Inclusive Infrastructure: People versus Place *

Clare Balboni †
London School of Economics

Gharad Bryan ‡
London School of Economics

Melanie Morten §
Stanford University and NBER

Caylee O'Connor ¶
Stanford University

Bilal Siddiqi ||
CEGA

March 20, 2025

Preliminary version: please check with authors for most recent draft before circulating

Abstract

Roads, rail, and other public transport in a city are "place-based," in that they are built in specific neighborhoods. Do such investments benefit individuals living in targeted locations? If people are mobile within a city, then any such place-based investment can lead to neighborhood changes, such as rent increases, which change who can afford to live near these investments and hence who benefits from them. Using panel data tracked on locations and individuals, we provide a decomposition to separate the causal effects of transportation from the compositional effects of population resorting. We illustrate the method using data from the Bus Rapid Transit (BRT) system expansion in Dar es Salaam, Tanzania.

Keywords: Transport, Urbanization, Gentrification, Tanzania, House Prices
JEL Classification:

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. We thank Jessica Mahoney, Juliana Aguilar, Justus Mugango, Rachel Steinacher, and Rachel Jones at IPA Tanzania for coordinating our data collection efforts in Tanzania. We also thank EDI for assistance with initial fieldwork. We thank Yonas Mchomvu at the World Bank and Ronald Lwakatare at the Dar es Salaam Rapid Transit Agency for institutional support. We also thank Nancy Lozano Garcia for sharing the World Bank LSMS data for Dar es Salaam. The research has been funded with UK aid from the UK government through the DIME ieConnect for Impact program at the World Bank. Additional funding is gratefully acknowledged from IGC, 3ie, and Stanford. Devika Lakhote, Helen Gu, and Betsenat Gebrewold provided excellent research assistance. Any errors are our own.

†Email: c.a.balboni@lse.ac.uk
‡Email: g.t.bryan@lse.ac.uk
§Email: memorten@stanford.edu
¶Email: cayleeoc@stanford.edu
||Email: bilal.siddiqi@berkeley.edu

1 Introduction

The world's urban areas are hubs of productivity and often a refuge for the poor (Glaeser 2011). They reflect a balance between two externalities. Agglomeration economies increase productivity and amenities, drawing people into cities, while congestion externalities reduce health, safety, and effective density, driving people out. A significant function of city governance is to provide local public goods, such as parks, transportation and police stations, that address congestion and allow more people to benefit from the positive externalities of density. But the partial-excludability of these local public goods makes them hard to target at specific groups, leading to a risk that the benefits of dense living are denied to some people. For example, local rent increases that follow the construction of transport infrastructure may price out the incumbent population (e.g., Balboni et al. 2021), and harassment of females on public transport may discourage use and mean that cities are effectively less dense for women (e.g., Borker et al. 2020; Kondylis et al. 2025).

If we can understand the characteristics of local public goods that stay with those that they target, rather than being captured by others, then we can work towards better design and delivery of public goods in the world's cities. But the data available to evaluate local public good programs, typically in the form of repeated cross sections measured at the household or location level, is ill-suited to study the targeting problem. For example, if we do not observe the outcomes of households initially living in targeted areas but who move away, or if data are not disaggregated by the sex of household members, then it is difficult to understand the impacts on targeted residents. These problems are compounded by limited data availability in developing country cities, where even repeated cross sections are rare. This scarcity of appropriate data in turn means that there is little direct evidence on whether public goods stay with those at whom they are targeted. Recent evidence on the distributional implications of investments takes a structural approach to overcome these challenges (Tsivanidis 2017; Couture et al. 2024), but often requires strong modeling assumptions.

We aim to address this gap in the context of the construction of Dar Es Salaam's Bus Rapid Transit System (BRT), using data that we collected specifically to understand the

targeting problem. The Dar es Salaam BRT, the first of its kind in East Africa, opened in 2016. It aims to provide rapid public transport via dedicated busways in one of Africa's fastest-growing cities. The BRT is World Bank funded, and it is hoped that programs like this will help East Africa's cities to become healthy and wealthy places to live for all. A particular focus is the aim that improved public transport will help to connect more households to work, leading to increased incomes.

Our study is based on two rounds of panel data that we collected through household surveys. The first round (baseline) was collected prior to the opening of the first BRT line and the second round (endline) approximately three years after the line started running. Our data has three key characteristics, designed to help answer the targeting question. First, in order to study the potentially gendered labor market impacts of the infrastructure, we sampled one random working age male and one random working age female in each of our baseline households. Second, we track all households (and the male and female respondents) in the original sample, including those that may have moved away from their original dwelling. Finally, we keep track of the original set of dwellings in our baseline sample, interviewing those who move in, including collecting retrospective data from these households about their outcomes at the time of our baseline survey, and once again drawing a random male and female respondent. Separating between individuals and locations is quantitatively important because of high within-city mobility: over three years, 40% of individuals (and more than 50% of renters) move house.

Our data permits estimation of casual effects of the BRT for several groups of *people* that are of policy interest. First is the population in a location that decides to stay after the BRT is built ('stayers'). Second is the set of people who decide to depart after construction of the BRT ('exiters'). Third are those who move into a location after construction of the BRT ('arrivers'). We use a differences-in-differences (DID) framework and show that, with suitable parallel trends and SUTVA assumptions, our data allows us to estimate the causal effect for stayers combined with exiters, a group that we refer to as *initial* residents, and also for stayers plus arrivers, a group that we refer to as *final* residents. The causal effect of the infrastructure on *initial* residents is of particular interest as this is the group who were initially targeted to receive access to the BRT infrastructure. Causal estimates

for final residents, on the other hand, tell us about the impact on those who are the actual recipients of the public good, rather than the targeted population. Furthermore, we can separately estimate the causal effect of the BRT on initial female and male residents, as well as final female and male residents.

We show that a DID estimator that establishes causal effects *at the location level* based on standard repeated cross section data does not estimate a causal effect for initial residents, nor final residents. There are two straightforward ways to decompose the location level effect. It either identifies the causal effect on initial residents plus a selection effect in the form of differences in *post-intervention* outcomes between exiters and arrivers; or alternatively identifies the causal effect on final residents plus a selection effect from the difference in *pre-intervention* outcomes between exiters and arrivers. These results are intuitive: in the absence of panel data that tracks respondents, it is not possible to observe outcomes for the exiter subset of initial residents post-treatment, nor the arriver subset of final residents pre-treatment. If selection effects are strong, a location level analysis may find a positive effect even though there is *no* causal impact on either initial or final residents. One common approach to avoid this problem in the existing literature is to control for changes in demographic variables, such as education of the household head, which are not likely to change over time and hence can be extracted from a repeated cross section. We can use our data to evaluate whether this strategy is effective in overcoming this selection challenge.

Applying this framework to our Tanzanian data, we find striking implications for understanding the labor market effects of the BRT. First, consistent with at least some households valuing the BRT, but also with concerns about targeting, there are large increases in reported rental values near the BRT, with a 19% increase in expected rent per room for the average 'highly-exposed' structure at the 80th percentile of the main treatment variable. Second, we show that location level analysis would suggest a large increase in labor force participation and incomes. Highly-exposed locations see a 7 percentage point increase in the probability of being employed (relative to a mean endline employment rate of 66%) and a 5,732 TSH increase in earnings (relative to a mean endline value of 25,530 TSH), with impacts concentrated among females.

The decomposition permitted by our tracking data reveals that much of this estimate is driven by a selection effect. At the household level, the casual estimate for initial residents is statistically indistinguishable from zero for both labor force participation and earnings, with point estimates that are 72% and 43% smaller respectively than the structure-level estimates. Descriptive analysis shows that the arrivers who replace the exiters were better educated and more likely to have worked at baseline. Disaggregating by gender, we continue to see statistically significant labor force participation impacts for initial female residents, but these are only about half the size of the location level estimates, while earnings effects are statistically insignificant. Final resident estimates reveal similar muted effects. Finally, we show that neither controlling for demographic characteristics of the type likely to be found in a repeated cross section, nor the more comprehensive set of time varying baseline and endline characteristics available in our data, helps to recover the causal effects on initial or final residents from the location level estimates.

These results have both empirical and methodological implications. On the empirical side, we document important impacts of transport infrastructure on female labor force participation, but also show clear evidence that infrastructure is captured by a set of people who are better off to start with along key dimensions. On the methodological side, we show that location level analyses can lead to incorrect conclusions about impacts both quantitatively and qualitatively. An incautious analysis of our data, even one that controls for demographic characteristics, would conclude that the BRT had strong impacts on household earnings and employment, but this would be erroneous: we see no causal impacts on household earning or labor force participation for either initial or final residents. The BRT did improve labor market outcomes for female initial residents, but by much less than the location level analysis would suggest, and we see no impacts on earnings. Overall, the BRT helped places more than it helped people.

The final section of our analysis provides suggestive evidence as to why the BRT is a piece of infrastructure that women are happy to use, and how this leads to an increase in female labor force participation. Self-reported data on perceptions of the BRT reveal that females are more satisfied with the improvement in public transport, accessibility and safety offered by the BRT than males. The impact of the BRT on responses to survey

questions regarding satisfaction with, and convenience of, transport options is also much stronger among female relative to male respondents. Overall, these results suggest that the BRT differentially improves public transport options to useful locations for women in the sample.

This paper contributes to several literatures. A fast-growing literature in urban economics uses quantitative spatial equilibrium models to study how investments in public transportation affect city structure and aggregate outcomes across a range of contexts (Ahlfeldt et al. 2016; Heblich et al. 2020; Zárate 2022), as well as distributional implications for inequality (Tsivanidis 2017; Barwick et al. 2024). We complement this literature by focusing on how population resorting and accurate tracking of mobile populations can affect estimates of the implications of transport infrastructure for labor market outcomes.

A small but growing literature uses a combination of repeated cross section and demographic data, together with quantitative models, to study gentrification. Gechter and Tsivanidis (2023) provide reduced form evidence for gentrification caused by the redevelopment of old mill cites in Mumbai, and use a dynamic spatial model to estimate welfare implications. Weiwu (2023) and Weiwu (2024) study the impact of the US interstate highway system, concentrating on the sorting of populations across locations as amenity and market access change, and again finds evidence for gentrification type effects. These studies do not leverage panel data but instead infer gentrification effects from changes in demographics in affected areas; in contrast, we leverage panel data tracking residents as they move in order to provide direct evidence for population resorting in response to the new infrastructure. Almagro et al. (2023) and Almagro and Domínguez-Iino (2024) use structural quantitative models to examine similar impacts of Chicago's public housing demolitions and mass tourism in Amsterdam, respectively. Finally, Couture et al. (2024) uses a quantitative model to estimate the extent to which gentrification in cities accentuates rising inequality. Perhaps most closely related to this paper, Autor et al. (2025) study the causal impact of the China shock using tax data that forms a panel. While the focus of their analysis is different, they have a related emphasis on the difference between impacts on people and impacts on places.

Also related, recent research underscores how household dynamics, location choices,

and public policy shape labor market outcomes, particularly for women. Venator (2022) shows that policies like unemployment insurance for trailing spouses can ease migration constraints, increasing mobility and long-term earnings, especially for women. Ranosova (2023) highlights how suburbanization and long commutes disproportionately affect married women's labor force participation while reinforcing household specialization. Moreno-Maldonado (2022) further demonstrates that city size influences gender gaps in employment, with women—especially mothers—participating less in labor markets in large cities due to longer commutes, occupational sorting, and higher childcare costs. Finally, Moreno-Maldonaldo and Santamaria (2024) link delayed childbearing to urban revival, as young, childless professionals concentrate in city centers, driving economic and spatial change. Together, these studies emphasize the role of mobility constraints, urban structure, and household decision-making in shaping labor market disparities—factors crucial for understanding the broader impact of place-based policies like mass transit expansion.

The rest of the paper is structured as follows. Section 2 describes the empirical setting and data collection. Section 3 discusses key descriptive statistics and motivating facts from our panel survey data. Section 4 outlines the framework we use to study the causal impacts of the BRT on initial and final residents in targeted locations, and distinguish these from selection effects. Section 5 discusses our results. Section 6 concludes.

2 Empirical setting and data collection

2.1 Dar es Salaam Bus Rapid Transit system

With a growth rate of 6.5%, Dar es Salaam is one of the fastest-growing cities in Africa (The World Bank (2019)). Congestion is a severe problem (Mpogole and Msangi (2016)), which costs the city an estimated $1.8 million each day in lost productivity (The World Bank (2019)). To help address these challenges, a six-phase BRT system of dedicated trunk lanes spanning 141km, shown in Figure 1, is being implemented in the city from 2005-2035, the first of its kind in East Africa.

