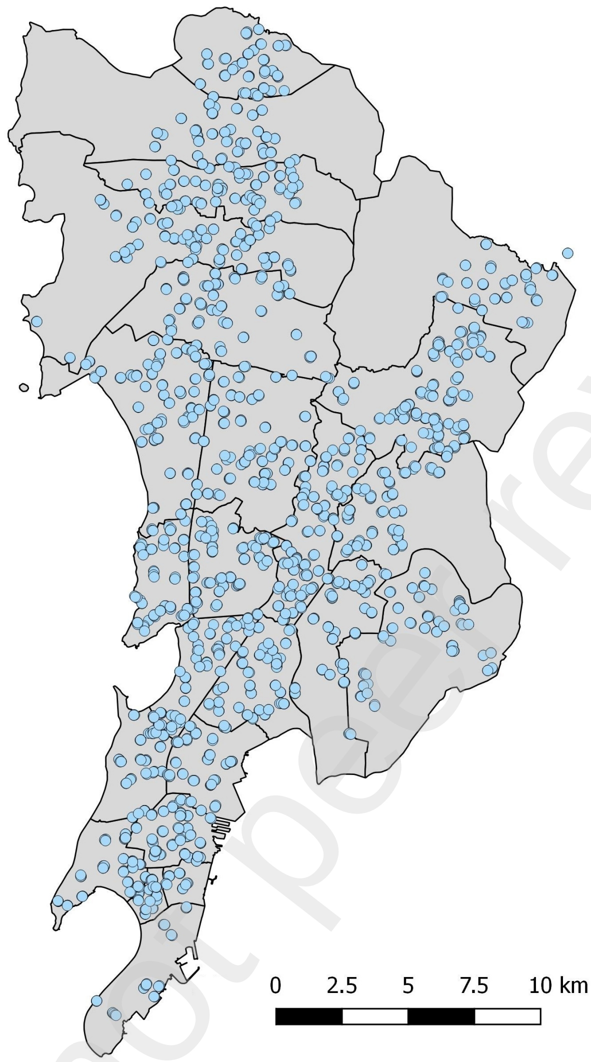
B Appendix: Figures and Tables

Figure B1: 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 B2: 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 B1: Sensitivity of the Commute Mode Choice model parameters to different definitions of IVT and OVT for Model 1 in Table 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***	0.025***	0.026***
	(0.002)	(0.002)	(0.002)
IVT	-0.019***	-0.021***	-0.015***
	(0.003)	(0.003)	(0.003)
OVT	-0.035***	-0.033***	-0.032***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Walk, Auto-rickshaw)	Omitted	Omitted	Omitted
(Two-wheeler)	0.734***	0.753***	0.748***
	(0.079)	(0.082)	(0.082)
(Car)	1.190***	1.202***	1.204***
	(0.147)	(0.152)	(0.154)
(Train, Bus)	-1.289***	-1.091***	-1.032***
	(0.095)	(0.078)	(0.077)
Dissimilarity Parameters:			
(Two-wheeler)	1	1	1
(Car)	1	1	1
(Train, Bus)	0.687***	0.545***	0.425***
	(0.104)	(0.066)	(0.054)
(Walk, Auto-rickshaw)	0.626***	0.665***	0.658***
	(0.063)	(0.073)	(0.075)
Individuals	2876.000	2854.000	2824.000
LR chi2	314.955	314.096	313.294
Log likelihood	-2982.274	-2943.460	-2904.154
IVT value (Rs. per minute)	0.772	0.823	0.592
OVT value (Rs. per minute)	1.412	1.314	1.233
Value of IVT (% wage)	39.7	42.5	30.6
Value of OVT (% wage)	72.6	67.8	63.8