Operations commenced on the system's first phase, connecting the city's central business district to residential areas in the city's northwest (Kinondoni), in May 2016. Within its first year of operations, the system was carrying 165,000 passengers per day. Later phases of the BRT system are planned along other radial routes connecting the central business district to the south (Temeke district, Phase 2), southwest (Ilala district, Phase 3), and north (Phase 4), as well as orbital and connecting routes (Phases 5 and 6).

We undertook household and individual level panel surveys that allow us to track both structures and residents from before the opening of the BRT until three years afterwards. We collect a baseline survey of 1748 structures and interview one male and one female respondent from each. We then track along two dimensions during our endline survey three years later. Appendix Figure A1 outlines the timeline of survey and operational activities. [1]

We resurveyed the household members interviewed at baseline, wherever they were living at endline (creating a panel of households), and we surveyed any new households residing in the original structures surveyed at baseline (creating a panel of structures). For new residents of original structures, we collect retrospective data on their key outcomes three years prior, in order to obtain baseline observations for all individuals in the panel.[2]

2.2 Baseline household survey

A baseline household survey was conducted in January-February 2016, before the first phase of the BRT began operations in April. We used a geographical sampling strategy to ensure coverage across the entire city of Dar es Salaam by selecting 141 clusters at equal intervals along 12 arcs at radii increasing at 1.5km intervals from the central business district, as shown in Figure A2. We conducted interviews at 125 of these clusters.[3] At each cluster location, a random walk was used to select 12-14 households to interview, yielding a total of 1748 households that were available for and consented to interviews.

[1] In addition to the primary baseline and endline survey rounds, we also completed some mobile surveys and an incomplete midline survey. These additional surveys were used primarily to assist with tracking households.

[2] Fuller et al. (2024) discuss the accuracy of retrospective survey data using data from Ethiopia.

[3] The remaining 16 clusters were military or special residence compounds, non-residential areas, or designated hazardous areas where we were not able to conduct interviews.

We conducted three interviews at each household. A household module was conducted with any knowledgeable household member found at the house, covering household demographics, dwelling information, assets, consumption, and summary education and employment information for all household members. A separate survey was then administered to one male and one female respondent aged above 17 years, randomly selected from their respective qualifying group in the household. This survey included more detailed questions on employment, income, commuting and neighborhood amenities. This survey was administered to a total of 3104 individuals. We construct weights to reweight our sample to be representative of the population of Dar es Salaam overall. All results are presented with population weights.[4]

2.3 Endline household survey

An endline household survey was conducted from February-May 2019.[5] All baseline respondents and structures were tracked as part of this survey. Hence, if a baseline household had moved out of its baseline structure, the household was tracked to its new structure, and interviews were conducted with both the original household in its new structure and with the new occupants of their original structure.[6] Baseline respondents who split from their baseline household and moved elsewhere were also tracked and a household survey was conducted for their new household.

Appendix Table A2 summarizes structure-, household- and individual-level recontact rates at endline.[7] Surveying households in a dynamic urban setting is challenging. We successfully located 91.8% of our baseline structures at the endline approximately three years later and successfully surveyed 76% of the sample. 10% of households living in structures refused (either at the midline survey or at endline), 8.2% were not found (either at midline or at endline), and 5.8% of structures were torn down or empty.

[4] Details of the weights are presented in Appendix B.

[5] We conducted a follow-up telephone survey in September-October 2019 to refine some variables given inconsistencies in some survey responses. Refer to the Data Appendix for full details.

[6] The only exception to this was households who had moved out of Dar es Salaam, who were not tracked.

[7] Appendix Table A2 shows the same table for ever-enrolled households. Midline and SMS surveys conducted in 2017-2018 were also conducted to keep in contact with as many respondents as possible between the baseline and endline surveys to reduce attrition.

The second panel shows the contact rates for households. Overall, we made contact with 89.8% of the baseline sample, successfully surveying 79.6% of them at the endline. The individual contact rate was 84.9%; we successfully surveyed 69.7% of the original individual sample at endline. For the sample of households who were successfully surveyed during the midline, Column (2) shows that we resurveyed 88% of structures, 92% of households, and 81% of individuals at endline. We present balance tests for attrition across the three samples in Appendix Tables A3-A5. We do not find the geography (i.e., distance to the BRT) predicts who was successfully surveyed, but, as perhaps expected in locations with a large amount of mobility, we are more likely to be able to recontact individuals who were owners rather than renters, and those who had lived in their structure for more years.

3 Motivating facts on commuting and housing

3.1 Baseline commuting patterns by gender

Panel B of Table 1 describes summary statistics of male and female survey respondents during both survey rounds. The baseline survey data reveals that 87% of males and 51% of females are employed. A proportionately higher share of working women relative to men work in the informal sector, which accounts for 79% of working females and 63% of working males. Poverty levels in the sample are relatively high, with 42% of respondents above the World Bank's poverty line of $5.50 per day.

Ninety one percent of working males and 72% of working females commute to work. Commute times are long, averaging 47 minutes across the sample, with especially long commutes among respondents living towards the outskirts of the city, and for males (55 minutes) relative to females (33 minutes).

In addition to aggregate commuting rates and durations, commuting modal choice differs markedly across males and females in the sample. Less than 10% of male or female commuters commute by car. A much larger share of women walk to work – 52% relative to 25% for men – while rates of commuting via public transit are correspondingly

higher among males. The main modes of public transit available in Dar es Salaam before the introduction of the BRT were shared minibuses (daladalas) and three wheeler taxis (bajajis), with smaller shares accounted for by motorbike taxis and car taxis.

3.2 Housing characteristics and household churn

Panel A of Table 1 describes structure-level summary statistics from the baseline and endline surveys. At baseline, 37% of surveyed respondents own their house, while 56% are renting. Dwellings comprise 3.44 households on average, with each household occupying an average of 2.45 rooms.

Rates of household churn in the data are high, suggesting that there may be significant scope for population resorting. Seventeen percent of structures surveyed had a different household residing in them at the time of the endline survey relative to the baseline survey date three years earlier. Between the two surveys, 40% (39%) of tracked women (men) moved house. These churn rates are especially pronounced among renters, with 52% (50%) of tracked women who were renting moving house over this period.

4 Framework

This section discusses how we combine our novel tracking data with key identification assumptions in order to isolate the causal effect of the BRT on well defined groups of people: initial residents (stayers and exiters) and final residents (stayers and arrivers). Impacts on initial residents are of particular policy interest as they are typically the group to whom the public good was targeted. Final residents may also be of interest given that they are the group that ultimately benefit from the public good, though it may be difficult for the planner to predict who these final residents will be. We argue that the parallel trends assumption typically invoked in estimation of location level effects will generally imply the parallel trends assumptions required for our initial and final resident estimates, but that standard location level analysis does not recover a causal effect for any group of people.

4.1 Identifying causal effects and selection

Let $\Delta y_{ji}(\Delta T) = \left(y_{ji}^{EL} - y_{ji}^{BL}\right)(\Delta T)$ denote the potential change in an outcome y (e.g., earnings) between baseline and endline for individual j who lived at an initial location i at the time of our baseline survey, where $\Delta T = \{\Delta T_1, \ldots, \Delta T_N\}$ is the change in travel times to the CBD caused by the BRT, for all initial locations $\{1, \ldots, N\}$. We assume parallel trends:

$$\Delta y_{ji}(0) \perp\!\!\!\perp \Delta T_i, \tag{1}$$

and a Stable Unit Treatment Value Assumption (SUTVA):

$$\Delta y_{ji}(\{\Delta T_1, \ldots, \Delta T_i = 0, \ldots \Delta T_N\}) = \Delta y_{ji}(\{\Delta T_1', \ldots, \Delta T_i = 0, \ldots \Delta T_N'\}) \ \ \forall i.$$

We also assume that there are no anticipation effects, although we discuss the implications of relaxing this assumption below. With these three assumptions, the difference-in-differences (DID) regression

$$\Delta y_{ji} = \delta^I + \beta^I \Delta T_i + \epsilon_{ji} \tag{2}$$

identifies the causal effect of the change in travel times caused by the BRT on the outcomes of initial residents, if the impact is linear.

Let $\Delta y_{jf}(\Delta T)$ define a potential change for individual j who will live at final location f at the time of our endline survey if the change in travel times is ΔT. A similar set of assumptions[8] imply that the DID specification

$$\Delta y_{jf} = \delta^F + \beta^F \Delta T_f + \epsilon_{jf} \tag{3}$$

identifies the causal effect of changes in travel times due to the BRT for final residents.

Finally, in order to identify causal effects of the BRT on locations, let $\Delta y_{jj'l}(\Delta T)$ be the potential change in outcomes of the occupant of location l if the change in travel times is ΔT. The notation jj' highlights that the occupant need not be the same person

[8]The parallel trends assumption for final residents is: $\Delta y_{jf}(0) \perp\!\!\!\perp \Delta T_f$; and the SUTVA is $\Delta y_{jf}(\{\Delta T_1, \ldots, \Delta T_f = 0, \ldots \Delta T_N\}) = \Delta y_{jf}(\{\Delta T_1', \ldots, \Delta T_f = 0, \ldots \Delta T_N'\}) \ \ \forall f$.

at baseline and endline. Similar parallel trends and SUTVA assumptions[9] imply that the DID specification

$$\Delta y_{jj'l} = \delta^L + \beta^L \Delta T_l + \epsilon_{jj'l} \tag{4}$$

identifies the causal effect of the BRT-induced change in travel time for the location l.

We next argue that the required parallel trends assumptions are likely to hold, at least conditionally, and are in fact weaker than those used in common location-level analyses. Recall that residents in a location l can be divided into three groups: stayers (who are in the location at baseline and endline), exiters (who are there at baseline but not endline), and arrivers (who are not there at baseline, but are at endline). Denoting p_l^s to be the probability that an individual in location l will be a stayer, S to be the set of stayers, A arrivers and E exiters, the expected location level trend in location l is:

$$\begin{aligned} \Delta y_{jj'l} &= p_l^s \Delta y_{j\in S} + (1-p_l^s)\left(y^{EL}_{j'\in A} - y^{BL}_{j\in E}\right) \\ &= p_l^s \underbrace{\Delta y_{j\in S}}_{stayer} + (1-p_l^s)\left(\underbrace{\Delta y_{j'\in A}}_{arriver} + \underbrace{\Delta y_{j\in E}}_{exiter} - \underbrace{(y^{EL}_{j'\in A} - y^{BL}_{j\in E})}_{selection}\right). \end{aligned} \tag{5}$$

If trends are parallel for all terms in (5) separately – corresponding to stayers, arrivers, exiters and selection – then the parallel trends assumption clearly holds at the location level. While we could have parallel trends at the location level without parallel trends for all sub-components, this would require that any deviation from parallel trends in one term in (5) is exactly offset by the other terms.[10] Given the implausibility of this knife-edge case, the assumption of parallel trends at the location level is highly likely to imply parallel trends for stayers, arrivers, exiters, and selection independently.

If the parallel trends assumption holds for each subgroup, we show that it then holds

[9]The parallel trends assumption for the location is $\Delta y_{jj'l}(0) \perp\!\!\!\perp \Delta T_d'$; and the SUTVA is $\Delta y_{jd}(\{\Delta T_1, \ldots, \Delta T_d = 0, \ldots \Delta T_N\}) = \Delta y_{ji}(\{\Delta T_1', \ldots, \Delta T_d = 0, \ldots \Delta T_N'\}) \ \forall d$.

[10]For example, if the earnings of stayers in BRT locations would have trended upward faster than those of stayers in other locations, we would require that the earnings of arrivers plus exiters and selection trend down faster by exactly the degree required to offset this.

for initial and final residents, by decomposing the trends for initial and final populations,

$$\begin{aligned}\Delta y_{ji} &= p_i^s \underbrace{\Delta y_{j\in S}}_{stayer} + (1-p_i^s) \underbrace{\Delta y_{j\in A}}_{arriver}, \text{ and} \\ \Delta y_{jf} &= p_f^s \underbrace{\Delta y_{j\in S}}_{stayer} + (1-p_f^s) \underbrace{\Delta y_{j\in E}}_{exiter}\end{aligned}$$

respectively. These equations make clear that, if the parallel trends assumption holds for stayers, arrivers and exiters separately, then it also holds for the initial and final populations of residents.

The location level estimates can be related to both the initial and final resident analysis by manipulation of (4), detailed in Appendix E, which yields:

$$\begin{aligned}E\Delta y_{jj'l} &= \beta^I + (1-p_l^s)\left(Ey^{EL}_{j'\in A} - Ey^{EL}_{j\in E}\right) = \beta^I + (1-p_l^s)\sigma^{EL} \\ &= \beta^F + (1-p_l^s)\left(Ey^{BL}_{j'\in A} - Ey^{BL}_{j\in E}\right) = \beta^F + (1-p_l^s)\sigma^{BL}\end{aligned}$$

where, as above, p_l^s is the proportion of residents in dwelling l that are stayers. In this equation, σ^{EL} and σ^{BL} are selection terms. σ^{EL} reflects the fact that those who arrive in location l after the treatment may have different *endline* characteristics from those who exit. As such, one interpretation of the dwelling level regression is the causal effect on initial residents, plus a selection effect related to changes in *endline* characteristics; here it is endline characteristics that are relevant because arrivers may differ from exiters partly due to the causal effect of the BRT on their endline outcomes. σ^{BL} reflects the fact that those who arrive in location l after the treatment may have different *baseline* characteristics from those who exit. Hence, an alternative interpretation of the location level estimate is the causal effect on final residents, plus selection on baseline characteristics.

An implication of this analysis is that attempts to control for selection, in order to recover causal effects for initial residents, would either need to control for unchanging demographic characteristics, or have access to data on endline characteristics of exiters. Below we consider whether the inclusion of such controls can help to recover the policy-relevant causal impacts on people (initial or final residents) from estimates of the impact

on location, or place.

4.2 Difference-in-differences with post double selection lasso

Given the practical difficulties associated with targeting the BRT towards areas with differential trends in specific outcomes, it is likely that the parallel trends assumption may be plausible in our empirical setting. Nevertheless, we describe here the approach we take to relax the parallel trends assumption somewhat. To do so, we include controls for a large number of possible baseline characteristics, employing a principled method for selecting controls without over-fitting by using the post double selection lasso approach suggested by Belloni et al. (2014).

We apply this approach directly to the first differenced regression specifications, for example implementing:

$$\Delta y_{ji} = \delta^I + \beta^I \Delta T_i + \gamma X_{ji} + \epsilon_{ji}$$

where X_{ji} are characteristics of those living in location i at baseline. The parallel trends assumption is then relaxed to the following conditional form:

$$\Delta y_{ji}(\Delta T) \perp\!\!\!\perp \Delta T_i \mid X_i$$

4.3 Analysis of pre-trends

To support our identification assumptions, we consider evidence for pre-trends in our key outcomes between 2015 and our baseline survey, using data from the World Bank's Measuring Living Standards Within Cities project, which was collected in Dar es Salaam in 2014 and 2015.[11]This dataset captures different households to those included in our survey data collection, and includes location identifiers at the level of enumeration areas. Our analysis of pre-trends is therefore conducted at the level of enumeration areas.

To link the two datasets, we assign each structure in our baseline survey to the nearest enumeration area in the World Bank data. Some enumeration areas from the 2015 World

[11]To do this, we accessed anoymized restricted-access data that contained geographical information of households at the level of enumeration areas. We thank Nancy Lozano Garcia at the World Bank for facilitating data access.

Bank survey are never assigned to structures in our baseline survey due to distance, and data from these areas are excluded from the pre-trends analysis.

We focus on variables that were consistently collected across both datasets, including imputed rent per room, duration in the house, typical number of days worked per week, typical number of hours worked per day, and household asset ownership. Due to differences in income reporting between the two surveys, income is not included in our pre-trends analysis.

Using the survey data, we generate a residualized rent variable by regressing log rent per room on household and structure characteristics, including access to electricity for lighting, use of a non-latrine toilet, availability of piped water, cooking fuel (oil, gas, or electricity), and whether the structure has cement walls or a slate roof. We also construct a wealth index using Principal Component Analysis (PCA) on household asset data, with the first principal component serving as the index.

The pre-trends analysis follows the same approach as our main analysis, using the post double selection lasso method to control for initial characteristics. As in the main analysis, our dependent variables are expressed as difference variables.

The results, presented in Table A7, indicate no significant differences in pre-existing trends across treatment intensities before the BRT's implementation.

4.4 Measuring exposure to the BRT

Exposure to the BRT is measured as the decrease in log travel time from the structure to the central business district (CBD, assigned to the centroid of Gerezani ward) as a result of the construction of BRT phase 1. This is calculated using the Network Analyst tool in ArcGIS and travel time data collected at baseline and endline as described at Appendix D. Predicted baseline travel times from each structure to the CBD are calculated by assigning to each stretch of the road network a travel speed equal to the average travel speed in the baseline travel time survey. Predicted endline travel time calculations isolate changes in travel times resulting from the BRT by assigning the same travel speeds to all stretches of the road network *except* for stretches along the BRT Phase 1 route, which are assigned

average BRT travel speeds. The implied reduction in log travel time to the CBD using this measure is visualized in Figure 2 panel A.

A recent literature highlights potential challenges in estimating the causal effects of travel time reductions based on transportation improvements if neighborhood characteristics are correlated with exposure to the treatment (Borusyak and Hull (2023)). For example, if BRT lines are more likely to be constructed along radial routes that connect the CBD, more treated locations may be expected to see differential travel time improvements relative to less treated locations even in the absence of BRT construction. To address this, we follow the method proposed in Borusyak and Hull (2023) by demeaning the measure shown in Figure 2 panel A – capturing exposure to the BRT Phase 1 line – by the expected reduction in travel time implied across the first four proposed lines of the BRT system, shown in Figure 2 panel B.

This procedure yields our key treatment variable, the demeaned predicted reduction in log travel time from each location to the CBD, shown in Figure 2 panel C. The distribution of the demeaned exposure variable, shown in Appendix Figure A3, reveals a median value of -0.27 and an eightieth percentile value of 0.101. When interpreting coefficient magnitudes, we consider results at this value of the exposure variable to capture estimated effects for an average respondent who was highly exposed to Phase 1 of the BRT.

4.5 Anticipation effects

As discussed in Section 4.1, estimation of the causal effect of the BRT using our framework rests on the assumption that there were no anticipation effects. This implies that our baseline survey data should not have been influenced by respondents' anticipation that the BRT was going to start operating.

The assumption that anticipation effects were not first order in our empirical setting may be plausible for two key reasons. First, as with many infrastructure projects of this type in low income countries, concerns remained right up until the opening of the line that it might not be implemented (Krüger et al., 2021; Jacobsen, 2022; Daily News Reporter,

2024).[12] This uncertainty is likely to have muted anticipatory behaviors.

Second, we can test for anticipation effects ahead of the construction of Phase 2 of the BRT system, for which funds were secured in late 2015, with construction scheduled to begin in 2019. This period coincides with our survey data collection timeline, allowing us to test whether there was evidence for changes in outcomes along the Phase 2 line in anticipation of its subsequent opening.

Importantly, the likely direction of bias resulting from plausible forms of anticipatory behavior suggests that such effects are unlikely to alter our qualitative conclusions. In particular, the most plausible form of anticipation is that those believing that areas served by the BRT will subsequently become more desirable move in to these areas, bidding up local rents. In this case, the group that our data identifies as initial residents may miss some initial residents who have already been pushed out of areas nearby the BRT, and include some individuals who should instead be included among the group of final residents. This should tend to bias upwards our estimates of causal impacts on initial residents, and bias downwards our estimates of selection. Given that our central qualitative finding is that impacts on initial residents are muted while selection effects are large, anticipation effects are, if anything, likely to *strengthen* our central conclusions.

5 Results

5.1 BRT usage and impacts on local dwelling characteristics

Table 2 reveals that a large share of individuals have ridden the BRT at the time of the endline survey, with mean values of 92% across all males and 87% across all females. Individuals residing in locations with stronger exposure to the BRT – measured as those respondents for which the demeaned predicted decrease in log travel time from the neighborhood to the CBD as a result of the opening of BRT Phase 1 exceeds the 80th percentile value of 0.101 (hereafter 'highly exposed'), as described in Section 4.4 – are intuitively

[12] Krüger et al. (2021) highlight that such concerns remained even in the months preceding the start of operations on Phase 1 of the BRT. In 2015, once construction was complete, "the service could not be launched because a provider for the bus operation was still missing".

more likely to use it.

BRT exposure is associated with especially pronounced increases in the use of the BRT to get to work for male respondents. In highly exposed locations, 31% of males use the BRT to get to work, but only 13% of females, relative to 4% and 2% respectively in low exposure locations. Conversely, BRT exposure drives stronger effects on the use of the BRT for shopping among females: 33% of highly exposed females and 30% of highly exposed males used the BRT for shopping in the last seven days, relative to 8% and 10% respectively in less exposed locations.

We next consider the impact of the BRT on characteristics of structures in treated locations, shown in Table 3. Reassuringly, BRT exposure reduces respondents' reported travel time to the CBD, with highly exposed respondents reporting a sizable reduction of 10 minutes relative to a mean endline value of 71 minutes. Monthly household expenditures on transportation also increase slightly. For treated respondents, average expected residualized rent per room[13] increases, suggesting that households value access to the new BRT infrastructure. The average highly-exposed structure sees a 19% increase in expected rent per room. The results in column 4 reveal that there are not differential rates of household churn in locations that see a larger reduction in predicted travel time to the city center as a result of the BRT.

5.2 Impacts on labor market outcomes

The central set of results in Table 4 turn to establishing the causal effects of the BRT on labor market outcomes, and the key methodological contribution of distinguishing causal impacts from selection effects using our tracking data.

The first column of Table 4 displays the estimate of the impact of the BRT on employment status that we obtain using a dwelling-level specification estimable using standard

[13]Expected log rent per room is constructed by dividing self-reported expected log rent by the self-reported number of rooms. The residualized version of this variable is generated by regressing expected log rent per room on the following variables and then obtaining the residuals from the regression: an indicator for having electricity as the main source of lighting in the household; an indicator for a non-latrine toilet being the main type of toilet used by the household; an indicator for the main source of household drinking water being piped water; an indicator for the main source of household cooking energy being oil, gas, or electricity; an indicator for the primary building material for the exterior walls being cement; and an indicator for the primary building material for the roof being slate.

repeated cross-sectional data. This suggests sizable and statistically significant impacts on labor force participation, driven entirely by females. On average, highly exposed individuals experience a 7 percentage point increase in the probability of being employed, relative to a mean endline value of 66%. This is driven by an 11 percentage point increase for females relative to a mean endline employment rate of 54%, while males have a much higher mean endline value of 82% with no significant impact of the BRT. The corresponding estimates in column (4) of Table 4 reveals that this translates into large increases in female earnings. The average effect for highly exposed females is 8,306 TSH, a 57% increase relative to the mean endline value of 14,686, while males again see no statistically significant change.

These findings at the dwelling level may be consistent with both a causal impact of the BRT and with differential selection into treated areas of women with higher ex-ante labor force participation rates or earnings. Regression coefficients on the treatment variable comparing the observable characteristics of those living in each dwelling at the time of the endline versus the baseline survey is shown in Table 5, shown separately where characteristics are measured at baseline or endline. These comparisons indeed suggest that final residents of dwellings in locations more strongly treated by the BRT are positively selected, especially for females. In particular, treated females are more educated and more likely to commute outside the home based on endline characteristics, and more likely to be employed, commute, and earn more based on baseline characteristics. For males, differences tend to be weaker and are generally insignificant.

As outlined in Section 4, our tracking data allows us to distinguish causal from selection effects and isolate the true causal effect of the BRT on initial residents in treated locations, which in many contexts may be the most policy-relevant object of estimation. Based on the estimand defined in Equation 2, columns (2) and (5) of Table 4 reveal that there is no longer evidence of a statistically significant causal effect of the BRT on labor force participation or weekly earnings at the household level. The insignificant point estimate for labor force participation of initial residents is 64% smaller than the structure-level estimate, and equality of the two coefficients can be rejected with a p-value of 0.139. We cannot reject equality of the coefficients at conventional levels for weekly earnings,

though again the initial resident estimates lose statistical significance and the point estimate sees a 40% decline relative to the structure-level estimates.[14]

These comparisons highlight that a conventional structure-level analysis would lead researchers interested in impacts on initial residents to conclude erroneously that the BRT had positive causal impacts on labor force participation and earnings. In contrast, our tracking data reveals that, while impacts may be positive at the *location* level, *initial residents* do not benefit from the new infrastructure.