This Table presents estimated preference parameters for the nested logit model in equation 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. 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. In Column (3), out-of-vehicle time is from the survey. Std. errors are in parentheses. Estimated parameters are based on the nesting structure in Model 1 of Table 4: (Walk, Auto-rickshaw) , (Car), (Two-wheeler), (Train, Bus). Estimated parameters based on Model 2 are in Table B2. Dissimilarity parameter is constrained to be ≤1 so that predictions are consistent with equation 3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B2: Sensitivity of the Commute Mode Choice Model parameters to different definitions of IVT and OVT for Model 2 in Table 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	0.023^{***}	0.024^{***}	0.024^{***}
	(0.001)	(0.001)	(0.002)
IVT	-0.019^{***}	-0.021^{***}	-0.015^{***}
	(0.003)	(0.003)	(0.003)
OVT	-0.034^{***}	-0.032^{***}	-0.031^{***}
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Walk, Auto-rickshaw)	Omitted	Omitted	Omitted
(Car, Two-wheeler)	0.796^{***}	0.808^{***}	0.802^{***}
	(0.075)	(0.078)	(0.077)
(Train, Bus)	-1.270^{***}	-1.083^{***}	-1.023^{***}
	(0.093)	(0.076)	(0.075)
Dissimilarity Parameters:			
(Car, Two-wheeler)	1	1	1
	(constrained)	(constrained)	(constrained)
(Train, Bus)	0.681^{***}	0.535^{***}	0.420^{***}
	(0.103)	(0.066)	(0.054)
(Walk, Auto-rickshaw)	0.596^{***}	0.628^{***}	0.617^{***}
	(0.058)	(0.065)	(0.066)
Individuals	2876.000	2854.000	2824.000
LR chi2	318.729	317.518	317.055
Log likelihood	-2987.271	-2948.212	-2908.934
IVT value (Rs. per minute)	0.824	0.865	0.634
OVT value (Rs. per minute)	1.446	1.345	1.268
Value of IVT (% wage)	42.4	44.7	32.8
Value of OVT (% wage)	74.4	69.4	65.6

This Table presents estimated preference parameters for the nested logit model in equation 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. 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. In Column (3), out-of-vehicle time is from the survey. Std. errors are in parentheses. Estimated parameters are based on the nesting structure in Model 2 of Table 4: (Car, Two-wheeler), (Walk, Auto-rickshaw), (Train, Bus). Estimated parameters based on Model 1 are in Appendix Table B1. Dissimilarity parameter is constrained to be ≤ 1 so that predictions are consistent with equation 3.

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

Table B3: Preference Parameters from Nested Logit Models of Commute Mode Choice with income entering non-linearly

	(1) Model 1	(2) Model 2	(3) Model 3
Cost/Wage	-0.042***	-0.041***	-0.042***
	(0.002)	(0.002)	(0.003)
IVT	-0.025***	-0.025***	-0.018***
	(0.004)	(0.004)	(0.004)
OVT	-0.038***	-0.038***	-0.038***
	(0.002)	(0.002)	(0.002)
Intercepts:			
(Walk, Auto-rickshaw)	Omitted	Omitted	Omitted
(Two-wheeler)	0.989***		
	(0.077)		
(Car)	1.244***		
	(0.140)		
(Train, Bus)	-1.120***	-1.108***	
	(0.108)	(0.108)	
(Car, Two-wheeler)		1.028***	0.983***
		(0.074)	(0.094)
(Train)			-0.989***
			(0.090)
(Bus)			-1.620***
			(0.118)
Dissimilarity Parameters:			
(Two-wheeler)	1		
(Car)	1		
(Train, Bus)	0.909***	0.910***	
	(0.135)	(0.135)	
(Walk, Auto-rickshaw)	1	1	1.044***
	(constrained)	(constrained)	(0.113)
(Car, Two-wheeler)		1	1
		(constrained)	(constrained)
(Train)			1
(Bus)			1
Individuals	2876.000	2876.000	2876.000
LR chi2	492.290	494.239	366.308
Log likelihood	-2995.816	-2997.480	-2969.753
IVT value (% wage)	58.4	60.2	42.3
OVT value (% wage)	89.7	90.7	90.6

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: (Walk, Auto-rickshaw), (Car), (Two-wheeler), (Train, Bus); Model 2: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train, Bus); Model 3: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train), (Bus). Dissimilarity parameter has been constrained to 1 so that predictions are consistent with equation 3. Walk also includes bicycle. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B4: Predicted mode shares of Nested Logit Models from Tables 4 (income entering linearly), B3 (income entering non-linearly), and a Mixed Logit Model (correlated random coefficients)