Given evidence for positive structure-level results but not for positive impacts on initial residents, the question arises as to whether other groups may have benefited from the BRT, or whether the positive structure-level results reflect a pure selection effect. To investigate this, we use the estimand in Equation 3 to estimate causal impacts on *final residents*, shown in columns (3) and (6) of Table 4. These estimates find no statistically significant impact on labor force participation or earnings among final residents. As such, the structure-level estimates are likely to be driven largely by selection effects.

Figure 3 considers how close to the true causal estimate of BRT impacts on initial residents (in columns (2) and (5) of Table 4) we might get with easier to obtain data, by controlling for observable respondent characteristics. For some characteristics which are unlikely to vary non-deterministically over time, such as age and (to a large extent) education, these control variables may be observed even without access to tracking data. For other characteristics, estimating these results still require our tracking data, given that the estimation relies on differences in observable characteristics between baseline and endline respondents in the dwelling as *measured at endline* (see Section 4). Nevertheless, this comparison is useful in elucidating how far the influence of selection effects may be driven by observable versus unobservable characteristics.

The bottom estimate in Figure 3 (labeled 'structure') shows the structure-level estimate for employment (panel (a)) and earnings (panel (b)) impacts, corresponding to the estimates in columns (1) and (4) respectively in Table 4. Moving upwards in the figure, we run the same specification but add sequential controls for differences between the

[14]The results restricting to females show that initial resident estimates remain statistically significant for labor force participation, though not for earnings. For both outcomes, point estimates and statistical significance are substantially attenuated in the initial resident, relative to the structure-level, estimates.

baseline and endline respondent within each dwelling. The first of these (demographic) controls for differences in age and education; the second for employment and commuting status as well as earnings (employment); the third for total expenditure (expenditure); and the next (all controls) for all of these together.

Adding in these controls, separately or together, can be seen to result in marginal changes in the point estimates and statistical significance of the structure-level estimates, but a consistent picture remains of sizable impacts, which are mostly statistically significant. In contrast, the true causal impact on initial residents (corresponding to the estimates in columns (2) and (5) in Table 4) are much smaller and statistically insignificant for both outcomes. This can be seen especially starkly in the equivalent plots restricting attention to female respondents in Figure 4. In this case, adding the controls described results in almost no change in the estimated structure-level effects, while the estimated causal effects on initial residents are much lower. These results suggest that the conflation of causal and selection effects in the structure-level estimates are largely driven by selection on unobservable characteristics.

Policy concern that incumbent residents may not benefit from place-based investments targeted at them often centers on rental rate increases (Kennedy and Leonard 2001), which may be expected to displace initial residents. We examine how far this concern may be warranted in our empirical setting by running the same set of specifications presented in Table 4, but restricting attention to the 56% of respondents who were living in rented accommodation at the time of the baseline survey. The results suggest a similar pattern of results for this group. Results are presented in Appendix Table A1. Labor force participation results (shown in columns (7) to (9)) show statistically significant positive estimates at the level of both structures and initial residents, though the initial resident estimate is 65% smaller and statistically distinguishable from the structure estimate. Earnings estimates (in columns (10) to (12)) again reveal that a structure-level analysis would imply significantly positive effects, while analysis at the level of initial residents would not.

5.3 Why does the BRT increase female labor force participation?

The results in the previous subsection suggest that the only statistically significant *causal* effect of the BRT on initial residents is an increase in female labor force participation, while there is no evidence for labor market impacts for males. In this section, we use other information contained in our survey data to consider potential mechanisms that may help to explain these divergent impacts by gender.

The endline survey asked highly-exposed respondents about their perceptions of the BRT after it had opened. Responses to these questions are summarized by gender in Table 6, where the first two columns consider all respondents, and the last two columns restrict attention to employed respondents only. These results suggest that there is evidence for gender differences in satisfaction with BRT amenities. Females report at higher rates that the BRT improves public transport options from the house and makes it easier to get their daily jobs done, and are more satisfied than males with BRT station accessibility and safety of driving and the vehicle. Among reported constraints that prevent respondents from using the BRT, females relative to males cite safety and security concerns with slightly higher frequency, and service reliability with lower frequency.

Table 7 examines how the BRT affects self-reported satisfaction with transportation options, all reported on a 1 to 10 scale and broken down by gender. Columns (1) and (2) consider initial residents' responses to the question 'How happy are you with public transportation options from the house?'. The estimates reveal that the BRT yields a statistically significant increase in female respondents' reported satisfaction with transportation options from their house, while the effect is less than half the magnitude and insignificant for males. Responses to the question 'How much do you agree with the statement "Where I live is convenient for where I want to go to"' in columns (3) and (4) show a similar pattern. Gender differences are apparent in the point estimates, but less pronounced than for other variables, in response to the question 'How much do you agree with the statement "I find it difficult to get my daily jobs done because of congestion"', shown in columns (5) and (6).

These findings are consistent with the BRT improving public transport, and serving

useful locations, differentially for women versus men. Columns (7) to (12), however, suggest that the BRT does not increase commuting or use of public transport to get to work among females. Taken together, these results suggest that – while the BRT does increase female labor force participation – this is not predominantly in jobs that require them to use the BRT to get to work. This could be consistent with, for instance, the BRT serving useful non-work locations that females frequent, but where increases in labor force participation are driven by employment in jobs that are sufficiently close to home for them to be accessed via foot (column (9)).

6 Conclusion

Urban infrastructure investments are typically place-based, yet policymakers often aim to improve outcomes for specific people rather than locations. In contexts with mobile populations, targeting locations does not guarantee benefits for initial residents. This paper develops a framework to separate selection effects from the true causal impacts of infrastructure investments, using panel data that tracks both individuals and structures before and after the introduction of East Africa's first BRT system.

Our results highlight the limitations of standard cross-sectional data in assessing the effects of place-based investments. A conventional neighborhood-level analysis would suggest large gains in labor force participation and earnings. However, by tracking individuals over time, we show that these estimates are primarily driven by selection, rather than real economic improvements for initial residents.

The only statistically significant causal effect we identify is an increase in female labor force participation. Survey evidence suggests that women perceive the BRT to be more accessible and safer than previous transport options, which may explain their increased workforce engagement. However, these effects are substantially smaller than would be inferred from neighborhood-level analysis.

More broadly, our findings illustrate how infrastructure investments can reshape the composition of affected neighborhoods, often benefiting a different population than initially targeted. This has important implications for the design and evaluation of place-

based policies: failing to account for population mobility can lead to misleading conclusions about who actually benefits.

References

Ahlfeldt, Gabriel, Stephen Redding, and Daniel Sturm, "The Quantitative Evaluation of Urban Transport Infrastructure Improvements," *Mimeo*, 2016, pp. 1–23.

Almagro, Milena and Tomás Domínguez-Iino, "Location sorting and endogenous amenities: Evidence from amsterdam," Technical Report, National Bureau of Economic Research 2024.

_ , Eric Chyn, and Bryan A Stuart, "Urban renewal and inequality: Evidence from chicago's public housing demolitions," Technical Report, National Bureau of Economic Research 2023.

Autor, David, David Dorn, Gordon H Hanson, Maggie R Jones, and Bradley Setzler, "Places versus people: the ins and outs of labor market adjustment to globalization," Technical Report, National Bureau of Economic Research 2025.

Balboni, By Clare, Gharad Bryan, Melanie Morten, and Bilal Siddiqi, "Could Gentrification Stop the Poor from Benefiting from Urban Improvements ?," 2021, pp. 1–8.

Barwick, Panle Jia, Shanjun Li, Andrew Waxman, Jing Wu, and Tianli Xia, "Efficiency and equity impacts of urban transportation policies with equilibrium sorting," *American Economic Review*, 2024, *114* (10), 3161–3205.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *Review of Economic Studies*, 2014, *81* (2), 608–650.

Borker, G, G Kreindler, and D Patel, "Women's urban mobility barriers: Evidence from delhi's free public transport policy," *IGC Final Report*, 2020.

Borusyak, Kirill and Peter Hull, "Non-Random Exposure to Natural Experiments," *Econometrica*, 2023, *91* (6), 2155–2185.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst, "Income growth and the distributional effects of urban spatial sorting," *Review of Economic Studies*, 2024, *91* (2), 858–898.

Daily News Reporter, "State warns against delays in BRT projects," *Daily News*, 2024.

Fuller, Rob, Simone Lombardini, and Cecilia Poggi, "How accurate are retrospective survey data? Evidence from rural Ethiopia," *Journal of Development Effectiveness*, 2024, *16* (1), 142–158.

Gechter, Michael and Nick Tsivanidis, "Spatial spillovers from high-rise developments: Evidence from the Mumbai Mills," *Unpublished manuscript*, 2023.

Glaeser, Edward, *Triumph of the city: How urban spaces make us human*, Pan Macmillan, 2011.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm, "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133.

Jacobsen, Malve, "Dar es Salaam's BRT: Why it has been a long bumpy ride," *The Citizen*, 2022.

Kennedy, Maureen and Paul Leonard, "Dealing With Neighborhood Change: A Primer on Gentrification and Policy Choices," Technical Report, Brookings Institution Center on Urban and Metropolitan Policy 2001.

Kondylis, Florence, Arianna Legovini, Kate Vyborny, Astrid Zwager, and Luiza Andrade, "Demand for "Safe Space": Avoiding harassment and complying with norms," *Journal of Development Economics*, 2025, *174*, 103392.

Krüger, Fred, Alexandra Titz, Raphael Arndt, Franziska Groß, Franziska Mehrbach, Vanessa Pajung, Lorenz Suda, Martina Wadenstorfer, and Laura Wimmer, "The Bus Rapid Transit (BRT) in Dar es Salaam: A pilot study on critical infrastructure, sustainable urban development and livelihoods," *Sustainability*, 2021, *13* (3), 1058.

Moreno-Maldonado, Ana, "Mums and the City: Household labour supply and location choice," *Available at SSRN 3466171*, 2022.

Moreno-Maldonaldo, Ana and Clara Santamaria, "Delayed Childbearing and Urban Revival: A Structural Approach," 2024.

Mpogole, Hosea and Samira Msangi, "Traffic Congestion in Dar Es Salaam: Implications for Workers' Productivity," *Journal of Sustainable Development*, 2016, *9* (6), 103.

Ranosova, Tereza, "Essays on Housing Economics and Family Dynamics." PhD dissertation 2023.

The World Bank, "Untying Dar Es Salaam's Traffic Knots, One Feeder Road at a Time," https://www.worldbank.org/en/news/feature/2019/04/01/untying-dar-es-salaams-traffic-knots-one-feeder-road-at-a-time 2019.

Tsivanidis, Nick, "The Aggregate and Distributional Effects of Urban Transit Infrastructure : Evidence from Bogotá's TransMilenio," 2017.

Venator, Joanna, *Dual-earner migration decisions, earnings, and unemployment insurance*, Boston College, 2022.

Weiwu, Laura, "Unequal access: Racial segregation and the distributional impacts of interstate highways in cities," Technical Report, Discussion paper, Technical report, MIT. 62 2023.

_ , "Opportunity in Motion: Equilibrium Effects of a Place-Based Policy on Economic Mobility," 2024.

Zárate, Román D, "Spatial misallocation, informality, and transit improvements: Evidence from mexico city," 2022.

Tables and figures

Table 1: Summary stats

			Female		Male	
	(1) Baseline	(2) Endline	(3) Baseline	(4) Endline	(5) Baseline	(6) Endline
Structure/household-level						
Electricity in house for lighting	0.71	0.77				
Street has lights	0.060	0.079				
Road is paved	0.15	0.25				
HH uses non-latrine toilet	0.51	0.74				
Number of households in dwelling	3.44	2.65				
Rooms household occupies in dwelling	2.45	2.91				
Number of rooms per household member	0.71	0.96				
Monthly rent expected per room, Tsh	36103.8	32297.9				
Own house	0.37	0.45				
Rented	0.56	0.47				
Share of consumption on (imputed) rent	0.17	0.19				
Share of consumption on foot	0.45	0.50				
Share of consumption on transportation	0.11	0.11				
Above WB 5.50 poverty line	0.42	0.44				
N	1517	1800	.	.	.	.
Individuals						
Age			35.4	38.4	38.3	40.0
Years education			7.90	8.05	8.93	9.10
Worked for pay last 7 days			0.29	0.32	0.58	0.54
Operate any self-employed business or activity last month			0.35	0.36	0.51	0.40
Employed (as per occupation variable)			0.51	0.50	0.87	0.82
Mainjob is formal (if employed)			0.21	0.18	0.37	0.27
Mainjob is informal (if employed)			0.79	0.81	0.63	0.72
Typical days worked per week			5.89	5.63	5.85	5.76
Typical hours worked per day			7.77	8.21	9.07	9.64
Wages from the past 7 days (TSH) (1000s TSH)			23.6	16.6	47.7	47.3
Job distance from home greater than 0 (if employed)			0.72	0.57	0.91	0.84
Walk main mode of transportation to get to work (if employed)			0.52	0.57	0.25	0.27
Publtrans main mode of transportation to get to work (if employed)			0.40	0.33	0.70	0.63
Commute time (mins)			33.2	28.5	55.4	99.6
Happiness public transport (scale 1-10)			3.22	6.26	3.20	6.20
Moved house			.	0.40	.	0.39
Moved house (renting BL)			.	0.52	.	0.50
N	.	.	1622	1999	1482	1825

Notes: Table shows summary stats for the sample of tracked male and female respondents. Commuting variables are conditional on being employed (as per classification variable). Statistics weighted by constructed survey weight.

Table 2: BRT usage

	Male		Female	
	Low exposure	High exposure	Low exposure	High exposure
BRT usage				
Have you ever ridden on the BRT?	0.75	0.96	0.63	0.98
Brt used to get to work (if employed)	0.04	0.31	0.02	0.13
Brt main mode of transportation to get to work (if employed)	0.01	0.22	0.01	0.13
Daladala main mode of transportation to get to work (if employed)	0.55	0.45	0.33	0.16
Walk main mode of transportation to get to work (if employed)	0.27	0.23	0.54	0.62
Used BRT to get to work last 7 days	0.11	0.33	0.06	0.14
Used BRT for shopping last 7 days	0.10	0.30	0.08	0.33
Used BRT to get to education last 7 days	0.02	0.06	0.01	0.03
Used BRT to get to health clinic last 7 days	0.05	0.25	0.09	0.22
Used BRT for leisure last 7 days	0.10	0.30	0.05	0.20
Used BRT to visit family last 7 days	0.14	0.29	0.13	0.36
Used BRT for other reasons last 7 days	0.02	0.00	0.01	0.02
Reduction in predicted traveltime				
Demeaned reduction in tt (phases 1-4) using BL speeds	-0.02	0.16	-0.02	0.16
Reduction in log travel time (log phase x - log phase 0) using BL speeds	-0.03	-0.22	-0.03	-0.22
Mean change log phases 1-4 using BL speeds	-0.05	-0.07	-0.05	-0.07
N	1270	302	1376	346

Notes: Table shows summary stats for male and female respondents, measured at EL. Treated is households with larger than 10% decrease in (demeaned) travel time; approximately the 80th percentile of the distribution. Ever used BRT, number of trips on BRT, and whether the BRT is used for purpose x in the last 7 days are tabulated for entire sample. Mode transportation to work questions are tabulated conditional on working. Statistics weighted by constructed survey weight.

Table 3: Change in structure characteristics

	(1) Self-reported time to CBD (minutes)	(2) Log rent per room (expected)	(3) Monthly household expenditures on transportation (in 1,000 TSH)	(4) Different household in structure at EL
Demeaned reduction in tt (phases 1-4) using BL speeds	-94.709	1.872	33.605	-0.111
	20.610***	0.614***	20.048*	0.224
N	1284	1213	1202	1567
Mean EL value (in levels)	70.689	9.887	58.004	0.099
Effect for average highly-exposed	-9.57	0.19	3.39	-0.01

Notes: Dependent variable is change unless noted. Observation is a structure at endline. Standard errors clustered at aggregate spatial unit. De-meaned variable is equal to predicted decrease in travel time to downtown along BRT Phase 1 minus average predicted decrease in travel time to downtown along planned BRT Phases 1-4. There are no significant differences when looking at men and women separately. The effect for average highly-exposed is the coefficient evaluated at 0.101, the mean exposure for a highly-exposed household (as defined in the text).

Table 4: Treatment effects on wages and employment: by gender

	Employment (All)			Earnings (All)		
	(1) Structure	(2) Initial resident	(3) Final resident	(4) Structure	(5) Initial resident	(6) Final resident
Household						
Reduction in log traveltime (demeaned)	0.649	0.202	0.295	57.259	34.283	32.410
	0.229***	0.170	0.227	19.354***	29.204	20.437
N	1320	1399	1088	1307	1380	1074
P-value (structure=individual)		0.117			0.512	
Mean EL value	0.663	0.688	0.684	25.571	29.135	26.374
Average effect highly-exposed	0.066	0.020	0.030	5.783	3.463	3.273
Males						
Reduction in log traveltime (demeaned)	-0.274	-0.220	0.114	7.590	48.010	37.541
	0.278	0.221	0.223	46.233	39.446	43.855
N	944	958	772	901	916	737
P-value (structure=individual)		0.879			0.506	
Mean EL value	0.816	0.841	0.833	42.061	46.534	44.014
Average effect highly-exposed	-0.028	-0.022	0.011	0.767	4.849	3.792
Females						
Reduction in log traveltime (demeaned)	1.106	0.655	0.619	82.235	32.724	40.584
	0.206***	0.283**	0.360*	22.223***	21.926	28.432
N	1146	1117	907	1131	1100	893
P-value (structure=individual)		0.197			0.113	
Mean EL value	0.543	0.550	0.559	14.686	15.059	14.055
Average effect highly-exposed	0.112	0.066	0.063	8.306	3.305	4.099

Notes: Dependent variable is change unless noted. Observation is a structure at endline. Standard errors clustered at aggregate spatial unit. De-meaned variable is equal to predicted decrease in travel time to downtown along BRT Phase 1 minus average predicted decrease in travel time to downtown along planned BRT Phases 1-4.

Table 5: Baseline and endline differences between initial and final residents

	(1) Years of education	(2) Age	(3) Employed	(4) Wages (1000 TSH)	(5) Commutes	(6) Total spending (1000 TSH)
Males: BL-BL						
Demeaned reduction in tt (phases 1-4) using BL speeds	-0.023	-0.148	-0.021	13.747	0.056	-1.816
	0.025	0.158	0.165	35.404	0.086	104.429
N	1035	1066	1224	1201	1027	1234
Mean EL value						
Average effect highly-exposed	-0.002	-0.015	-0.002	1.388	0.006	-0.183
Females: BL-BL						
Demeaned reduction in tt (phases 1-4) using BL speeds	-0.165	0.564	0.446	54.064	0.273	83.957
	0.136	0.491	0.154***	27.290**	0.160*	142.153
N	1126	1148	1373	1362	792	1388
Mean EL value						
Average effect highly-exposed	-0.017	0.057	0.045	5.461	0.028	8.480
Males: EL-EL						
Demeaned reduction in tt (phases 1-4) using BL speeds	3.293	-2.317	-0.253	-50.408	0.274	48.065
	1.584**	2.380	0.142*	36.917	0.236	62.578
N	852	849	859	821	496	1234
Mean EL value						
Average effect highly-exposed	0.333	-0.234	-0.026	-5.091	0.028	4.855
Females: EL-EL						
Demeaned reduction in tt (phases 1-4) using BL speeds	2.934	2.048	0.377	37.056	0.556	-166.869
	1.097***	4.219	0.263	23.122	0.261**	79.391**
N	1011	1014	1029	1011	361	1388
Mean EL value						
Average effect highly-exposed	0.296	0.207	0.038	3.743	0.056	-16.854

Notes: Dependent variable is change unless noted. Observation is a structure at endline. Standard errors clustered at aggregate spatial unit. De-meaned variable is equal to predicted decrease in travel time to downtown along BRT Phase 1 minus average predicted decrease in travel time to downtown along planned BRT Phases 1-4. Years of education is measured as the change in years of education between BL respondent and EL respondent (structure-level, endline data - not collected retrospectively). Age is the change in age between BL respondent and EL respondent (structure-level, endline data - not collected retrospectively). Employed is the change in whether individual is employed between BL respondent and EL respondent (structure-level, retrospective data). Wages is the change in individual's wages (1000 TSH) between BL respondent and EL respondent (structure-level, retrospective data). Commutes is the change in whether individual commutes between BL respondent and EL respondent (structure-level, retrospective data). Total spending is the change in total spending (1000 TSH) between BL respondent and EL respondent (structure-level, retrospective data).

Table 6: BRT usage and preferences in high-exposure locations, by gender

	All		Employed	
	(1) Male	(2) Female	(3) Male	(4) Female
BRT usage				
Have you ever ridden on the BRT?	0.96	0.98	0.98	0.99
Brt used to get to work (if employed)	0.31	0.13	0.32	0.14
Used BRT to get to work last 7 days	0.33	0.14	0.40	0.21
Used BRT for shopping last 7 days	0.30	0.33	0.36	0.36
Number BRT trips last 7 days	3.15	1.98	3.65	2.40
Agree with statements				
BRT station is easily accessible	6.76	6.72	6.75	6.80
BRT driver/vehicle safe	7.99	8.33	7.84	8.11
BRT is cheap	5.60	5.66	5.64	5.44
BRT goes to places useful to me	7.60	7.31	7.73	7.03
BRT makes easier daily jobs done	7.41	7.49	7.46	7.70
BRT improves public transport options	7.00	7.74	6.89	7.63
Main constraint from using BRT				
Route not useful	0.20	0.21	0.18	0.22
Service not reliable	0.17	0.12	0.16	0.14
Limited accessibility station	0.12	0.11	0.13	0.11
Vehicle or driver safety, personal security, harrassment	0.01	0.04	0.00	0.06
N	228	273	182	172

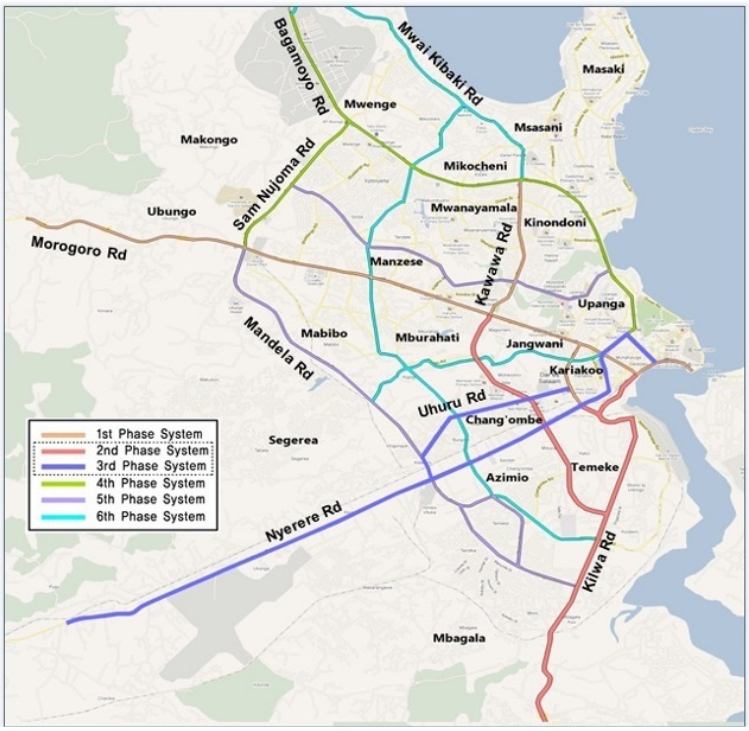
Notes: Responses measured on a scale of 1-10, 10 is favorable. Sample is individuals living in high-exposure locations, measured at endline. Ever used BRT, number of trips on BRT, and whether the BRT is used for purpose x in the last 7 days are tabulated for entire sample. Mode transportation to work questions are tabulated conditional on working.