Travel modes	True shares	Income entering linearly Model 1	Income entering linearly Model 2	Income entering linearly Model 3	Income entering non-linearly Model 1	Income entering non-linearly Model 2	Income entering non-linearly Model 3	Mixed Logit
Walk	32.55	29.08	28.99	28.39	28.13	28.04	27.17	28.00
Train	15.99	12.54	12.52	15.99	12.94	12.94	15.99	20.71
Bus	8.41	11.87	11.89	8.41	11.46	11.47	8.41	16.47
Auto-rickshaw	9.14	12.61	12.70	13.30	13.56	13.65	14.52	14.66
Own two-wheeler	29.45	29.45	30.21	30.23	29.45	29.90	29.92	17.91
Car	4.45	4.45	3.69	3.67	4.45	4.00	3.98	2.24

This Table compares sample commute mode shares with predicted mode shares from nested logit models in Table 4 where income-cost enters linearly, models in Table B3 where income enters non-linearly to compare model fit, and those predicted under a mixed logit specification. Nesting structure in Model 1: (Walk, Auto-rickshaw), (Car), (Two-wheeler), (Train, Bus); Model 2: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train, Bus); Model 3: (Walk, Auto-rickshaw), (Car, Two-wheeler), (Train), (Bus). 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 $\alpha_1^i, \alpha_2^i, \alpha_3^i \sim N(\alpha, \mathbf{\Sigma})$.

Table B5: Factor Loading for Variables in Housing Amenities Index

Variable	Factor Loading
Good roof	0.2488
Floorspace (in sqft.)	0.3809
Number of rooms	0.3744
Separate Kitchen	0.4499
Toilet inside the house	0.4333
Bathroom inside the house	0.3720
Piped water	0.3526

This Table presents factor loadings for the Housing amenities index variable used in the second stage of the housing choice model (Table 8). These variables are summarized in Table 6.

Table B6: 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 housing choice model (Column (5), Table 6). Each of these variables have four categories increasing in order of proximity.

Table B7: First Stage for 2SLS regression in Table 8

	(1)	(2)	(3)	(4)	(5)
Log(Annual assessed sale value)	3637.496***	3594.764***	3593.144***	3601.963***	3154.939***
	(572.026)	(561.310)	(576.687)	(574.384)	(499.932)
Housing Amenity Index	2232.026***	2225.880***	2225.741***	2245.628***	2224.245***
	(141.948)	(147.204)	(147.750)	(148.090)	(147.637)
Distance to coast (in km)		42.429	42.707	-109.139	-55.038
		(58.631)	(61.772)	(74.866)	(73.166)
Slum Classification Dummy		-961.339***	-961.750***	-1002.938***	-854.924**
		(337.903)	(336.040)	(329.238)	(340.533)
No. of Reported Crimes Against Women		-1.777	-1.729	-5.263	10.162
		(13.360)	(13.688)	(13.698)	(14.189)
Distance to the nearest railway station (in km)			-4.499	-473.355*	-480.123*
			(202.728)	(264.479)	(277.323)
Standardized Employment Accessibility Index				1072.819***	831.095***
				(306.753)	(303.340)
Proximity to Health Services Index					401.972***
					(135.074)
R-squared	0.373	0.378	0.378	0.387	0.403
Observations	2,170	2,170	2,170	2,170	1,989

This Table presents the first-stage estimates of the 2SLS specifications in Table 8. These parameters are in equation 12. 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 B8: Sensitivity of Long Run Benefits to First Stage Specification

Taste shifter	(1) None	(2) Educ	(3) Income	(4) Vehicle ownership
Mean E(CV) Line 1:				
Model 1 (in Rs. per month)	186.68	187.78	186.85	174.30
Model 1 (% Monthly Rent)	1.91	1.92	1.92	1.79
Model 2 (in Rs. per month)	188.94	194.91	188.72	178.84
Model 2 (% Monthly Rent)	1.94	2.00	1.93	1.83
Mean E(CV) Lines 2, 3 and 7:				
Model 1 (in Rs. per month)	535.26	553.23	526.92	508.50
Model 1 (% Monthly Rent)	5.49	5.67	5.40	5.21
Model 2 (in Rs. per month)	536.33	556.72	537.19	508.96
Model 2 (% Monthly Rent)	5.50	5.71	5.51	5.22

This Table shows expected compensating variation computed using specifications with taste-shifters in the first stage shown in Table 11. Column (1) is the base model without any taste-shifters. Taste-shifters in Columns 2, 3 and 4 are education, income, and vehicle ownership, respectively (corresponding to Table 7). The corresponding second stage specification for each model has the same controls as in Column (4), Table 8. The average monthly rent is Rs. 9757.147.