Table 7: Mechanisms by gender

	Happy public transport		Live convenient		Difficult congestion		Commutes		Walks to work		Public transport to work	
	(1) Female	(2) Male	(3) Female	(4) Male	(5) Female	(6) Male	(7) Female	(8) Male	(9) Female	(10) Male	(11) Female	(12) Male
Reduction in log traveltime (demeaned)	6.016	2.187	6.518	3.589	-7.266	-5.326	0.136	0.258	0.870	-0.059	-0.307	0.112
	2.456**	2.172	2.163***	2.662	2.016***	2.367**	0.340	0.385	0.234***	0.303	0.209	0.298
N	1116	954	1093	886	1093	886	1116	954	1116	954	1116	954
P-value (structure=individual)												
Mean EL value	5.880	5.690	6.075	6.159	6.059	6.501	0.186	0.512	0.255	0.166	0.160	0.563
Average effect highly-exposed												

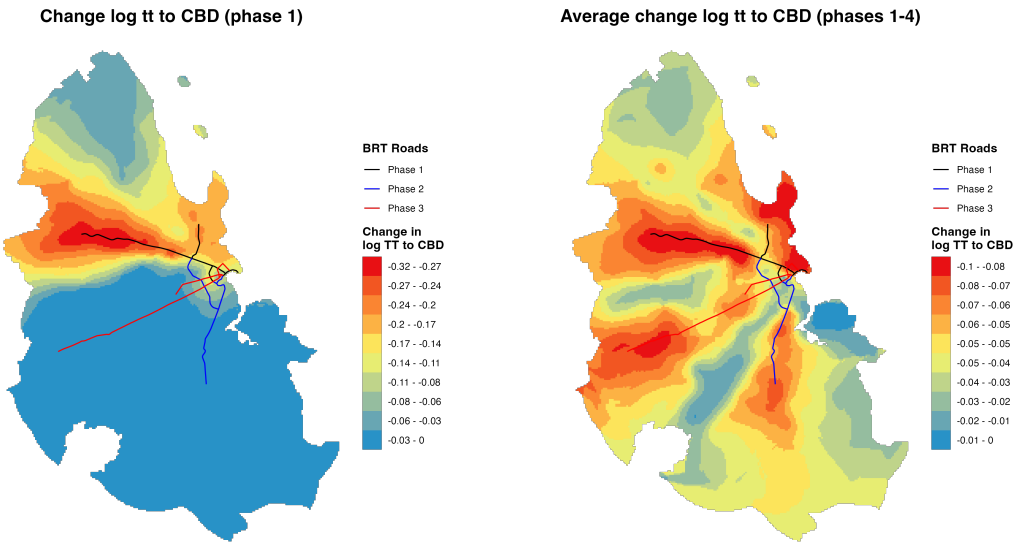
Notes: Unit of analysis is an initial resident. Variables collected at both baseline and endline. Commuting and mode of commuting are extensive margin variables i.e., set to zero if the person does not work.

Figure 1: BRT network in Dar es Salaam

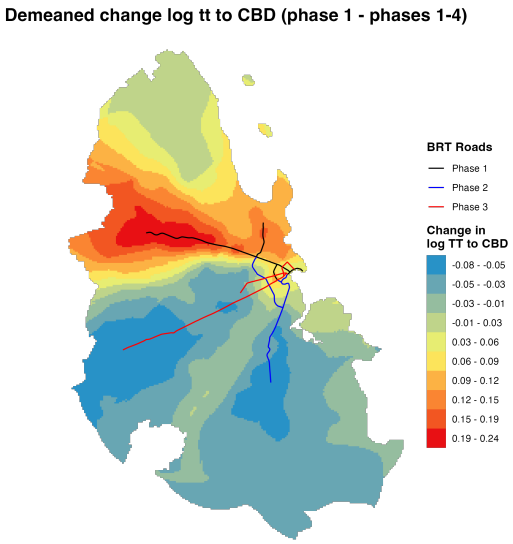


Notes: The figure shows the planned 6-phase BRT network in Dar es Salaam. We examine the impact of Phase 1 that runs along Morogoro Road. Image source: Dar Rapid Transit (DART).

Figure 2: Reduction in travel time to CBD due to BRT



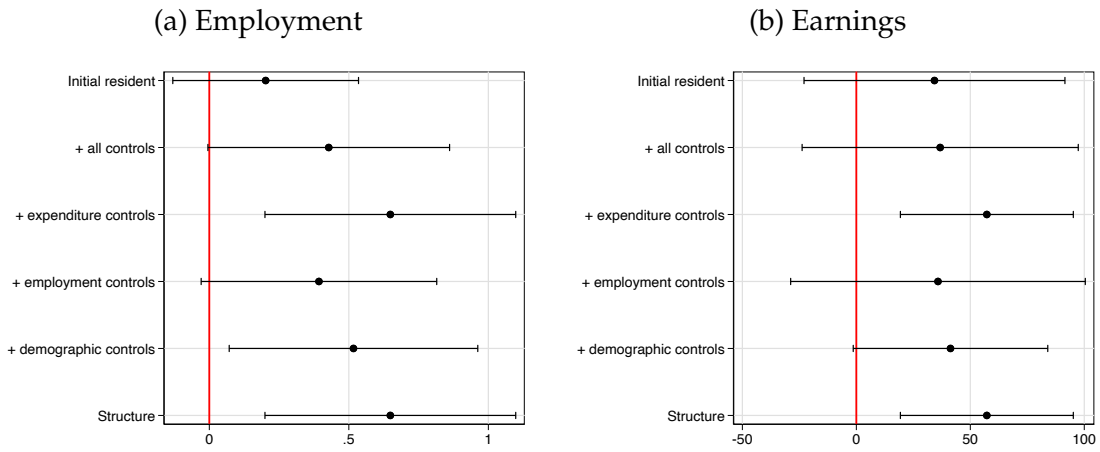
(a) Reduction from Phase 1 (b) Average reduction across Phases 1-4



(c) Demeaned travel time

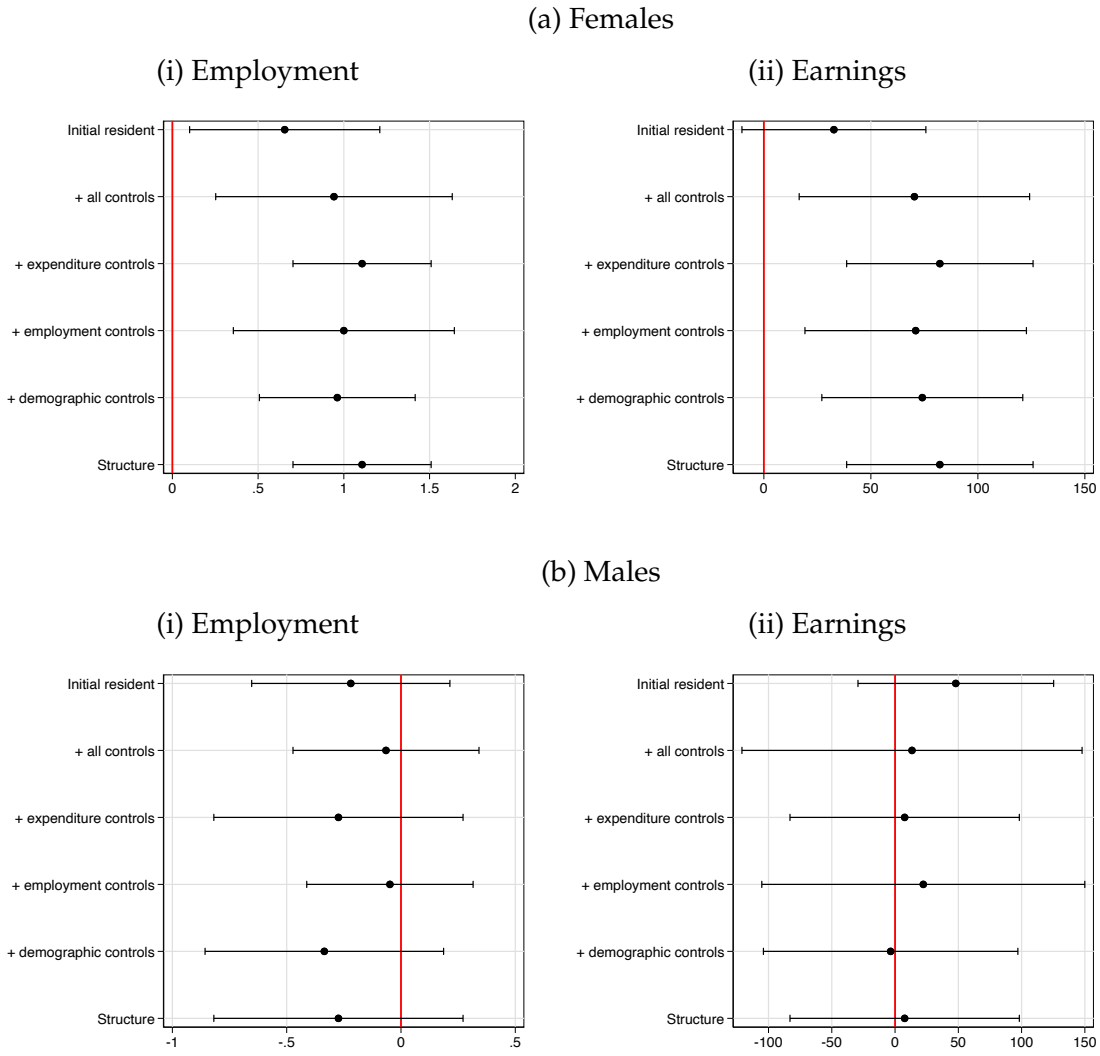
Notes: The figure shows the calculated reduction in travel time to the CBD. Panel (a) shows the raw value. Panel (b) shows the average reduction in travel time to CBD from Phases 1-4. Panel (c) shows the value after demeaning travel time by the average decrease for Phases 1 to 4, following Borusyak and Hull (2023)

Figure 3: Decomposing the selection effect



Notes: The figure plots the point estimate and 95th percentile confidence interval of the demeaned coefficient for the average individual. Individual is the estimated treatment effect for the individual living in the structure at baseline. Structure is the estimated treatment effect for the structure. Selection is accounted for by controlling for the difference in endline characteristics of the baseline and endline respondents with the structure. Demographic controls are demographic changes (education, age). Employed controls for earnings and commute status. Ependiture controls are controls for household consumption. All controls includes all controls listed above.

Figure 4: Decomposing the selection effect: by gender



Notes: The figure plots the point estimate and 95th percentile confidence interval of the demeaned coefficient. Individual is the estimated treatment effect for the individual living in the structure at baseline. Structure is the estimated treatment effect for the structure. Selection is accounted for by controlling for the difference in endline characteristics of the baseline and endline respondents with the structure. SelectionXdemog controls for demographic changes (education, age). SelectionXemployed controls for earnings and commute status. SelectionXcons controls for household consumption. SelectionXall controls for all differences.

APPENDIX: FOR ONLINE PUBLICATION ONLY

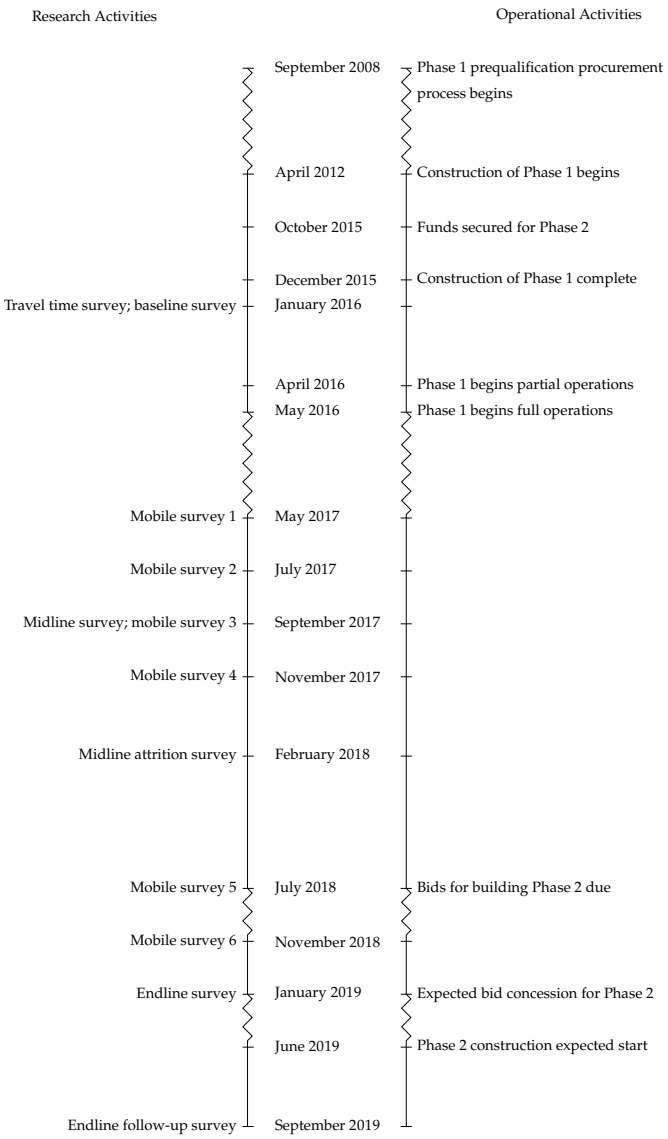
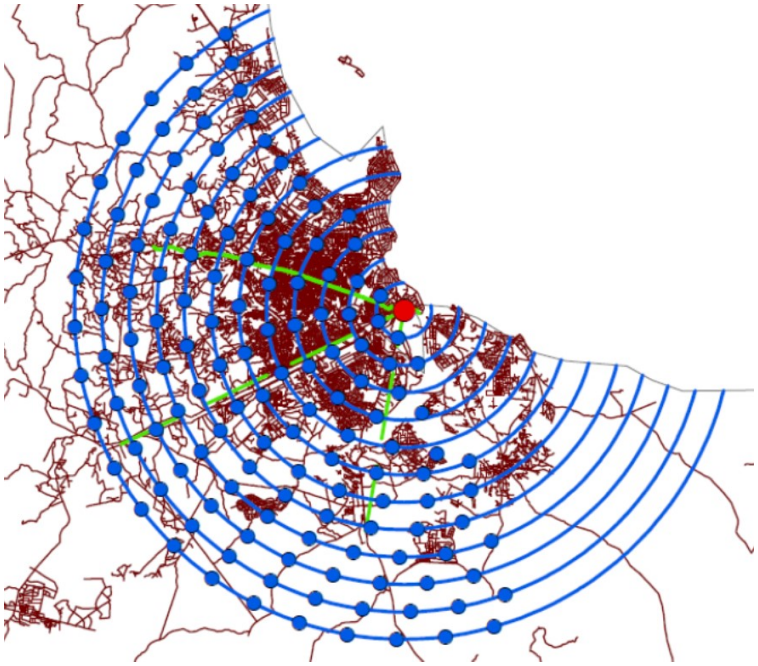


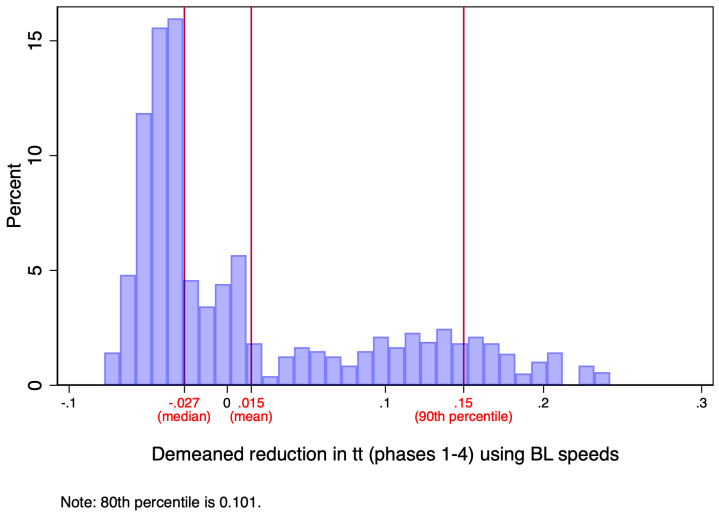
Figure A1: Timeline of activities

Figure A2: Spatial sampling frame



Notes: The figure shows the spatial sampling frame used to construct the sample. We constructed 12 arcs at radii increasing at 1.5km intervals from the CBD, shown in the figure as a red dot. We enrolled households in 125 of the 141 clusters. Each cluster enrolled 12-14 households, yielding a sample of 1748 households.

Figure A3: Distribution of demeaned travel time



Notes: This figure shows the distribution of demeaned travel time to the CBD. An unit of observation is a baseline structure.

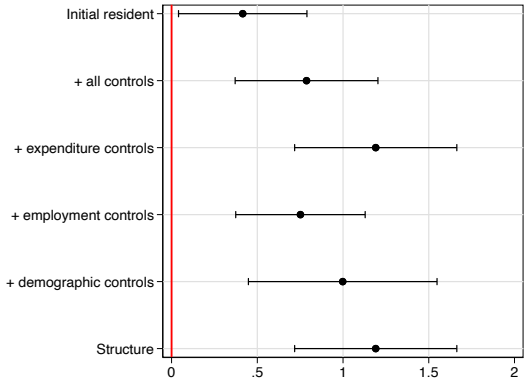
Table A1: Treatment effects on wages and employment: by gender and rental status

	Employment (All)			Earnings (All)			Employment (Renters)			Earnings (Renters)		
	(1) Structure	(2) Initial resident	(3) Final resident	(4) Structure	(5) Initial resident	(6) Final resident	(7) Structure	(8) Initial resident	(9) Final resident	(10) Structure	(11) Initial resident	(12) Final resident
Household												
Reduction in log traveltime (demeaned)	0.649	0.202	0.295	57.259	34.283	32.410	1.191	0.415	-0.196	64.071	24.264	-6.223
	0.229***	0.170	0.227	19.354***	29.204	20.437	0.241***	0.191**	0.336	37.871*	45.525	42.326
N	1320	1399	1088	1307	1380	1074	259	277	143	256	274	141
P-value (structure=individual)		0.117			0.512			0.012			0.501	
Mean EL value	0.663	0.688	0.684	25.571	29.135	26.374	0.659	0.731	0.738	30.014	34.348	33.884
Average effect highly-exposed	0.066	0.020	0.030	5.783	3.463	3.273						
Males												
Reduction in log traveltime (demeaned)	-0.274	-0.220	0.114	7.590	48.010	37.541	-0.297	-0.095	0.151	-21.493	74.825	-8.536
	0.278	0.221	0.223	46.233	39.446	43.855	0.430	0.437	0.409	83.055	82.671	87.710
N	944	958	772	901	916	737	165	186	97	157	177	93
P-value (structure=individual)		0.879			0.506			0.741			0.411	
Mean EL value	0.816	0.841	0.833	42.061	46.534	44.014	0.858	0.909	0.918	46.640	57.429	55.328
Average effect highly-exposed	-0.028	-0.022	0.011	0.767	4.849	3.792						
Females												
Reduction in log traveltime (demeaned)	1.106	0.655	0.619	82.235	32.724	40.584	1.761	1.050	1.184	112.701	36.529	49.978
	0.206***	0.283**	0.360*	22.223***	21.926	28.432	0.314***	0.416**	0.673*	29.819***	25.248	42.432
N	1146	1117	907	1131	1100	893	202	211	114	202	210	114
P-value (structure=individual)		0.197			0.113			0.172			0.051	
Mean EL value	0.543	0.550	0.559	14.686	15.059	14.055	0.475	0.517	0.518	17.252	14.442	14.529
Average effect highly-exposed	0.112	0.066	0.063	8.306	3.305	4.099						

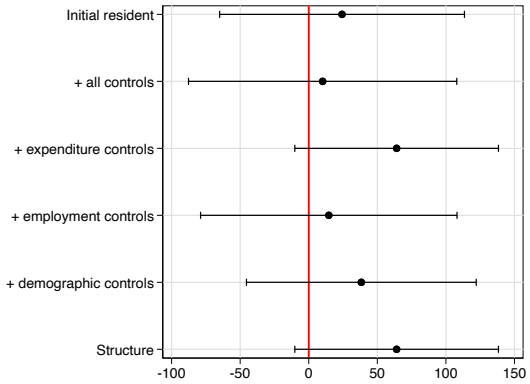
Notes: Dependent variable is change unless noted. Observation is a structure at endline. Standard errors clustered at aggregate spatial unit. De-meaned variable is equal to predicted decrease in travel time to downtown along BRT Phase 1 minus average predicted decrease in travel time to downtown along planned BRT Phases 1-4.

Figure A4: Decomposing the selection effect: renters

(i) Employment



(ii) Earnings



Notes: The figure plots the point estimate and 95th percentile confidence interval of the demeaned coefficient. Individual is the estimated treatment effect for the individual living in the structure at baseline. Structure is the estimated treatment effect for the structure. Selection is accounted for by controlling for the difference in endline characteristics of the baseline and endline respondents with the structure. SelectionXdemog controls for demographic changes (education, age). SelectionXemployed controls for earnings and commute status. SelectionXcons controls for household consumption. SelectionXall controls for all differences.

A Survey recontact rates

Table A2: Contact rates of sample at endline

	Baseline sample		Ever enrolled	
	(1) All	(2) Drop ML untracked	(3) All	(4) Drop ML untracked
Structures				
Found and survey complete	76.0	87.5	67.5	75.7
Found and refused /incomplete	10.0	3.6	9.6	3.7
Not found	8.2	2.1	18.1	15.3
Torn down / empty	5.8	6.7	4.7	5.3
N	1748	1517	2511	2241
Households				
Found and survey complete	79.6	92.2	70.8	80.5
Found and refused/incomplete	9.2	2.3	9.4	3.6
Not found	11.2	5.6	18.4	15.9
Splitoff returned original hh / left Dar			1.4	
N	1748	1510	2542	2237
Male/female respondents				
Found and survey complete	69.7	81.3	73.8	83.4
Found and refused /incomplete	13.3	6.0	11.8	5.7
Not found	15.1	10.6	13.0	9.2
Died	1.8	2.1	1.5	1.7
N	3104	2662	3891	3441

Notes: Table shows percent in each status. We initially enrolled 1748 structures and households. Up to two individual respondents (one male, one female) were enrolled per household. We stopped tracking structures after midline if they (i) refused, or (ii) we were not able to locate either the structure or the household after exhausting all contact information available. We did not attempt to find non-tracked sample at endline.

Table A3: Individual-Level Attrition balance for households enrolled at baseline

	Endline					Endline mop up	
	(1) Complete	(2) Incomplete	(3) Refused	(4) Not found	(5) Died	(6) Surveyed	(7) Surveyed if attempted
Log BL HH monthly gross income per capita	-0.028**	0.005	0.011	0.018**	-0.006	-0.024*	-0.006
	(0.012)	(0.005)	(0.009)	(0.007)	(0.004)	(0.012)	(0.011)
Log BL monthly rent per room	-0.020	0.002	0.012	0.006	0.001	-0.016	-0.000
	(0.013)	(0.005)	(0.009)	(0.006)	(0.004)	(0.012)	(0.012)
BL number of years lived in the structure	-0.001	-0.000	0.001	-0.000	0.000	-0.001	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
BL number of HH in the structure	0.010	-0.004**	0.002	-0.005	-0.002*	0.011*	0.010*
	(0.006)	(0.002)	(0.004)	(0.005)	(0.001)	(0.006)	(0.005)
Dwelling rented by HH	-0.135***	-0.003	-0.020	0.139***	0.019**	-0.096**	0.003
	(0.037)	(0.013)	(0.022)	(0.028)	(0.009)	(0.046)	(0.043)
Dwelling owned by HH	0.011	0.017	0.015	-0.049**	0.006	0.038	0.014
	(0.030)	(0.011)	(0.019)	(0.020)	(0.008)	(0.035)	(0.031)
BL distance to phase 1	0.002	-0.001	-0.002	0.001	-0.000	0.004	0.004*
	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.003)
BL distance to phase 2	-0.001	0.002**	0.000	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
BL Age	-0.000	0.000	-0.000	-0.001***	0.002***	0.001*	0.001
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Male dummy	-0.060***	0.030***	0.012*	0.009*	0.008*	0.003	0.026*
	(0.014)	(0.007)	(0.007)	(0.005)	(0.005)	(0.016)	(0.015)
N	2646	2646	2646	2646	2646	2646	2212
Mean dependent variable	0.697	0.034	0.100	0.151	0.018	0.589	0.786

Notes: Standard errors, clustered by aggregate spatial units, are reported in parentheses. Missing values for the independent variables are dummied out. Refused includes male/female respondents that were tracked at EL but refused to be surveyed and those that were not tracked at EL because either the household or male/female respondent refused at ML. Not found includes male/female respondents that were tracked but not found at EL and those that were not tracked at EL because the household was not found at ML.

Table A4: Household-Level Attrition balance for households enrolled at baseline

	Endline			Endline mop up	
	(1) Complete	(2) Refused	(3) Not found	(4) Surveyed	(5) Surveyed if attempted
Log BL HH monthly gross income per capita	-0.026*	0.010	0.016*	-0.032**	-0.011
	(0.013)	(0.009)	(0.009)	(0.013)	(0.010)
Log BL monthly rent per room	-0.022	0.016	0.006	-0.019	0.000
	(0.013)	(0.010)	(0.005)	(0.015)	(0.012)
BL number of years lived in the structure	-0.000	0.001	-0.001	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
BL number of HH in the structure	0.005	0.002	-0.006	0.011*	0.009*
	(0.006)	(0.004)	(0.005)	(0.006)	(0.005)
Dwelling rented by HH	-0.134***	-0.010	0.144***	-0.077	0.047
	(0.037)	(0.021)	(0.029)	(0.049)	(0.044)
Dwelling owned by HH	0.036	0.024	-0.060***	0.060	0.037
	(0.026)	(0.019)	(0.021)	(0.043)	(0.038)
BL distance to phase 1	0.000	-0.001	0.001	0.004	0.004*
	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
BL distance to phase 2	0.001	0.001	-0.002	0.001	0.000
	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)
N	1607	1607	1607	1607	1361
Mean dependent variable	0.796	0.092	0.112	0.682	0.858

Notes: Standard errors, clustered by aggregate spatial units, are reported in parentheses. Missing values for the independent variables are dummied out. Refused includes households that were either tracked at EL but refused to be surveyed or not tracked at EL because they refused to be surveyed at ML. Not found includes households that were tracked but not found at EL, not contacted at because EL we ran out of time, or not tracked at EL because either they were not found at ML or they splitoff outside Dar at ML. We dropped households that splitoff from BL structure at ML but returned to original HH at EL.

Table A5: Structure-Level Attrition balance for households enrolled at baseline

	Endline					Endline mop up	
	(1) Complete	(2) Incomplete	(3) Refused	(4) Not found	(5) Torn down/empty	(6) Surveyed	(7) Surveyed if attempted
Log BL HH monthly gross income per capita	-0.032***	0.001	0.004	0.022***	0.009	-0.027**	0.001
	(0.010)	(0.001)	(0.008)	(0.007)	(0.006)	(0.011)	(0.010)
Log BL monthly rent per room	-0.035***	-0.000	0.020**	0.007	0.005	-0.031**	-0.003
	(0.013)	(0.001)	(0.008)	(0.006)	(0.008)	(0.014)	(0.013)
BL number of years lived in the structure	0.002	-0.000*	0.001	-0.001**	-0.001	0.002	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
BL number of HH in the structure	0.006	0.000	-0.002	-0.009*	0.002	0.010	0.005
	(0.006)	(0.001)	(0.004)	(0.005)	(0.004)	(0.006)	(0.007)
Dwelling rented by HH	-0.145**	-0.002	-0.007	0.134***	0.008	-0.119**	-0.011
	(0.055)	(0.006)	(0.028)	(0.028)	(0.025)	(0.055)	(0.044)
Dwelling owned by HH	0.100**	-0.006	-0.006	-0.051**	-0.045**	0.128**	0.051
	(0.046)	(0.005)	(0.023)	(0.022)	(0.019)	(0.051)	(0.036)
BL distance to phase 1	0.004	0.000	-0.002	0.000	-0.001	0.007**	0.004
	(0.003)	(0.000)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
BL distance to phase 2	0.005	-0.000	-0.002	-0.003	-0.001	0.003	-0.000
	(0.004)	(0.000)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)
N	1748	1748	1748	1748	1748	1748	1325
Mean dependent variable	0.760	0.002	0.069	0.082	0.058	0.648	0.855

Notes: Standard errors, clustered by aggregate spatial units, are reported in parentheses. Missing values for the independent variables are dummied out. Incomplete includes structures that were found but whose household either did not complete the survey or refused to be surveyed. Refused includes structures that were either tracked and found at EL but the household refused to be surveyed at EL or not tracked at EL because the household refused to be surveyed at ML. Not found includes structures that were tracked but not found at EL, structures we did not look for at EL because we ran out of time, or structures that were not tracked at EL because they were not found at ML. We dropped observations for structures whose households splitoff from their BL structure at ML but returned to original HH at EL and those that were not tracked because they were outside Dar.

B Construction of household weights

We construct two weights: an observable-adjusted weight and a population weight. The observable-adjusted weight reweights our sample to be balanced on ownership structure and ward population. The weight sums to the number of households we survey. The population weight purely adjusts for the population distribution and sums to the number of households in the surveyed area (Dar es Salaam and two wards in Pwani district).

To measure ownership structure, we use the variable StatDwel. Though the response translations remained the same across all survey rounds (i.e. asking whether the structure is owned by a household, rather than by your household), the translation of the survey question changed a little at endline. Rather than asking about the status of the household ownership in general, it seems to be more directed at the respondent we're talking to; here are the Google translations, to give you a rough idea: "What is your legal ownership in these settlements?" at EL, versus "What is the legal status of the use of this residence?" at BL.

For everything that follows, we define owner and renter based on the household respondent's answer to the question "what is the legal status of the household ownership?" as this is the question used when determining the tenant and sub-letter module skips in the survey.

First, we have three questions of interest, all of which are only asked at endline:

- To renters: Does the owner live in this dwelling?
- To renters who say there are more than one households living in their dwelling: for each additional household over 1, the respondent is asked how many rooms tenant household x occupies, and asked how many total people live in tenant household x
- To owners or households living in their dwelling for free: Do you sublet any portion of this dwelling to other HHs? And if so, how many households do you rent out to?

Some numbers using the responses to these questions (again, all of these numbers are endline-only):

- 31% of renters live in their dwelling with the owner (N = 476 HH)
- 22% of owners sublet the dwelling they're living in (N = 1,167 HH)
- In 60% of the dwellings for which we know there were both owners and renters living together (N = 425 structures), we surveyed the owner

Note: all of the following tables only include endline data, given that we can only classify households as follows if we use endline data (as stated above).

B.0.1 Generating endline weights

Define a structure s such that the total population of structures in our data at time t, $t = \{0, 1\}$ is n_t. Within each of these structures, define survey household a_s, which is indexed by two additional dimensions: $b \in \{owner, resident\}$ and $c = \{1, 2, ..., 5+\}$, where c is a categorical variable for the number of households in structure s. N_hh_{st} is the total number of households living in structure s at time t, a inclusive.

Then, the total population of households at time $t = \sum_s N_hh_{st}$. Let I be the set of all households. At each time period t, we surveyed a subset $x \in I$, $x = \{a_1, a_2, ..., a_{n_t}\}$.

Define our actual surveying probability as $\tilde{\pi}_{bct}$ and the population probability as $\pi_{bct} \forall b, c$.

Then, the weights are given by $\Omega_{bct} = N_hh_{st}(\pi_{bct}/\tilde{\pi}_{bct})$. If a renter household lives in a structure with all renters, $\pi_{bct} = \tilde{\pi}_{bct}$ and thus $\Omega_{bct} = N_hh_{st}$.

B.0.2 Extending endline weight definition to baseline

At baseline, we did not ask renters who lived in dwellings with more than one household whether the owner was one of the other households that lived in the dwelling. As a result, we cannot determine the rate at which we over- or under-sampled owners at baseline. Thus, we ammend the previous section such that π_{bc} and $\tilde{\pi}_{bc}$ are both time-invariant, with $\pi_{bc} = \pi_{bc,t=1}$ and $\tilde{\pi}_{bc} = \tilde{\pi}_{bc,t=1}$.

B.0.3 Incorporating population weights using 2012 Census data

We know the number of individuals living in each of our 52 aggregate spatial units from the 2012 Tanzania Census; however, we want to know how many households reside in each unit. To convert from individuals to households, we divide the total number of individuals by the average number of members in a household in Dar (3.9), per the 2011/2012 Tanzania HIV/AIDS and Malaria Indicator Survey (AIS) (source: DHS statcompiler https://www.statcompiler.com/en/; see Tanzania_Paper/RawData/mean_num

B.0.4 Additional notes

In our survey, we group households that live in their structure for free with households that own their structure when asking questions related to the dwelling, i.e. do you sublet any part of this dwelling. To be consistent, we do the same when generating weights.

B.0.5 Weight variables in our analysis datasets

- owner_weight = $\pi_{bc}/\tilde{\pi}_{bc}$
- hh_size_weight = N_hh_{st}
- combined_weight = $\Omega_{bct} = N_hh_{st} * (\pi_{bc}/\tilde{\pi}_{bc})$
- population_weight = Λ_i, i = aggregate spatial unit of residence
- combined_weight_census = $\Lambda_i * \Omega_{bct}$; this is our preferred weight

For both of the combined weights listed above, we have two additional sets in our data: one which ends with _struc, and another which ends with _hh. These variables are equal to the endline combined weight variables, and are applied to both baseline and endline observations for the structure/household.

Table A6: Composition of household types

	BL data (BL weight)		EL data (EL weight)		EL data (BL weight)	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Owned	65.08	34.38	65.01	36.79	65.01	44.72
Lived in without paying rent	9.04	5.67	8.80	5.01	8.80	8.42
Rented	25.87	59.95	26.18	58.21	26.18	46.85
N	1747	1747	1795	1701	1795	1795

Notes: BL weight is computed from the distribution of owners and number of households at BL. EL weight is computed from the distribution of owners and number of households at EL. I prefer the BL weight as we can only compute EL weight for ppl we survey at EL, so that means any household that has attrited does not have a weight. We could potenitally reweight the EL sample though to make it look more similar to the BL one.

C Pretrend analysis

Table A7: Pre-trends analysis

	(1) Residualized log rent per room (expected)	(2) Lived in house < 3 yrs	(3) Typical number of days worked per week	(4) Typical number of hours worked per day	(5) Wealth index (PCA)
De-meaned pred. decrease TT CBD Ph1	0.872	0.466	-0.918	-0.650	0.092
	0.581	0.562	0.923	2.249	0.886
N	92	92	92	92	92
Mean BL value	0.219	0.308	5.924	8.744	-0.129

Notes: Unit of observation is EA. Standard errors clustered at relevant aggregate spatial unit. Treatment is defined at the EA level, where pre-period structures lie within an EA and post-period structures are assigned to the nearest EA, regardless of distance. Pre-period data come from WB Measuring Living Standards in Cities 2014-2015 survey, and post-period data come from Dar BRT BL survey.

D Calculating travel speeds

We calculate travel speeds for travel by (1) car, (2) public transport excluding BRT, and (3) public transport including BRT (endline only).

Baseline travel speeds At baseline, we use the travel time survey conducted in January 2016 to calculate the average travel speed by car and by public transport (the latter is the average of speeds by daladala and bajaji - rickshaw). The average travel speed by taxi - private car is 16.6185 km/hr = 4.61625 m/s, while the average travel speed by public transport is 13.4261 km/hr = 3.72947 m/s.

Endline travel speeds We obtain the average travel speed along the BRT routes using data from DART. DAR's 2007 planning document gives the length of each BRT route. DART also provided data on travel times along five of the seven BRT routes each month from November 2017 - April 2018. We use the data from January 2018 to align with the month of the baseline travel time survey (since there may be some seasonality in travel speeds, for example due to monsoon timing). This gives an average travel speed of 19.565 km/hr = 5.43472 m/s.

E Deriving the selection effect

This section derives the selection terms.

Recall, we have three objects of interest:

1. Impact of BRT on dwelling (neighborhood)

$$\beta^D = \Delta y_{j \to j',D} = y^{t+1}_{j',D} - y^t_{j,D}$$

2. Impact of BRT on initial residents

$$\beta^I = \Delta y_{j,I \to f} = y^{t+1}_{j,f} - y^t_{j,I}$$

3. Impact of BRT on final residents

$$\beta^F = \Delta y_{j',i \to F} = y^{t+1}_{j',F} - y^t_{j',i}$$

It is easy to show that the dwelling effect is the initial resident effect and a selection effect measured at $t+1$:

$$\begin{aligned}
\beta^D = \Delta y_{j \to j',D} &= y^{t+1}_{j',D} - y^t_{j,D} \\
&= y^{t+1}_{j',D} - y^t_{j,D} + y^{t+1}_{j,f} - y^{t+1}_{j,f} \\
&= \underbrace{\left(y^{t+1}_{j,f} - y^t_{j,D}\right)}_{\text{initial resident}} + \underbrace{\left(y^{t+1}_{j',D} - y^{t+1}_{j,f}\right)}_{\text{selection effect, pd t+1}} \\
&= \beta^I + \sigma^{t+1}_{j,j'}
\end{aligned}$$

Alternatively, it is easy to show that the dwelling effect is the final resident effect and a selection effect measured at t:

$$\begin{aligned}
\beta^D = \Delta y_{j \to j',D} &= y^{t+1}_{j',D} - y^t_{j,D} \\
&= y^{t+1}_{j',D} - y^t_{j,D} + y^t_{j',i} - y^t_{j',i} \\
&= \underbrace{\left(y^{t+1}_{j',D} - y^t_{j',i}\right)}_{\text{final resident}} + \underbrace{\left(y^t_{j',i} - y^t_{j,D}\right)}_{\text{selection effect, pd t}} \\
&= \Upsilon^F + \sigma^t_{j,j'}
\end{aligned}$$